

SANDAG Commercial Vehicle Model and Heavy Truck Model Update

Model Design

prepared for

San Diego Association of Governments

prepared by

Cambridge Systematics, Inc.

SANDAG Commercial Vehicle Model and Heavy Truck Model Update

Model Design Draft Technical Memorandum

prepared for

San Diego Association of Governments

prepared by

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

date

December 2024

Table of Contents

1.0	Introduction	1-1
2.0	Design Objectives and Overview.....	2-1
2.1	CVM Overview and Flow Between Components.....	2-2
2.1.1	Pre-Processes	2-3
2.1.2	Route Generation and Starting Conditions	2-4
2.1.3	Route Stop Simulation Loop	2-6
3.0	CVM Component Design Details	3-8
3.1	Pre-Processes	3-8
3.1.1	Establishment Synthesis.....	3-8
	Method	3-8
	Inputs	3-9
	Outputs	3-10
3.1.2	Household Attraction Generation.....	3-11
	Method	3-11
	Inputs	3-11
	Outputs	3-11
3.1.3	Establishment Attraction Generation	3-12
	Method	3-12
	Inputs	3-14
	Outputs	3-14
3.1.4	Accessibility	3-15
	Method	3-15
	Inputs	3-16
	Outputs	3-17
3.2	Route Generation and Starting Conditions	3-18
3.2.1	Route Generation -- Establishments.....	3-18
	Method	3-18
	Inputs	3-19
	Outputs	3-20
3.2.2	Route Generation -- TNCs.....	3-20

	Method	3-20
	Inputs	3-21
	Outputs	3-21
3.2.3	Route Vehicle, Purpose, and Customer Type Allocation.....	3-22
	Method	3-22
	Inputs	3-23
	Outputs	3-23
3.2.4	Route Start Time Choice	3-23
	Method	3-23
	Inputs	3-26
	Outputs	3-26
3.2.5	Route Origin and Termination Location Choices.....	3-26
	Method	3-28
	TNC Variation	3-29
	Inputs	3-29
	Outputs	3-29
3.3	Route Stop Simulation Loop	3-30
3.3.1	Next Stop Purpose Choice	3-30
	Method	3-32
	Inputs	3-32
	Outputs	3-32
3.3.2	Next Stop Location Choice	3-32
	Method	3-33
	Inputs	3-35
	Outputs	3-35
3.3.3	Stop Duration Simulation	3-35
	Method	3-35
	Inputs	3-38
	Outputs	3-38
3.3.4	Stops and Trips Complete Outputs.....	3-38
4.0	Heavy Truck Model.....	4-39
4.1	HTM Workflow	4-40
4.2	Major HTM Updates.....	4-41

4.3 Detailed Steps in Running the Updated HTM..... 4-42

4.4 HTM Outputs..... 4-43

List of Tables

Table 3.1-1 CVM Establishment Industry Groups	3-9
Table 3.1-2 Establishment Size Class Definitions	3-9
Table 3.1-3 Example of LUZ Target Distributions	3-10
Table 3.1-4 Example of Synthetic Establishment Records	3-10
Table 3.1-5 Household Attractions Variables Added to Household Records	3-12
Table 3.1-6 Household Attraction Variables Added to Land Use Records	3-12
Table 3.1-7 Establishment Attractions Model Group Segmentation.....	3-13
Table 3.1-8 Establishment Attraction Variables Added to Establishment Records.....	3-14
Table 3.1-9 Establishment Attraction Variables Added to Land Use Records	3-14
Table 3.1-10 Accessibility Proportional Usage and Value of Time by Vehicle Type.....	3-17
Table 3.1-11 Establishment Accessibility Variables	3-17
Table 3.1-12 TNC Accessibility Variables	3-18
Table 3.2-1 Establishment Route Generation Model Group Segmentation	3-19
Table 3.2-2 Calculation of TNC Route Expansion Weights.....	3-21
Table 3.2-3 Calculation of TNC Route Generation Rates	3-21
Table 3.2-4 Route Starting Time Choice Business Group Definitions	3-23
Table 3.2-5 CVM Route Attributes.....	3-30
Table 3.3-1 Stop Duration Model Segmentation for Regular Establishments.....	3-37
Table 3.3-2 Stop Duration Model Segmentation for TNCs.....	3-38
Table 3.3-3 CVM Stop and Trip Attributes.....	3-39

List of Figures

Figure 2.1-1 CVM Components..... 2-3

Figure 2.1-2 CVM Route Simulation Process..... 2-6

Figure 3.1-1 Accessibility Impedance Curve Example 3-16

Figure 3.2-1 Route Starting Time Distributions: Bus. Grp. 1 - Agriculture and Utilities Goods/Services Purpose and Maintenance/Other Purpose 3-24

Figure 3.2-2 Route Starting Time Distributions by Vehicle Type: Bus. Grp. 2 - Retail and Various Service Industries Goods/Service Purpose 3-25

Figure 3.2-3 Route Starting Time Distributions by Vehicle Type: Bus. Grp. 3 - Manufacturing, Wholesale & Transportation Goods/Service Purpose 3-25

Figure 3.2-4 TNC Route Starting Time Distributions by Customer Type 3-26

Figure 3.2-5 Route Origin and Termination Location Choice Structure 3-27

Figure 3.3-1 Next Stop Purpose Feasible Choice Sets for Establishment Routes 3-31

Figure 3.3-2 Next Stop Purpose Feasible Choice Sets for TNC Routes..... 3-31

Figure 3.3-3 Next Stop Location Choice Structure 3-33

Figure 3.3-4 Example of "Rubber Banding" in Next Stop Location Choice 3-35

Figure 3.3-5 Observed Stop Durations (15-minute intervals) Regular Establishments..... 3-36

Figure 3.3-6 Observed Stop Durations (15-minute intervals) TNCs..... 3-37

Figure 4.1-1 Previous HTM Workflow in ABM2+ 4-40

Figure 4.1-2 Updated HTM Workflow in ABM3 4-41

1.0 Introduction

The SANDAG Commercial Vehicle Model (CVM) simulates the weekday demand patterns of commercial vehicle movements throughout the San Diego region. The CVM is an important part of the complete travel demand modeling system for the region, representing a market of travel that dominates the middle part of most weekdays and has been steadily growing as consumer demands for home deliveries and personal services have increased.

The primary sources of data for developing the CVM were the 2022 SANDAG Commercial Vehicle Establishment and TNC-driver surveys, which focused on goods, services, and maintenance trips, and obtained travel diaries from employees of these establishments whose jobs involve routine travel for either goods pickup and delivery or for service provision. The TNC driver survey used an identical travel diary format to the Establishment survey, the difference being that individual TNC drivers were surveyed as their own establishments who worked on behalf of an online pickup and delivery service. The Establishment and TNC surveys provided detailed travel pattern data for individual drivers and vehicles, which formed the basis for estimating and calibrating model components. The categorical definitions of attribute variables in the two surveys set the possibilities for segmentation of the model system, such as establishment industry sectors; trip origin and destination purposes, land uses, and place types; and vehicle types.

The geographic scope of the CVM are internal-to-internal trip movements. The market scope of the model includes commercial goods movements (pickup and deliveries) as well as trips made for commercial and public services. Trips made for other purposes, namely maintenance and personal, are also included in the CVM if these trips are made in the context of a commercial vehicle tour pattern. The CVM explicitly distinguishes between residential and non-residential customer types, and between three vehicle types—light, medium, and heavy—consistent with the definitions used in the Establishment Survey.

The CVM does not cover the types of work-related travel that would be expected to be covered in the ABM3 Resident model, namely workers traveling for meetings, sales calls, out-of-town travel, and similar activities. The CVM also does not model long-distance freight truck movements that enter and exit the region, which are covered by the Heavy Truck Model (HTM).

The HTM covers long-distance freight movements into and out of San Diego County. The source of the demand in the HTM are commodity flows between shippers and receivers throughout North America, focusing on those with either a trip end (shipper or receiver) in San Diego County or which pass through San Diego County, for example, between Mexico and Los Angeles. Commodity flows are derived from the Federal Highway Administration (FHWA) Freight Analysis Framework version 5 (FAF5). The model design assumes that freight truck trips between establishments within San Diego County are covered by the CVM, which has been designed to explicitly account for truck movements involving warehouse and distribution centers and port facilities.

2.0 Design Objectives and Overview

The 2024 CVM design was conceived as an improvement upon the SANDAG CVM created in 2014, which used a similar approach for simulating commercial vehicle tours. The new CVM design differs from the previous model in several important ways. Improvement objectives of the new design include:

- Better accuracy for representing total commercial vehicle demand, as represented in the new surveys, particularly more accurate accounting of vehicle miles traveled (VMT) by vehicle size;
- Leveraging the new survey data to capture contemporary e-commerce trends in direct deliveries between warehouse fulfillment centers and consumers;
- Distinguishing between residential and non-residential customer types;
- Creating linkages between the CVM and the Resident model to better represent household demand for food and package deliveries and for services;
- Leveraging the TNC driver survey to represent this recently emerged type of on-line establishment and the “gig” workers it employs.

In addition, the HTM update to the most current version of FAF5 was important to representing commodity trading and supply chain trends after the COVID-19 pandemic. The HTM update also includes improvements to process efficiency and accuracy.

2.1 CVM Overview and Flow Between Components

At its core, the CVM is a dynamic simulation of commercial vehicle travel patterns during a representative weekday in the San Diego region. The CVM creates travel patterns composed of trips to stop locations, with each stop defined by a purpose, customer type if appropriate, the vehicle type used, geographic location (MGRA and TAZ identifications), and arrival and departure times. The model uses the term “route” to describe a day’s worth of trips made by the same vehicle and is consistent with the operational notion of commercial vehicle route planning and scheduling. A commercial vehicle route is comprised of one or multiple stops and may include one or more full or partial tours. The CVM is dynamic in the sense that it generates routes as simulation elements with starting conditions, then simulates the stops and trips on the route incrementally while considering time of day, elapsed time, and the travel distance to a pre-chosen terminal location for the route.

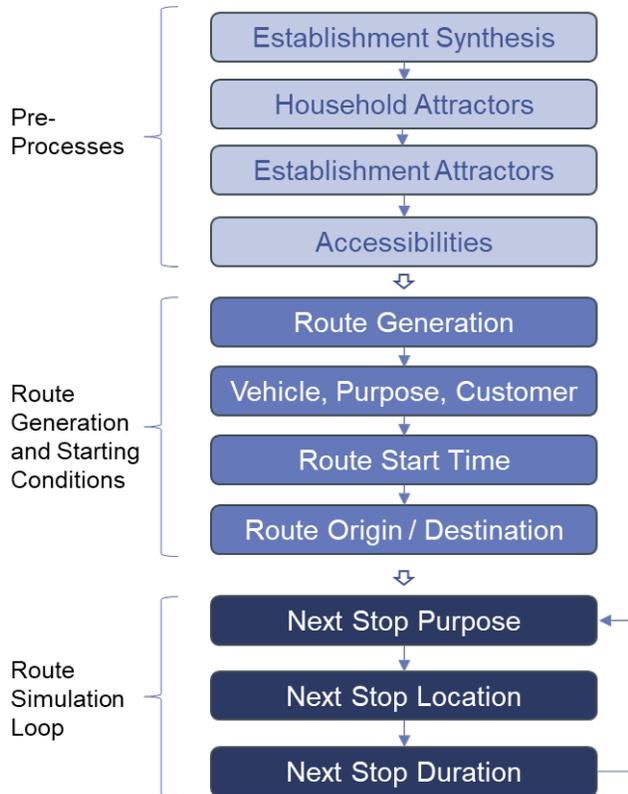
The new CVM differs from the previous version of the model in that it does not use the concept of complete tours, with all tours beginning and ending at establishment, as an organizing principal and constraint. The reasons for this are that commercial vehicle diary surveys revealed that for about one-third of vehicle days the vehicle did not return to the same place that it started, and the starting or ending locations for vehicle days were often not the establishment itself. Therefore, the decision was made to allow for “open jaw” tours whereby the simulation does not force a vehicle to begin and end the simulation day at the establishment, but rather models possible choices of non-establishment starting and ending locations. Open jaw tours are common for businesses that have multiple establishment locations, particularly those with separate warehouses and vehicle maintenance yards or garages.

The CVM represents establishments as individual simulation entities, which differs from the predecessor model’s generation and attraction of commercial vehicle trips from zone-level aggregate employment. This was an important design feature because, as the establishment surveys showed, there is a non-linear decreasing relationship between the number of employees at an establishment location and the number of commercial vehicle trips that location produces and attracts. Small businesses generate more trips per employee than larger businesses, a relationship that is lost when generating commercial vehicle trips from

aggregate zonal employment. Moreover, over time, trends in establishment sizes change, with recent smaller-size establishments becoming more prevalent in most parts of the U.S.

A flow diagram portraying the CVM system is shown in Figure 2.1-1. The CVM comprises three primary modeling stages: (1) pre-processing models which create inputs variables that are essential to the simulation; (2) generation of vehicle routes and their starting conditions; and (3) the dynamic simulation of stops and travel for each route.

Figure 2.1-1 CVM Components



2.1.1 Pre-Processes

The purpose of the pre-processing steps is to create scenario-specific variables that are important to how the CVM represents demand and its policy sensitivity.

Establishment Synthesis creates records for individual establishments, which form the basis for generating establishment trip attractions and for generating commercial vehicle routes in downstream models. Representation of establishments by industry and size enables the CVM to more accurately represent the non-linear decreasing rates of both trip attractions and route productions.

Household Attraction Generation links the CVM to the Resident model. This step generates commercial vehicle trip attractions to households, reading in the synthetic households and tours from the Resident model. These attractions include home deliveries of food and packages as well as various types of service visits. Household attractions are used in the creation of accessibility variables as a measure of residential customer demand and used as attractor variables in Stop Location Choice models where the customer type is residential.

Establishment Attraction Generation uses the synthetic establishments to generate commercial vehicle trip attractions to establishments. These attractions include goods deliveries and various types of service visits. Establishment attractions are used in the creation of accessibility variables as a measure of non-residential customer demand and are used as attractor variables in Stop Location Choice models where the customer type is non-residential.

Accessibility creates a set of variables to represent access to household attractions and access to establishment attractions. Accessibility variables are used in Route Generation and Stop Generation models to represent potential customer demand and affect the propensity of establishments to generate more vehicle routes or make additional stops on a route. Accessibility variables allow demand to vary throughout the region based on the density of households and businesses.

2.1.2 Route Generation and Starting Conditions

The purpose of the route starting conditions step is to generate the commercial vehicle routes to be simulated and to create parameters for those routes that represent contextual variables. There are two different sets of models, one for regular establishments and another for TNCs, which are not establishment-based.

Route Generation begins by generating some number of commercial vehicle routes for each establishment. Even within the same industry and firm, some establishment locations will generate daily commercial vehicle routes while others will not, which reflects different functions at each site. For example, corporate offices, which are unlikely to produce commercial vehicle goods and service trips, may be in a different location from production facilities, warehouses, or service facilities, which do produce commercial vehicle trips. Accordingly, the route generation model for establishments is a two-stage model. Stage 1 predicts whether an establishment will generate at least one commercial vehicle route, and Stage 2 predicts how many routes will be generated given at least one. The CVM software then creates separate simulation objects equal to the number of routes predicted for each establishment.

The route generation model for TNCs works differently. TNC drivers are effectively independent agents who work for online pickup and delivery services and are treated as their own establishments in the model. The demand for TNC pickup and delivery services is a function of the establishments which hire them through the online services, such as restaurants, grocery stores, and other businesses. For TNCs, the CVM generates a total number of TNC routes for an entire land use zone (LUZ), based on equations specified to estimate demand for pickups by the types of businesses that use them. As described below, the CVM then chooses a starting and ending location for each TNC route. The idea behind the choice of the LUZs was to generate TNC routes spatially as a function of the collective demand of businesses that use TNCs within the same general area. Although the model restricts the starting and ending locations of the TNC routes to the LUZ that generated them, the downstream model that chooses the location of pickup and delivery stops is not restricted to the starting LUZ.

Route Vehicle, Purpose, and Customer Type models are applied to each route generated in the simulation. The model is formulated as the multinomial joint choice across these three dimensions, which are closely correlated. Choice of vehicle type will depend on whether the route purpose involves goods pickup/delivery or service provision. For goods deliveries, if the customer type is a residential, a smaller or medium-size vehicle is more likely to be used, whereas multi-unit (tractor trailers) are seldom used to make deliveries to residences.

For TNCs, each route is assumed to use a light-duty vehicle type and the purpose to be for goods pickup/delivery, which represents 95 percent of the route observations in the TNC driver survey. TNC routes are also assumed to be eligible to serve a mix of residential and business customer type.

Route Start Time Choice models are applied to each CVM route, after the vehicle, purpose, and customer types have been chosen. The route starting time represents the departure time for the first trip on the route. The model is specified as an empirical distribution derived from the commercial vehicle survey and segmented by four groups of industries and by vehicle type, which were chosen for being statistically distinct from one another. For TNCs, distributions were derived based on the type of businesses being served by the TNC driver—restaurant, retail, or other.

Starting times are specified at 15-minute quarterly intervals within each hour of the day. The model is applied by drawing a starting time interval from the distribution appropriate to each combination of business and vehicle type. Note that the ending times for commercial vehicle routes are not pre-determined by a choice model, but rather determined by the sequence of downstream model outcomes—choices of next stop type, including to terminate the route; travel times between stops; and the time spent at each stop.

Route Origin and Termination Location Choice models are applied to choose the starting and ending location for each CVM route. The CVM allows for vehicles to begin or end their day at locations other than their home establishment. Separate models are used to choose locations for the origin of the route and for the location where the route will terminate. This is formulated as a two-level choice in which the first level is a location type choice, and the second level is the choice of zone (MGRA) given the type. The location type alternatives are:

- Establishment Location
- Other Warehouse/Distribution Center Land Use
- Other Residential Land Use
- Other Commercial/Public Land Use

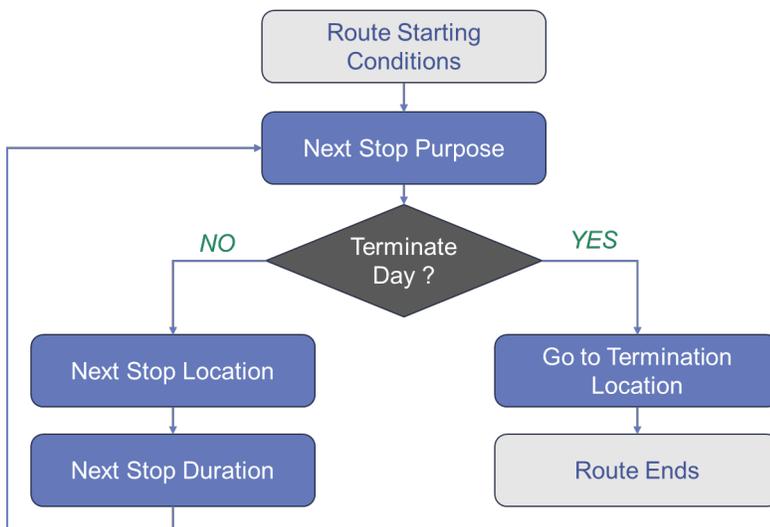
The choice of location type for the termination of the route considers the location type choice made for the origin of the route. If the establishment location is chosen, then the zone (MGRA) is given by the establishment. The non-establishment location types were specified to support more accurate representation of commercial vehicle parking locations at the start and end of each day. While a majority of vehicle routes will originate and terminate at the establishment location, other common location types are warehouse and distribution centers, most belonging to the same company; residential locations, which are commonly a driver's home; and other commercial or public land uses, such as maintenance facilities or customer sites. If the non-establishment location type has been chosen, the model will then choose an actual zone (MGRA), with the choice alternatives constrained by the availability of the particular land use type (warehousing, residential, other non-residential). The model also considers the various attraction variables such as employment or households in the MGRA as well as the distance from the establishment.

For TNCs, which are generated by LUZ, a single location is chosen to be both the origin of the route and the termination point of the route. This simple model which chooses a location within the LUZ based on household and employment attributes. In the TNC survey, some drivers recorded route origin or termination locations as "home." Others began or ended their recorded routes at non-home locations, which appear to be based on when they began and ended working for their particular TNC service.

2.1.3 Route Stop Simulation Loop

The route stop simulation models produce the actual trips made by commercial vehicles. This is dynamic model incrementally creates each stop on the vehicle route, given the starting conditions of vehicle, purpose, and customer type; starting time of day; and the starting and ending locations for the route. Figure 2.1-2 shows the relationship between the three principal steps in the simulation—the choices of the next stop purpose, location, and activity duration. As each stop is simulated, the CVM updates the location of the vehicle, the time of day, and the elapsed time on the route. These three factors are considered in each of the route simulation sub-models. The three models are run in the order shown until the “terminate route” purpose is chosen as the next stop purpose, which will create a trip to the pre-determined termination location and end the vehicle route.

Figure 2.1-2 CVM Route Simulation Process



Next Stop Purpose Choice decides what will happen next on each commercial vehicle route. This is the choice of one of seven stop purposes:

- **Goods Pickup** – a stop to pick up goods
- **Goods Delivery** – a stop to deliver goods to a customer
- **Service** – a stop to provide a professional service to a customer
- **Maintenance/Other** – a stop for either vehicle maintenance/refueling or driver breaks
- **Base Establishment** – a trip the establishment where the vehicle is based, without ending the route
- **Home** – a trip to the driver’s home, without ending the route
- **Terminate Route** – a trip to the final stop on the route where the vehicle will be parked, ending the route. This is the location chosen by the Route Termination Location Choice model describe above.

The next stop purpose model applies logical transition constraints, such that it is not possible to go from base to base or from home to home, since base and home each refer to a unique location. The model does not impose a limit on the number of stops that can be made; however, it is specified so that as the elapsed time on the tour increases, the probability of choosing the “terminate route” alternative increases. In addition, the terminate route alternative is more likely during certain time periods, irrespective of the starting time of the route, to reflect common business hours.

The next stop purpose model for TNCs is very similar to regular establishments, except that TNCs are assumed to not have the “service” alternative, with customer interactions limited to goods pickup and delivery.

Next Stop Location Choice determines the destination of the trip, where the next stop purpose will take place. Similar to the route origin and termination location choices, this is a two-level model. The first level is the choice of a land use type, and the second level is the choice of an actual zone (MGRA) given the land use type. The model is not applied if the next stop purpose is either “base” (establishment) or “terminate route” because those locations are already known. For other stop purposes, the location type alternatives are:

- Residential Land Use
- Warehouse/Distribution Center Land Use
- Port Facilities (Seaport, Airport, Intermodal Rail)
- Other Commercial/Public Land Use

These land use types were specified to enable more accurate representation of the demand attracted to each of these types through model calibration. The location type choice is a function of many factors, including the stop purpose, customer type, vehicle type, elapsed time, and time of day. For example, if the stop purpose is “home” or the customer type for a goods delivery or service stops is “residential,” then the residential land use type should be chosen. Other location land use types would be more likely for certain business types than others.

Given the choice of a location land use type, the CVM then chooses an actual zone (MGRA) from among the zones eligible for that land use type. The residential location type is expected to include only those zones with households; the warehouse/distribution center location type considers only zones with employment in transportation/warehousing; the port facilities type includes only zones with designated seaport, airport, or intermodal rail facilities; and the other commercial/public land use type must contain employment in any industry sector, other than transportation/warehousing.

The location choices model is specified to consider the household and establishment attractors in each zone, as predicted by those pre-processing models as well as other employment variables, depending on the stop purpose. The model also considers travel times and vehicle operating costs, including the travel time from the current location to alternative next stop locations and the travel time from the eventual terminal location to the alternative next stop locations, and the elapsed time on the route. As the elapsed time on the route increases, the model will be more likely to choose locations closer to the terminal location.

Next Stop Duration Simulation determines the simulated time spent at each location. For goods purposes, the stop duration would include time spent loading or unloading items and interacting with customers. For services, the duration represents the amount of time spent at the customer site providing the service, which

will vary widely depending on the business type, such as installation/repairs, landscaping, construction, health care, and cleaning, to name a few common service types. For maintenance/other purposes, the stop duration could involve vehicle refueling or repairs or it could involve driver meal breaks or other personal or company business. For the “home” purpose, duration is assumed to be personal time spent at the driver’s home.

For regular establishments, the duration is simulated using a model that considers the stop purpose, the vehicle type used, and whether the stop begins before or after 12 p.m., with stops later in the day expected to be more time constrained, particularly for service provision. For TNCs, the duration model considers the stop purpose and the land use density at the stop location, with increased density assumed to require longer stop times, as would be expected for deliveries to multi-family residential building and to office buildings.

3.0 CVM Component Design Details

This section provides additional details for the individual CVM components. For a better understanding of the flow between components, the reader is encouraged to refer to Section 2.1.

3.1 Pre-Processes

The CVM includes four pre-processing steps, the first of which, establishment synthesis is implemented in a standalone Python program, while the remaining three steps are integrated directly into the main CVM program in the ActivitySim program format.

3.1.1 Establishment Synthesis

The Establishment Synthesizer creates a table of establishment records. Each establishment has the attributes of a geographic location in an MGRA, an industry group, and the number of employees based at the site.

Method

The creation of establishments uses a mathematical programming approach to generate individual establishment records which meet two primary conditions:

1. The sum of employees for all establishments in an MGRA must equal the MGRA control totals for employment by industry group as forecast by SANDAG for the model scenario year.
2. The distribution of establishment sizes should come as close as possible to matching distribution of establishments by Land Use Zone (LUZ) for seven establishment size categories. The size distribution of establishments by LUZ was created by SANDAG for the base year of the model. In the absence of an alternative distribution for future years, the base-year distribution of establishment sizes is used by default.

As an outcome of satisfying these two conditions, the program also determines the total number of establishments for each industry group within a LUZ. The total number of establishments is not provided as an input.

Table 3.1-1 shows the industry group names and definitions used in the CVM and synthetic establishments.

Table 3.1-1 CVM Establishment Industry Groups

Code	Industry Group	Abbreviation	MGRA File Fields
1	Agriculture/Mining	AGM	emp_ag_min
2	Manufacturing	MFG	emp_mnf
3	Industrial/Utilities	IUT	emp_utl
4	Retail	RET	emp_ret
5	Wholesale	WHL	emp_whl
6	Construction	CON	emp_con
7	Transportation	TRN	emp_trn_wrh
8	Info/Finance/Insurance/Real Estate/Prof. Services	IFR	emp_fin_res_mgm, emp_bus_svcs
9	Education/Public/Other Services	EPO	emp_educ, emp_gov, emp_oth
10	Medical/Health Services	MHS	emp_hlth
11	Leisure/Accommodations and Food Services	LAF	emp_ent, emp_accm, emp_food
12	Military	MIL	emp_mil

The establishment size categories used in the CVM are shown in Table 3.1-2. These definitions were adapted from the size classes used by National Business Employment Dynamics Data and are consistent with the definitions used in the U.S. Census's County Business Patterns. Although these sources defined up to nine sizes classes, the Establishment Survey lacked observations in the two largest size classes; therefore, the CVM used the seventh class as the largest.

Table 3.1-2 Establishment Size Class Definitions

Code	Class Definition
1	1 to 4 employees
2	5 to 9 employees
3	10 to 19 employees
4	20 to 49 employees
5	50 to 99 employees
6	100 to 249 employees
7	250 to more employees

Inputs

The Establishment Synthesizer takes two sources of input data:

- Zonal land use file containing employment by industry sector in each zone. These employment values are industry-specific targets for the scenario year in question. The model will create and allocate establishments to each zone that, in sum, meet those targets. The Establishment Synthesizer was specified to synthesize establishments at the MGRA geographic level and used the file **mgra15_based_input2022.csv** for model development. Table 3.1-1, above, shows the MGRA employment fields in the rightmost column that were used for each industry group.
- LUZ data file containing baseline target percentage distributions of establishments by size and industry for each of 245 LUZs covering the SANDAG model area. The Establishment Synthesizer was specified to use the distributions created by SANDAG's land use and economics team who

created the input file **percent of establishments by luz-size-emp_cat.xlsx**. This file was prepared in August 2023 using the MGRA employment fields and size classes shown above in Table 3.1-1 and Table 3.1-2. An example excerpt from this file for the first seven LUZs for the Manufacturing sector is shown in Table 3.1-3. For the sake of simplicity in model development, it was assumed that these baseline distributions would carry forward to future years; however, it is possible to develop and use a LUZ data file for any scenario year if there is information suggesting that the distribution should be different from the baseline.

Table 3.1-3 Example of LUZ Target Distributions

Sector:	emp_mnf						
Size Class	1	2	3	4	5	6	7
LUZ	emp_mnf1	emp_mnf2	emp_mnf3	emp_mnf4	emp_mnf5	emp_mnf6	emp_mnf7
1	58%	25%	8%	0%	8%	0%	0%
2	86%	0%	0%	14%	0%	0%	0%
3	75%	0%	0%	0%	25%	0%	0%
4	47%	21%	16%	11%	5%	0%	0%
5	39%	39%	11%	6%	6%	0%	0%
6	90%	10%	0%	0%	0%	0%	0%
7	70%	10%	20%	0%	0%	0%	0%

The Establishment Synthesizer loops through each of the 17 MGRA employment types shown in the rightmost column of Table 3.1-2. For each employment category, it loops through each of the 245 LUZs. In effect, the model tries to solve a separate optimization problem for each employment type and LUZ, skipping LUZs where there is no employment of a particular type.

Outputs

The outputs of the Establishment Synthesizer are a table of establishment records. Table 3.1-4 is an example of the information these records contain for the first LUZ for the Manufacturing sector. The synthetic establishments are used directly in two downstream processes—Establishment Attractor Generation and Route Generation.

Table 3.1-4 Example of Synthetic Establishment Records

Establishment_ID	Industry_No	Industry_Name	LUZ	MGRA	Employees	Size_Class
87191	2	MFG	1	533	1	1
87192	2	MFG	1	3576	2	1
87193	2	MFG	1	3576	2	1
87194	2	MFG	1	3748	3	1
87195	2	MFG	1	5286	13	3
87196	2	MFG	1	5581	6	2
87197	2	MFG	1	6482	99	5
87198	2	MFG	1	7610	2	1
87199	2	MFG	1	7628	2	1
87200	2	MFG	1	8346	4	1

3.1.2 Household Attraction Generation

The Household Attraction Generation model is applied to the synthetic households in ABM3 to predict whether each household will attract a commercial vehicle trips of three types:

- Package delivery
- Food delivery (grocery or meals)
- Service stops

The outcomes of these predictions are then summed across all the households in a land use zone (MGRA) to create household attraction variables for that zone. This creates zone-level attributes for each of the three attraction types.

Method

The dependent variables in the model were based on responses to questions in the 2022 SANDAG household travel survey, which asked whether (not how many) commercial vehicle trips were made to the household for each of the three types on the survey day.

A separate model was estimated for each of the three attraction types. The form of the model is a binary logit model as shown in the equation below. The probability of attraction type A for household n is a function of estimated parameters β and household attributes X .

$$Prob_n(A) = \frac{1}{1 + e^{-\beta X_n}}$$

Specification testing of the models considered several household attributes that might influence their propensity to attract commercial vehicle trips of each type as well as the potential influence of the number of tours generated by the household for dining, shopping, and other purposes. Details of the final model specification may be found in the Model Estimation and Calibration technical report.

Inputs

Inputs to the Household Attraction Generation model are the synthetic households from ABM3 and the number of tours made by these households in the resident model simulation, which is assumed to run prior to the CVM.

Outputs

The outputs of the Household Attraction Generation model are predictions for individual households and MGRA-level summaries of attractions. Individual household records are updated with three new columns in a CVM output file, **final_households.csv**, a sample of which is shown in Table 3.1-5.

Table 3.1-5 Household Attractions Variables Added to Household Records

household_id	home_zone_id	has_attraction_food	has_attraction_package	has_attraction_service
30	100	FALSE	TRUE	FALSE
31	100	FALSE	TRUE	FALSE
32	100	FALSE	FALSE	FALSE
33	100	FALSE	TRUE	FALSE
34	100	FALSE	TRUE	FALSE
35	100	FALSE	FALSE	FALSE
36	100	FALSE	FALSE	FALSE
37	100	TRUE	FALSE	TRUE
38	100	FALSE	TRUE	FALSE
39	100	FALSE	FALSE	FALSE

MGRA land use records are updated with the three new columns in the CVM output file, **final_land_use.csv**, as shown in the example in Table 3.1-6. The MGRA summaries of attractions are used as input variables to two subsequent modeling steps—Accessibility Variables and Stop Location Choice.

Table 3.1-6 Household Attraction Variables Added to Land Use Records

mgra	num_hh_food_delivery	num_hh_package_delivery	num_hh_service
1	8	52	12
2	0	9	2
3	6	70	6
4	0	0	0
5	2	23	2
6	1	17	6
7	1	20	4
8	1	12	5
9	12	135	26
10	12	111	22

3.1.3 Establishment Attraction Generation

The Establishment Attraction Generation model is applied to the synthetic establishments created by the Establishment Synthesizer to predict whether and how many commercial vehicle trips will be attracted to each establishment for either goods pickup or delivery or service stops. The outcomes of these predictions are then summed across all the establishments in a land use zone (MGRA) to create establishment attraction variables for that zone, specific to each of the twelve CVM industry groups (Table 3.1-1).

Method

The model consists of two prediction steps:

- Step 1: Probability of at least one stop (binary logit)

- Step 2: Given at least one stop, predict the number of attractions per employee (linear regression)

The dependent variables in the model were based on responses to questions in the 2022 SANDAG commercial vehicle survey which asked how many inbound commercial vehicle trips were made to an establishment on an average day for either goods or services. Notably, a substantial proportion of establishment reported zero commercial vehicle visits on an average day, which led to the two-stage model.

The form of the first model is a binary logit model as shown in the equation below. The probability of at least one stop S for establishment of n in industry group g is a function of estimated parameters β and the number of employees X .

$$Prob_n(S > 0) = \frac{1}{1 + e^{-\beta_g X_n}}$$

The model was segmented into four attractions model groups, based on initial specification testing which revealed that some industry groups had similar commercial vehicle attraction rates per employee and were indistinguishable for purposes of statistical estimation. The four groupings are as shown in Table 3.1-7.

Table 3.1-7 Establishment Attractions Model Group Segmentation

Code	Industry Group	Abbreviation	Attractions Model Group
2	Manufacturing	MFG	1
4	Retail	RET	1
5	Wholesale	WHL	1
3	Industrial/Utilities	IUT	2
11	Leisure/Accommodations and Food Services	LAF	2
1	Agriculture/Mining	AGM	3
6	Construction	CON	3
7	Transportation	TRN	3
9	Education/Public/Other Services	EPO	3
10	Medical/Health Services	MHS	3
8	Info/Finance/Insurance/Real Estate/Prof. Services	IFR	4

Note that Military establishments were not included in the commercial vehicle survey, and attraction models were not specified to be sensitive to military employment. The decision not to simulate common commercial vehicle trips to military bases was based partly on the lack of supporting data and partly on the nature of their operations, which tend to be self-contained. Areas in and around military bases can attract general purpose commercial vehicle trips in the CVM to the extent that non-military employment types or housing are present.

The form of the second model is a linear regression model which was specified to predict the number of commercial vehicle attractions y to an establishment n , given there is at least one, as a function of its employment X :

$$y_n = F(\beta_g, X_n)$$

The model is segmented by all eleven of the industry groups found in Table 3.1-7. The role of the attractions generated in the CVM is to attract trips but not to force a specific integer number of trips. Accordingly, the establishment attraction generation model was specified to allow for fractional attractions per establishment.

Specification testing of more complex limited-dependent variables models, including count models such as Poisson regression, revealed that these models predicted distributions of outcomes that did not fit the observed data as well. In addition, various non-linear forms of the employment variable were tested and a variable of the form square-root of employment was ultimately adopted. Details of the final model specification may be found in the Model Estimation and Calibration technical report.

Inputs

Inputs to the Establishment Attraction Generation model are the synthetic establishments generated by the Establishment Synthesizer. The key attributes used to predict attractions are the establishment’s industry group and number of employees (size of the establishment).

Outputs

The Establishment Attraction Generation model generates predictions which may be found in two CVM output files. The CVM produces an output file, **final_establishments.csv**, to which it will add fields indicating whether a synthetic establishment will attract commercial vehicle trips and how many. Example entries for these fields are shown in Table 3.1-8.

Table 3.1-8 Establishment Attraction Variables Added to Establishment Records

establishment_id	industry_name	zone_id	employees	size_class	has_attraction	attractions
1	EPO	6716	6	2	TRUE	2.3
2	EPO	6716	11	3	TRUE	3.1
3	EPO	6716	11	3	TRUE	3.1
4	EPO	6716	13	3	FALSE	0.0
5	EPO	6716	25	4	TRUE	4.7
6	EPO	6716	25	4	TRUE	4.7
7	EPO	6716	25	4	TRUE	4.7
8	EPO	6716	26	4	TRUE	4.8

MGRA land use records are updated with the new attraction columns in the CVM output file, **final_land_use.csv**. There is a separate column for each CVM industry group as shown in the example in Table 3.1-9. The MGRA summaries of attractions are used as input variables to two subsequent modeling steps—Accessibility Variables and Stop Location Choice.

Table 3.1-9 Establishment Attraction Variables Added to Land Use Records

mgra	Total	AGM	CON	EPO	IFR	IUT	MFG	MHS	MIL	RET	TRN	WHL
1	11.2	0	0	0	2.3	0	0	9.0	0	0	0	0
2	21.4	0	7.0	9.5	0	0	0	2.2	0	0	0	0
3	131.9	0	0	21.1	0	0	0	106.2	0	2.5	1.1	0
4	66.3	0	4.5	4.0	11.7	0	0	19.8	0	0	0	0
5	0.9	0	0	0	0	0	0	0	0	0	0	0
6	12.5	0	1.2	0	9.1	0	0	0	0	0	0	0

mgra	Total	AGM	CON	EPO	IFR	IUT	MFG	MHS	MIL	RET	TRN	WHL
7	1.0	0	0	0	1.0	0	0	0	0	0	0	0
8	2.7	0	0	0	0	0	1.5	0	0	0	0	0
9	1.5	0	0	0	0	0	1.5	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0

3.1.4 Accessibility

The Accessibility model creates accessibility variables representing the access to household and establishment attractors weighted by a continuous function of travel time impedance. These variables are used in the downstream models for commercial vehicle route generation and stop purpose generation, allowing commercial vehicle demand to vary partially as a result of potential customer interactions, which are expected to vary across the region based on household and employment density. For forecasting growth scenarios, it is expected that accessibility will increase as a function more households and business, resulting in greater demand, but this should be offset to some degree by congested travel times.

Method

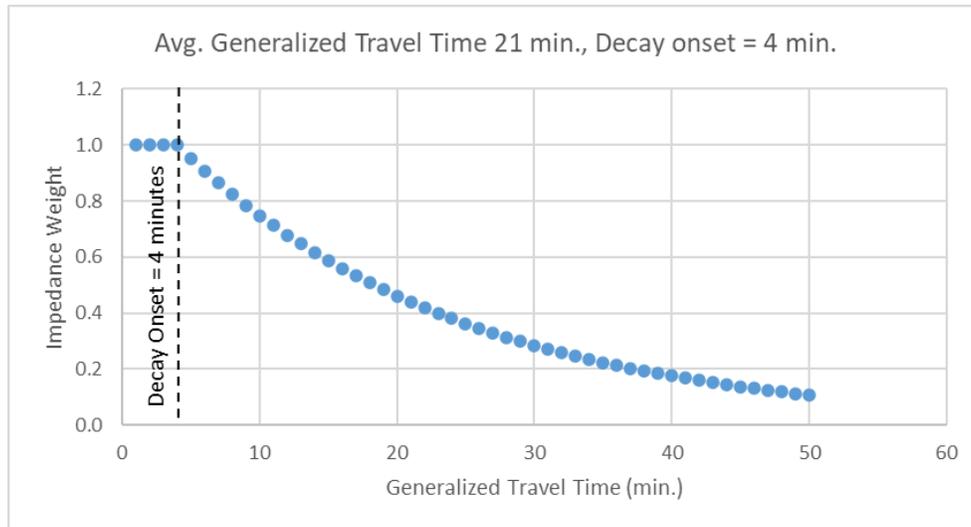
Accessibility variables are generated specific to each TAZ and assumed to apply uniformly to all MGRAs within each TAZ. The Accessibility model takes the following form, composed of two parts, attractors and impedance, similar to a gravity model formulation:

$$Accessibility_i^k = \ln \left(\sum_{\forall j \in J} Attractors_j^k \times \exp(\theta_k * (\max(0, GenTravelTime_{ij} - DecayOnsetTime_k))) \right)$$

The accessibility of zone *i* to attractors of type *k* is measured by the sum of the attractors in all zones *j*, weighted by the impedance of generalized travel time between zone *i* and each zone *j*. The entire righthand side of the expression is enclosed in a natural log function for consistency with the composite utility variables used in discrete choice models.

The above formula includes a decay onset time, which assumes that drivers are indifferent to travel times less than a certain amount (e.g., 4 to 6 minutes). θ_k is a parameter used to control the rate of decay in attractiveness as the generalize travel time increases, which could be specific to each attractor type. The value of θ_k is estimated based on the assumption of a negative exponential distribution and the weighted mean generalized travel time observed for individual trips in the 2022 Commercial Vehicle Survey. Figure 3.1-1 is an illustrative example of what the impedance curve looks like. Calculation details may be found in the Model Estimation and Calibration technical report.

Figure 3.1-1 Accessibility Impedance Curve Example



The generalized travel time is represented by the congested travel time plus the cost of tolls. Toll costs are converted into equivalent in-vehicle minutes using values of time specific to commercial vehicle types. The values of time used in the Accessibility model should be consistent with those used in the ABM3 network assignment for specific vehicle types.

Inputs

The inputs to the Accessibility model are attraction variables and travel time and toll cost skims. The household and establishment attraction variables created in the above steps and found in the augmented land use file as shown above in Table 3.1-6 for household attractors and

Table 3.1-9 Establishment Attraction Variables Added to Land Use Records

for establishment attractors.

The primary use of the accessibility variables is to influence route generation, which is the next step in the sequence of models. At the point at which commercial vehicle routes are generated, the type of vehicle and the times of day covered by the route are unknown. Therefore, the Accessibility model makes the following simplifying assumptions about travel time and cost skims:

- Use midday (MD) travel time and toll skims, which are likely to cover most route operating hours;
- Skimmed travel times and costs should use a weighted average of light-, medium-, and heavy-duty vehicle types (LHDT, MHDT, HHDT). Weights are based on the proportion of commercial vehicle routes using each type as found in the 2022 Commercial Vehicle Survey as follows:

$$\begin{aligned}
 AccessTravelTime = & GenTravelTime_{LHDT} * Proportion_{LCV} \\
 & + GenTravelTime_{MHDT} * Proportion_{SUT} \\
 & + GenTravelTime_{HHDT} * Proportion_{MUT}
 \end{aligned}$$

The proportional use of vehicles by type and the corresponding assumed values of time are shown in Table 3.1-10. The values of time were specified to be the consistent with those using in the ABM3 network assignment procedures. LCVs used the value of time for light-heavy duty trucks (LHDT), SUTs used the value of time for medium-heavy duty trucks (MHDT), MUTs used the value of time for heavy-heavy duty trucks (HHDT), and TNC passenger vehicles used the value of time for high-income single-occupancy vehicle passenger cars. For TNCs, the skim market segments are assumed to use auto single-occupancy vehicle (SOV) and the value of time corresponding to the “high” income group.

Table 3.1-10 Accessibility Proportional Usage and Value of Time by Vehicle Type

Vehicle Type	Proportion Used	Value of Time (\$/hr.)
Establishments		
Light Commercial Vehicle (LCV)	54.4%	\$67.00
Single Unit Truck (SUT)	34.1%	\$68.00
Multi-Unit Truck (MUT)	11.5%	\$89.00
TNCs		
Passenger Vehicle	100.0%	\$85.00

Outputs

The Accessibility model calculates new variables which are created by the CVM in a separate file, **final_commercial_accessibility.csv**. These variables are then available for use in downstream models, such as route generation and stop purpose choice. Table 3.1-11 shows a sample of outputs for regular Establishments. The industry groups 1 – 4 shown in this table were specified to align with the segmentation by industry group of the route generation model described below (see Section 3.2.1). Table 3.1-12 shows accessibility variables for TNCs, which were developed separately.

Table 3.1-11 Establishment Accessibility Variables

zone_id	Ind. Group 1	Ind. Group 2	Ind. Group 3	Ind. Group 4	HH Food	HH Packg.	HH Service
1	10.0	9.4	10.8	10.0	10.1	12.3	10.5
2	10.1	9.6	10.9	10.3	10.2	12.3	10.6
3	9.8	9.1	10.6	9.8	9.9	12.1	10.3
4	9.8	9.1	10.4	9.9	9.8	11.9	10.3
5	9.8	9.2	10.5	10.0	9.8	12.0	10.3
6	9.8	9.1	10.4	9.9	9.7	11.9	10.3
7	9.9	9.4	10.7	10.1	10.0	12.1	10.4
8	9.9	9.4	10.7	10.1	10.0	12.2	10.5

Table 3.1-12 TNC Accessibility Variables

zone_id	Establishments All	HH Food	HH Package
1	11.6	10.1	12.3
2	11.7	10.2	12.4
3	11.4	10.0	12.1
4	11.4	9.8	12.0
5	11.4	9.9	12.1
6	11.3	9.8	12.0
7	11.5	10.0	12.2
8	11.6	10.1	12.2

3.2 Route Generation and Starting Conditions

The set of models described in this section generate commercial vehicle routes; determine the purpose, customer type, and vehicle type for each route; set a starting time for the route; and choose the starting and ending locations of the route, which may or may not be the location of the establishment itself.

For each of these model components, a separate, parallel set of models covers TNCs, which operate as independent driver agents. TNC routes are not linked to one particular commercial establishment but are generated as an outcome of expected demand from potential customers.

3.2.1 Route Generation -- Establishments

The Establishment Route Generation model is applied to the synthetic establishments created by the Establishment Synthesizer to predict whether and how many commercial vehicle routes the establishment will generate for a simulation day.

Method

The model consists of two prediction steps:

- Step 1: Probability of at least one route (binary logit)
- Step 2: Given at least one route, predict the number of routes per employee (linear regression)

The dependent variables in the model were based on responses to questions in the 2022 SANDAG Commercial Vehicle Survey which asked how many outbound commercial vehicle trips were made from an establishment on an average day for either goods or services. As with trip attractions, a substantial proportion of establishment reported zero commercial vehicle routes on an average day, which led to the two-stage model. In addition, the number of routes reported in driver travel diaries can be used to develop the model; however, about half of the establishments in the survey did not provide driver diaries, which does not mean that they did not generate commercial vehicle routes. Thus, the “how many...” question which was reported by all establishments was viewed as a more reliable indicator of which establishments produced commercial vehicle trips.

The form of the first model is a binary logit model as shown in the equation below. The probability of at least one route R for establishment of n in industry group g is a function of estimated parameters β and the number of employees X .

$$Prob_n(R > 0) = \frac{1}{1 + e^{-\beta_g X_n}}$$

The model was segmented into four route generation model groups, based on initial specification testing which revealed that some industry groups had similar route generation rates per employee and were indistinguishable for purposes of statistical estimation. The four groupings, which are different from the groupings used in the establishment attractions model, are shown in Table 3.2-1.

Table 3.2-1 Establishment Route Generation Model Group Segmentation

Code	Industry Group	Abbreviation	Route Gen. Model Group
1	Agriculture/Mining	AGM	1
2	Manufacturing	MFG	1
5	Wholesale	WHL	1
3	Industrial/Utilities	IUT	2
4	Retail	RET	2
6	Construction	CON	2
7	Transportation	TRN	2
8	Info/Finance/Insurance/Real Estate/Prof. Services	IFR	3
11	Leisure/Accommodations and Food Services	LAF	3
9	Education/Public/Other Services	EPO	4
10	Medical/Health Services	MHS	4

The form of the second model is a linear regression model which was specified to predict the number of commercial vehicle routes y generated by an establishment n , given there is at least one, as a function of its employment X and accessibility variables A :

$$y_n = F(\beta_g, X_n, A_n)$$

The model is segmented by all eleven of the industry groups found in Table 3.2-1. The CVM creates a route equal to the number of routes predicted for each establishment, rounded to the nearest integer. Specification testing of more complex limited-dependent variables models, including count models such as Poisson regression, revealed that these models predicted distributions of outcomes that did not fit the observed data as well. Alternative non-linear forms of the employment variable were tested and a variable of the form square-root of employment was ultimately adopted. Details of the final model specification may be found in the Model Estimation and Calibration technical report.

Inputs

Inputs to the Establishment Route Generation model are the synthetic establishments generated by the Establishment Synthesizer and the Accessibility variables. For purposes of route generation is assumed that the establishment inherits the accessibility variables for its MGRA. The key attributes used to predict attractions are the establishment’s industry group, number of employees, and accessibility to both establishment and household attractions.

Outputs

The outcomes of the Establishment Route Generation model are route simulation objects, captured in table form in the output file **final_routes.csv**. Subsequent modeling steps, described below, then add attributes of route purpose, customer type, vehicle, type, starting time, origin location, and termination location. An example listing the complete list of route attributes is discussed and shown below in Table 3.2-5.

3.2.2 Route Generation -- TNCs

The CVM generates a total number of TNC routes for an entire LUZ, based on equations specified to estimate demand for pickups by the types of businesses that use them. TNC drivers are effectively independent agents who work for online pickup and delivery services and are treated as their own establishments in the model. The demand for TNC pickup and delivery services is a function of the establishments which hire them through the online services, such as restaurants, grocery stores, and other businesses. The choice of LUZ as a spatial organizing device was to simulate operating territories for TNC drivers whose home bases are not consistently reported or easy to derive from the TNC survey.

Method

Estimating the total demand for TNC usage for commercial pickup and delivery services was challenging because the TNC portion of the Commercial Vehicle Survey did not weight the sample due to lack of data on the total number of TNC drivers working for commercial delivery services in the San Diego region at the time of the survey.

To overcome this limitation, information from the Establishment part of the survey was used. Each establishment in the survey was asked if they used TNCs for delivering goods and how many outbound cargo trips they made on an average day. In addition, each establishment was assigned a sampling weight based on the survey processing recommendation to use estimates of total establishments by type in the region¹. From this information it was possible to attribute a weighted number of TNC pickups for each establishment and to summarize these pickups by the establishment's industry group for the entire region.

In the TNC survey, respondents were asked to identify the service they were representing (e.g., Instacart, Uber Eats, Amazon Flex, etc. As part of the CVM data cleaning process, these services were recoded into three categories:

- Restaurant
- Retail
- Non-Restaurant or Retail

Each establishment in the establishment survey which reported using TNCs was also classified into these same categories. Assuming that each TNC pickup episode at a single establishment represents a unique TNC driver/route, expansion weights were calculated as shown in Table 3.2-2.

¹ SANDAG Data Expansion Tech Memo Feb 22 2023.docx

Table 3.2-2 Calculation of TNC Route Expansion Weights

TNC Customer Type	TNC Client Pickups	Establishment Weighted Deliveries using TNCs	TNC Expansion Weights
Restaurant	636	8,184	12.868
Retail	197	3,766	19.117
Non Restaurant or Retail	152	575	3.783

These TNC expansion weights were then applied to the route and trip records in the TNC portion of the Commercial Vehicle Survey to produce a sum of weighted routes for entire model region for each TNC customer type. Weighting the TNC survey produced an estimate of 5,019 TNC routes and 60,932 TNC trips per day across all customer types.

The next step in developing a route generation model for TNCs was to create a set of route generation rates based on the total employment in the region for each of the TNC customer types. Employment figures were taken from the 2022 MGRA land use file. The results of these calculations are shown in Table 3.2-3.

Table 3.2-3 Calculation of TNC Route Generation Rates

TNC Customer Type	TNC Weighted Routes	Establishment Employment	TNC Routes Generated Per Job
Restaurant	2,844	138,803	0.0204895
Retail	1,816	143,864	0.0126230
Non Restaurant or Retail²	359	1,282,601	0.0002799

Inputs

The main input to the TNC Route Generation model is the MGRA land use file. The TNC route generation rates shown in the rightmost column of Table 3.2-3 are applied to the total employment for each of the three customer types at the LUZ.

Outputs

The TNC Route Generation model produces an expected number of TNC routes for each of the three customer types for each LUZ. As with regular establishments, the CVM creates TNC route simulation objects, which appear as table records in the output file **final_routes.csv**. Rather than industry group codes,

² Non-restaurant or retail employment includes all employment types in the MGRA file other than food (restaurant) and retail. Also not included are military employment and the non-wage-and salary employment categories.

TNC routes are labeled by the three customer types. Subsequent modeling steps, described below, add attributes of route purpose, customer type, vehicle, type, starting time, origin location, and termination location for each route. An example listing the complete list of route attributes is discussed and shown below in Table 3.2-5.

3.2.3 Route Vehicle, Purpose, and Customer Type Allocation

The Route Vehicle, Purpose, and Customer Type Allocation (VPC) model assigns each of these attributes to each route generated by the Route Generation model. This model was specified as a multinomial joint choice across these three dimensions recognizing that they are interrelated. The choice dimensions for the model are as follows:

- **Vehicle Type:**
 - Light Commercial Vehicle (LCV)
 - Single-Unit Truck (SUT)
 - Multi-Unit Truck (MUT)
- **Purpose Type:**
 - Goods: customer-oriented; may include Services and other stop types
 - Service: customer-oriented; does not include Goods stops; may include other stop types
 - Maintenance/Other: refueling, driver breaks, vehicle/equipment repositioning, buying supplies, other; does not include Goods or Services stops; not customer oriented
- **Customer Type:**
 - Residential Only: households, including multi-family buildings (Goods or Services purposes)
 - Non-residential Only: commercial, public/government (Goods or Services purposes)
 - Mixed Residential and Non-residential: (Goods or Services purposes)
 - No Customer: (Maintenance/Other purpose)

Note that there is not separate VPC model for TNCs. Due to the homogeneity of TNC usage found the Commercial Vehicle Survey, it was assumed that all TNC routes would be automatically assigned the LCV vehicle type, the Goods purpose, and a Mixed Residential and Non-residential customer type.

Method

The VPC model is formulated as a multinomial logit model in which the choice alternatives represent the 21 feasible combinations of vehicle, purpose, and customer type as defined above. The model is formulated as a joint choice in which all three dimensions are represented at the same level (no nesting). As expressed in the equation below, the probability of a particular combination of vehicle, purpose, and customer type (*vpc*) for establishment *n* is a function of estimated parameters β related to specific dimensions (either vehicle type, purpose, or customer type) and establishment attributes *X*, such as industry groups.

$$Prob_n(vpc) = \frac{e^{\beta_k X_n}}{\sum_{vpc} e^{\beta_k X_n}}$$

Details of the final model specification may be found in the Model Estimation and Calibration technical report.

Inputs

The primary inputs to the VPC model are the routes generated in the previous step, which should include any relevant establishment attributes.

Outputs

The VPC adds three new attributes to each simulated route—vehicle type, route purpose, and customer type. These results may be viewed in the final outputs in the file, **final_routes.csv**. An example listing the complete list of route attributes is discussed and shown below in Table 3.2-5.

3.2.4 Route Start Time Choice

The CVM simulates each stop on a commercial vehicle route dynamically, beginning with a starting clock time for each route. The route starting time represents the departure time for the first trip on the route and becomes an attribute of the route.

Method

The method used to choose start times for each route is to draw times from a smoothed empirical distribution of route starting times using a weighted distribution from the Commercial Vehicle Survey. Intervals of 15 minutes were chosen as the alternative time choice units, which were judged to provide sufficient resolution for the simulation without stretching the observed data too thin. Based on a careful study of the data which considered alternative ways of grouping routes, the time of day choices for establishments were segmented based on similarities between their distributions and maximizing the differences between groups. For routes with goods or service as the primary purpose, routes were segmented into business groups, and some further segmented by vehicle types. There is a separate group for routes with maintenance/other purposes Table 3.2-4 lists the business group definitions used for the segmentation.

Table 3.2-4 Route Starting Time Choice Business Group Definitions

Business Groups	Business Types included
Business Group 1	Agriculture / Mining Industrial / Utilities
Business Group 2	Construction Education / Other Public Services Info / Finance / Insurance / Real Estate Leisure / Accommodations & Food Medical / Health Services Retail
Business Group 3	Manufacturing Transportation Wholesale

The following segments were used for the final model:

- BusGrp1
- BusGrp2 x LCV
- BusGrp2 x MUT
- BusGrp2 x SUT
- BusGrp3 x LCV
- BusGrp3 x MUT
- BusGrp3 x SUT
- Maintenance/Other Purposes

The route starting time distributions for these segments are plotted below in Figure 3.2-1, Figure 3.2-2, and Figure 3.2-3. For TNCs, time of day choices were segmented by the three customer types—restaurant, retail, and non-restaurant or retail as shown in Figure 3.2-4. Notably, routes involving heavy trucks (MUT) tend to start earlier in the day; maintenance/other purposes start later. TNCs used for non-restaurant or retail customers tend to start earlier than those used for restaurants or retail deliveries.

Figure 3.2-1 Route Starting Time Distributions: Bus. Grp. 1 - Agriculture and Utilities Goods/Services Purpose and Maintenance/Other Purpose

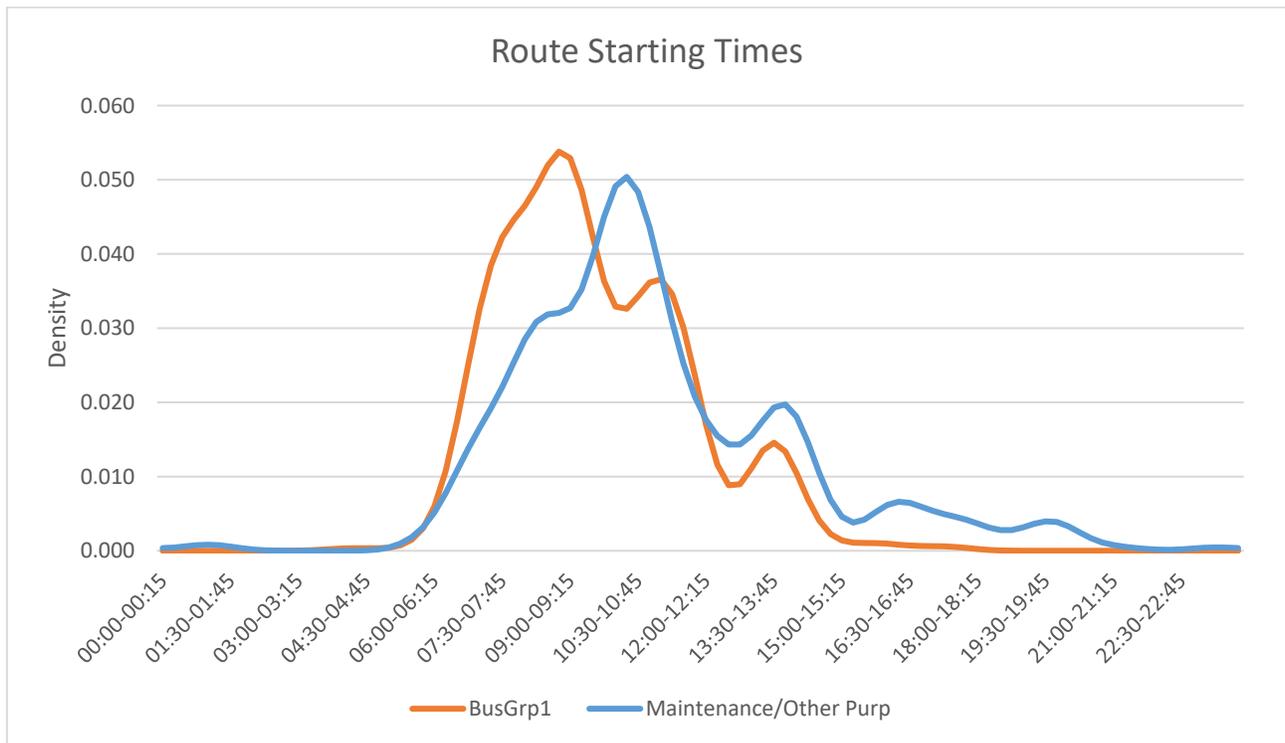


Figure 3.2-2 Route Starting Time Distributions by Vehicle Type: Bus. Grp. 2 - Retail and Various Service Industries Goods/Service Purpose

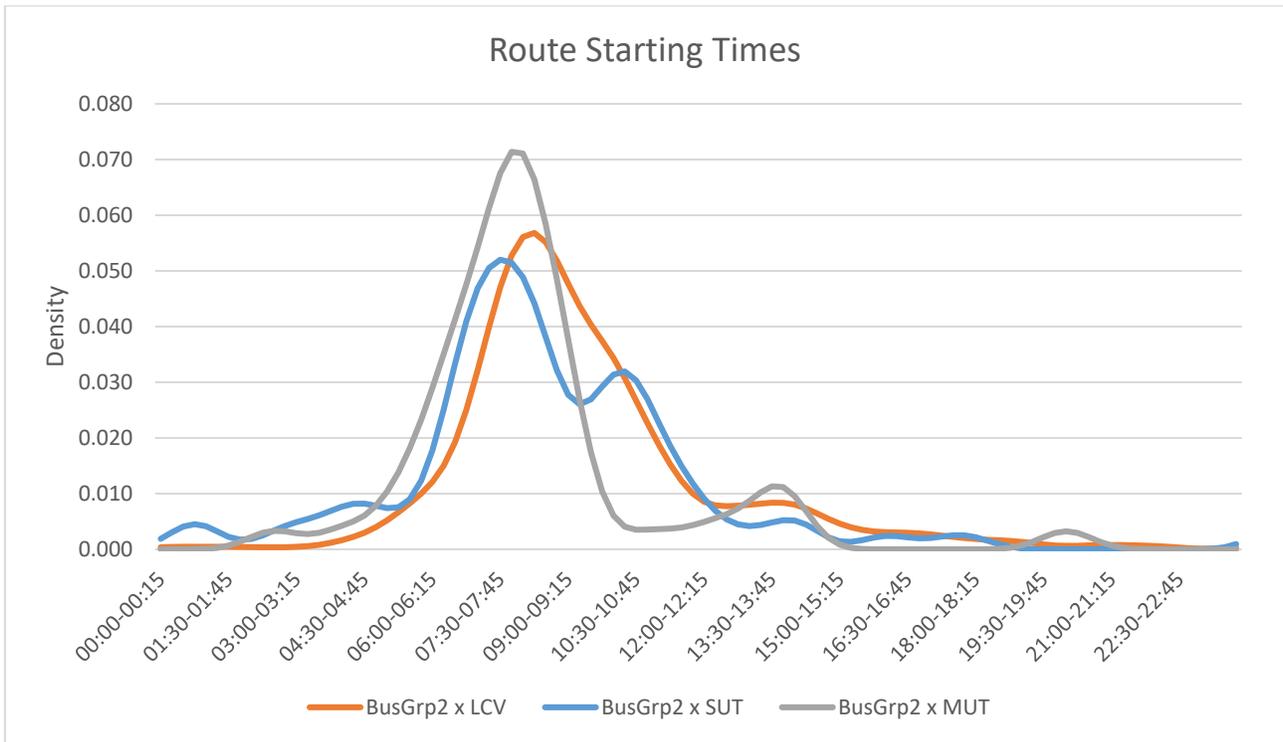


Figure 3.2-3 Route Starting Time Distributions by Vehicle Type: Bus. Grp. 3 - Manufacturing, Wholesale & Transportation Goods/Service Purpose

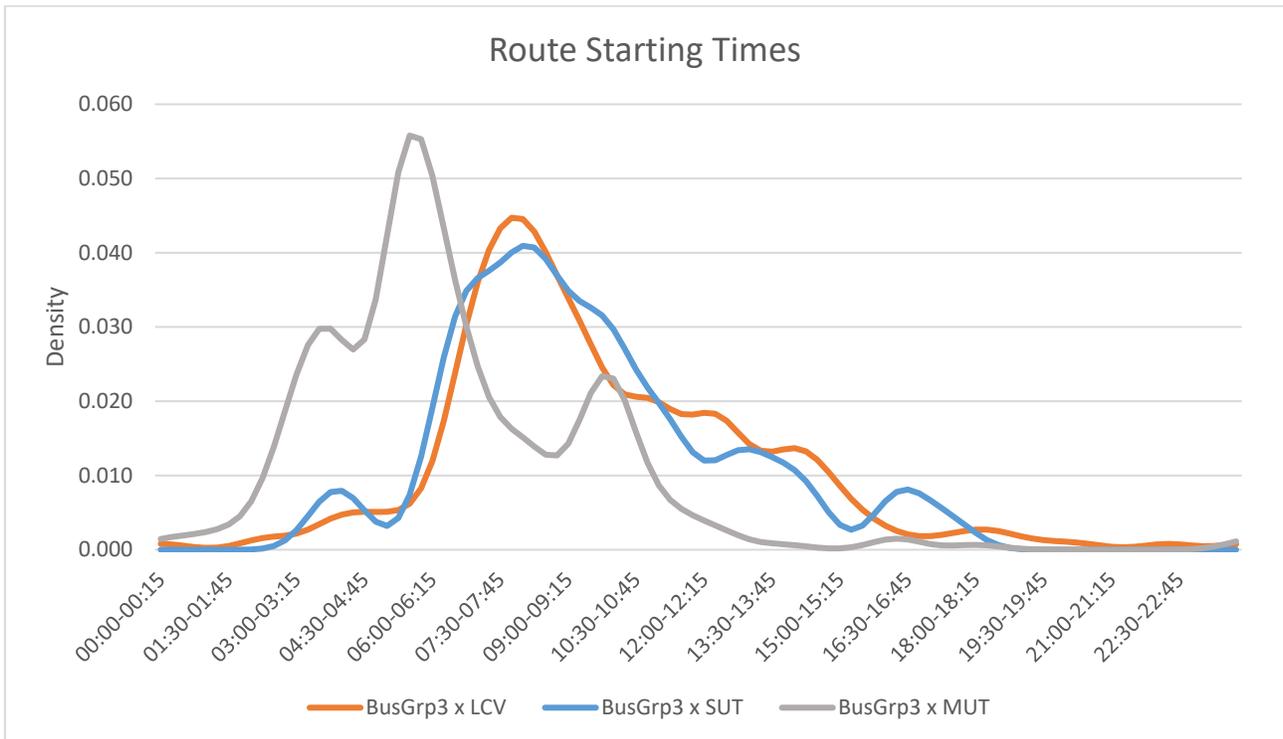
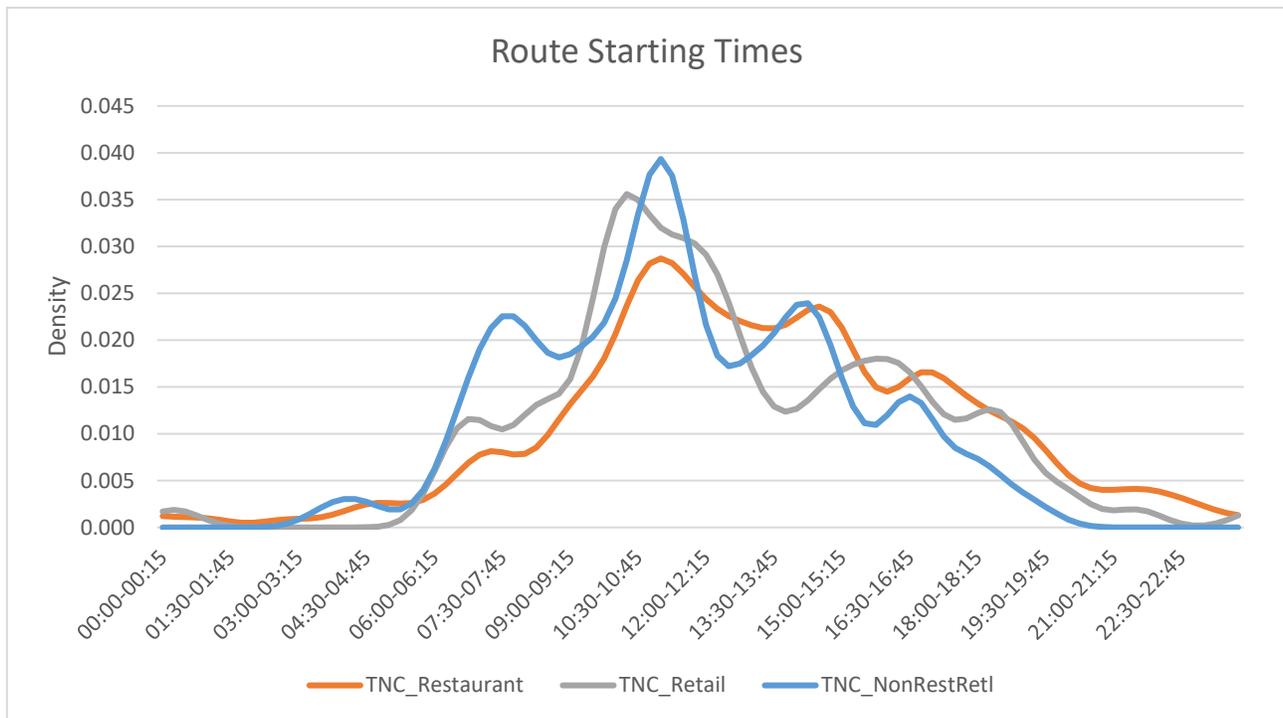


Figure 3.2-4 TNC Route Starting Time Distributions by Customer Type



Inputs

The primary input to the Route Start Time Choice model are the routes generated in the previous step, which should include any relevant establishment attributes as well as the route’s vehicle type, purpose, and customer type, as determined from the VPC model step.

Outputs

The Route Start Time Choice adds one new attributes to each simulated route, the starting time of day. These results may be viewed in the final outputs in the file, **final_routes.csv**. An example listing the complete list of route attributes is discussed and shown below in Table 3.2-5.

3.2.5 Route Origin and Termination Location Choices

The Route Origin and Termination Location Choice step consists of a set of complementary models, which together determine the starting and ending locations of commercial vehicle routes. This model applies to regular establishments. A different model structure is used for TNC routes since they do not have an establishment as a base.

Commercial vehicle routes represent an entire day’s worth of travel for a commercial vehicle. While the establishment location may be the origin and termination point for a majority of routes, a substantial portion of the routes in the Commercial Vehicle Survey began or ended their day at a different location. Using models to predict the starting and ending locations of routes permits the flexibility to represent these variations.

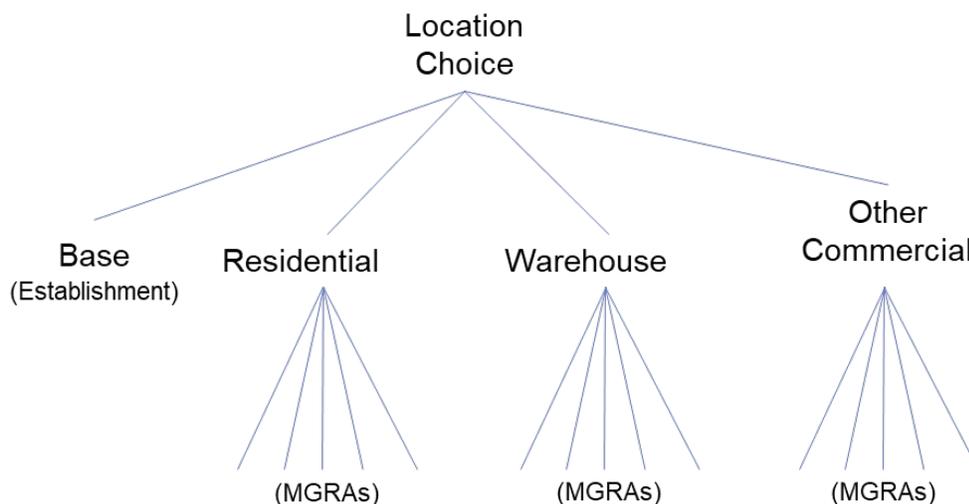
Combined with the Stop Location model described below (Section 3.3.2), this design enables the CVM to model multiple tours on the same vehicle route as complex patterns such as a complete establishment-based tour, followed by a journey to another non-establishment location (e.g., a maintenance facility) at the end of the day. In addition, the termination location of the route plays an important role in the Stop Location Choice model as an anchor point for choosing the locations of other stops, which produces more sensible, internally consistent stop sequences and tours (e.g., drivers choose stop locations at the end of the day that are closer to the route termination location).

The Route Origin and Termination Location Choice step includes the following sub models:

- **Route Origin Location Choice**
 - a. Choice of Location Type
 - b. Given Location Type, Choice of Zone
- **Route Termination Location Choice**
 - a. Choice of Location Type
 - b. Given Location Type, Choice of Zone

The general choice structure for both Route Origin and Termination Location choices is depicted in Figure 3.2-5. The upper level choice is between the establishment base and other potential locations, such as a residential location, a warehouse/distribution center, or another commercial location. The choice of a location type is intended to provide more accurate predictions of where commercial vehicles start and end their days and another mechanism for controlling the locations of vehicle flows in model validation exercises. As found in the Commercial Vehicle Survey, many drivers start or end their workday at their place of residence, particularly for small businesses. It is also common for freight shippers to leave trucks overnight at a company warehouse, distribution center, or maintenance facility that may be different from their base establishment location.

Figure 3.2-5 Route Origin and Termination Location Choice Structure



The upper-level choice of a location type conditions the lower-level choice of zones (MGRAs). If Base is chosen, then the location is that of the establishment and no further choice need be made. If Residential is chosen, the model chooses from among zones with households. If Warehouse is chosen, then the choice set is restricted to zones with employment in transportation and warehousing. If Other Commercial is chosen, the choice set includes zones that have any type of employment besides transportation and warehousing.

Method

The same method is applied to both Route Origin Location Choice and Route Termination Location choice. These two modeling steps are applied in sequence, Origin followed by Termination. The results of the Origin location type choice may be used to inform the choice of the Termination location type. In both models, travel times and costs are with reference to the base zone of the establishment.

The upper-level choice of location type is a simple multinomial logit model. The lower-level choice is a destination choice model. In the initial estimation of these models, the two choices were explicitly joined in a nested structure, in which the composite utility of the lower-level zone choices fed upward into the utility of the choice of location type. In the course of implementing the model within the ActivitySim software framework, however, this integrated approach presented software engineering challenges which led to the adoption of the simpler strategy of allowing the two levels be applied independently.

Location Type (multinomial logit) – as expressed in the equation below, the probability of selecting location type t for establishment n is a function of estimated parameters β specific to the location type and attributes X of the establishment and route, such as industry group and route purpose, customer type, and vehicle type.

$$Prob_n(t) = \frac{e^{\beta_t X_n}}{\sum_{\forall t \in T} e^{\beta_t X_n}}$$

Location Zone (destination choice) – as expressed in the equation below, the probability of choosing j , given an establishment location i and location type t is a function of the size of potential attractions S of type t in alternative zone j , and the impedance of travel X between zone i (the establishment) and zone j (*route origin or termination*). The parameters β measure the disutility of travel times and costs.

$$Prob(j|i, t) = \frac{S_j^t * e^{\beta X_{ij}}}{\sum_{\forall j \in J_t} S_j^t * e^{\beta X_{ij}}}$$

The size function, S_j^t , shown below, represents a linear combination of k variables representing quantities, such as households, employment, or other attraction variables that satisfy the conditions of location type t . The θ is a scaling constant which is customarily fixed to one, and the γ are parameters representing the attractiveness of each quantity and must be positive-valued.

$$S_j^t = \theta * \ln \left(\sum_k \gamma_k Q_j^k \right)$$

Details of the final model specifications may be found in the Model Estimation and Calibration technical report.

TNC Variation

As described above in Section 3.2.2, TNC routes are generated through a different process from regular establishments. TNC routes are created for each LUZ in the model area and are assumed to be for goods pickup and delivery, serve a mix of both residential and non-residential customers, and use light commercial vehicle types (passenger car equivalents). The choice of route starting and ending locations for routes can also be simplified. While it is assumed that all TNC drivers begin their days from a home location, about half of the TNC routes in the Commercial Vehicle Survey began at other locations, most likely when they switched to working for their pickup and delivery client, which may be at any point during the day. To simplify the model design, it was decided that TNC routes would originate and terminate at a single location, and the location choice method would be a simplified model in which a zone (MGRA) would be chosen from among all of the MGRAs within the LUZ which generated the route. For such a model, there is no impedance term and was formulated based purely on attractiveness as in the following equation:

$$Prob(j|luz) = \frac{S_j}{\sum_{\forall j \in J_{luz}} S_j}$$

Inputs

The primary inputs to both Origin and Termination choice models are the route simulation results from all previous steps, which include establishment industry group and route vehicle type, purpose, customer type, and starting time. For Termination Location, choice the outcome of the Origin location type choice is known. In addition, the choice of zones requires land use file inputs and travel time and cost skims.

Outputs

The outputs of the Route Origin and Termination Location Choice step complete the attributes needed to describe the route starting conditions at the beginning of the route stop simulation loop. The CVM records these results in the **final_routes.csv** file.

Table 3.2-5 lists these attributes (vertically) with examples of complete route records, which include the route purpose, customer type, vehicle type, starting time, origin stop type and zone, and termination stop type and zone. The rightmost record is a TNC route.

Table 3.2-5 CVM Route Attributes

establishment_id	2	2	2	2	120941
business_type	EPO	EPO	EPO	EPO	TNCNRR
route_id	209001	209002	209003	209004	12094113000
route_purpose	service	service	maintenance	service	goods
customer_type	mixed	mixed	na	mixed	mixed
vehicle_type	LCV	LCV	LCV	SUT	LCV
is_tnc	FALSE	FALSE	FALSE	FALSE	TRUE
vehicle_type_abm3	passenger_car	passenger_car	passenger_car	LHDT	passenger_car
start_time	13	8	15	10	14
route_start_time_period	MD	AM	MD	AM	MD
origination_stop_type	base	base	base	base	base
origination_zone	6716	6716	6716	6716	8861
terminal_stop_type	base	commercial	base	base	base
terminal_zone	6716	5581	6716	6716	8861

3.3 Route Stop Simulation Loop

The set of models described in this section simulate the decisions made for each stop in a commercial vehicle route. These include the choices of the next stop purpose and the location of the next stop as well as simulating the amount of time spent at the stop (duration). A complete description of how these three components work together as a system may be found in Section 2.1.3. For each of these model components, a separate, similarly structured set of models covers TNCs and are described below.

The primary outcome of the route stop simulation loop is a trip list. The trip list includes all of the trips simulated to represent travel from one stop to the next. Each trip will include the relevant origin and destination stop purposes, origin and destination location types and zone IDs, starting and ending times of day, stop duration (dwell time), travel time to the stop, elapsed time on the route while on the stop, and vehicle types. The CVM-ABM3 interface will convert these trip lists into trip tables by ABM3 vehicle types for network assignment. Next Stop Purpose Choice

3.3.1 Next Stop Purpose Choice

The Next Stop Purpose Choice model predicts the purpose of each stop on a commercial vehicle route. If this is the first stop on the route, the model will consider only the route starting conditions, which include the industry group of establishment making the route, the route start time, purpose, customer type, and vehicle type. For subsequent choices of the next stop purpose on the route, the model will additionally consider the purpose of the current stop, the current time of day, and elapsed time on the route since it began.

Definitions of stop purposes are as follows:

- **Goods Pickup** – a stop to pick up goods

- **Goods Delivery** – a stop to deliver goods to a customer
- **Service** – a stop to provide a professional service to a customer
- **Maintenance/Other** – a stop for either vehicle maintenance/refueling or driver breaks
- **Base Establishment** – a trip to the establishment where the vehicle is based, without ending the route
- **Home** – a trip to the driver’s home, without ending the route
- **Terminate Route** – a trip to the final stop on the route where the vehicle will be parked, ending the route. This is the location chosen by the Route Termination Location Choice model described in Section 3.2.5.

Figure 3.3-1 is a matrix showing the allowable transition between current stop purposes and potential next stop purposes for establishment routes. Infeasible transitions are shown with a red ‘x.’ The decision to terminate the route can never be the first purpose and once terminate route is chosen, a final trip to the termination location is created and the route simulation ends. Other transitions that are prohibited include going from base to base (same establishment location) and from home to home because they would involve no travel.

Figure 3.3-1 Next Stop Purpose Feasible Choice Sets for Establishment Routes

Current Purpose	Next Purpose						
	base	goods pickup	goods delivery	service	maint. / other	home	terminate route
base	X						
goods pickup							
goods delivery							
service							
maint. /other							
home						X	

A similar transition matrix is shown in Figure 3.3-2 for TNCs. The main difference is that TNCs are assumed not to provide services in the CVM, only pickup and deliver goods.

Figure 3.3-2 Next Stop Purpose Feasible Choice Sets for TNC Routes

Current Purpose	Next Purpose					
	base	goods pickup	goods delivery	maint. / other	home	terminate route
base	X					
goods pickup						
goods delivery						
maint. /other						
home					X	

In terms of sensitivity, the model specification should also consider accessibility variables that might affect the propensity to make another stop of a particular type. For example, the probability of making another goods delivery stop may be influenced by accessibility to households with food or package delivery demands. (See Section 3.1.4 for more about accessibility variables.)

The model should also consider time of day in two different forms. The model should be sensitive to the clock time or time period, such that certain activities are more likely during regular business hours and such that routes are more likely to terminate later in the day. In addition, as the elapsed time on the tour increase, the likelihood of choosing to terminate the route will increase, as may making a maintenance stop or the driver visiting their home.

Method

The Next Stop Purpose Choice model is formulated as a multinomial logit model in which the choice alternatives represent the feasible transitions between consecutive stop purposes, as defined above. The model is formulated as a multinomial logit model. As expressed in the equation below, given the current stop purpose p_n , the probability of choosing the next stop purpose p_{n+1} for route r is a function of estimated parameters β and establishment and route attributes X , such as industry group, the current stop purpose, time of day, elapsed time on the route, customer type, vehicle type, and potentially accessibility variables.

$$Prob_r(p_{n+1}|p_n) = \frac{e^{\beta p X_r}}{\sum_{\forall p \in P} e^{\beta p X_r}}$$

Details of the final model specification may be found in the Model Estimation and Calibration technical report.

Inputs

The primary inputs to the Next Stop Purpose Choice model are the route-level attributes (see Table 3.2-5) and the in-memory variables related to the current simulated clock time, elapsed time since the route began, and the current location (MGRA).

Outputs

The Next Stop Purpose Choice model has one primary output, the choice of the next stop purpose. If this purpose is “terminate route,” then the route stop simulation loop creates a final trip to the route termination location and ends the simulation of that route. For any other next stop purpose, the route stop simulation loop proceeds to the Next Stop Location Choice model. All trip level outputs are recorded in the file **final_trips.csv**.

3.3.2 Next Stop Location Choice

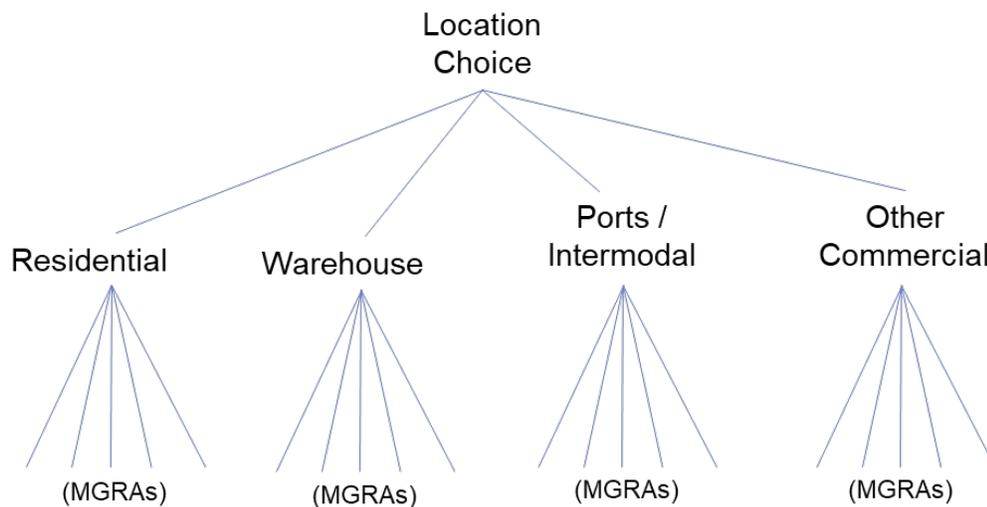
The Next Stop Location Choice model chooses the location of the next stop on the route, given the upstream choice of a stop purpose. The model is not applied if the next stop purpose is either “base” (establishment) or “terminate route” because those locations are already known. The same general model structure is used by both regular establishments and TNCs.

The Next Stop Location Choice model is formulated as a two-level model and consists of the following sub models:

- **Next Stop Location Choice**
 - a. Choice of Location Type
 - b. Given Location Type, Choice of Zone

The general choice structure for both Route Origin and Termination Location choices is depicted in Figure 3.3-3. The upper level choice is between location types such as residential locations, warehouse/distribution centers, ports/intermodal, or another commercial location. The choice of a location type is intended to provide more accurate predictions of commercial vehicles destinations to certain land use types, particularly warehouses and ports by making these explicit high-level choices. The choice of Location Type should be sensitive to the route and stop purposes (goods pickup/delivery, service provision, maintenance/other); route purpose; route customer type (residential, commercial, mixed), vehicle type (light, medium, heavy), and time of day factors (business hours).

Figure 3.3-3 Next Stop Location Choice Structure



The upper-level choice of a location type conditions the lower-level choice of zones (MGRAs). If Residential is chosen, the model chooses from among zones with households. If Warehouse is chosen, then the choice set is restricted to zones with employment in transportation and warehousing. If Port/Logistics Nodes is chosen, the choice is restricted to a small set of MGRAs contained in TAZs which have been identified as port facilities (marine, air, intermodal). If Other Commercial is chosen, the choice set includes zones that have any type of employment besides transportation and warehousing.

Method

The upper-level choice of location type is a simple multinomial logit model. The lower-level choice is a destination choice model. In the initial estimation of these models, the two choices were explicitly joined in a nested structure, in which the composite utility of the lower-level zone choices fed upward into the utility of the choice of location type. In the course of implementing the model within the ActivitySim software

framework, however, this integrated approach presented software engineering challenges which led to the adoption of the simpler strategy of allowing the two levels be applied independently.

Location Type (multinomial logit) – as expressed in the equation below, the probability of selecting location type t for route r is a function of estimated parameters β specific to the location type and attributes X of the establishment and route, such as industry group and route and stop purpose, customer type, vehicle type, and time of day.

$$Prob_r(t) = \frac{e^{\beta_t X_r}}{\sum_{\forall t \in T} e^{\beta_t X_r}}$$

Location Zone (destination choice) – as expressed in the equation below, the probability of choosing zone j , given current location i and location type t is a function of the size of potential attractions S of type t in alternative zone j , the impedance of travel X between zone i (current zone) and zone j (alternative next zone), and the impedance of travel X between zone j (alternative next zone) and zone k (terminal zone). The parameters β measure the disutility of travel times and costs.

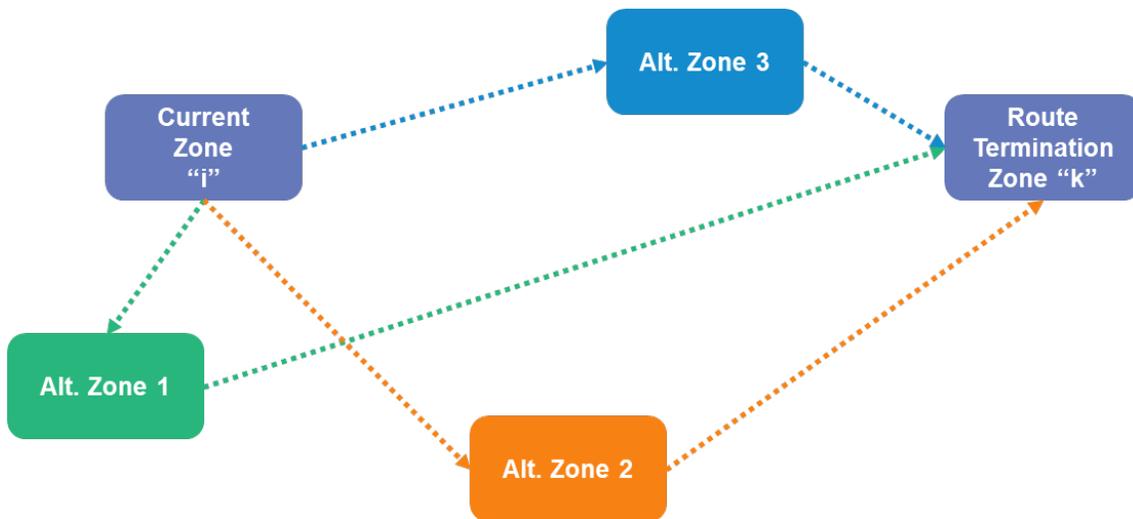
$$Prob(j|i, t) = \frac{S_j^t * e^{\beta X_{ij} + \beta X_{jk}}}{\sum_{\forall j \in J_t} S_j^t * e^{\beta X_{ij} + \beta X_{jk}}}$$

The size function, S_j^t , shown below, represents a linear combination of k variables representing quantities, such as households, employment, and other attraction variables that satisfy the conditions of location type t . The θ is a scaling constant which is customarily fixed to one, and the γ are parameters representing the attractiveness of each quantity and must be positive-valued.

$$S_j^t = \theta * \ln \left(\sum_k \gamma_k Q_j^k \right)$$

The impedance portion of the Next Stop Location Choice model considers the travel time and costs to each feasible zone alternative and the travel time and costs from each alternative to the route termination zone, which acts as an anchor point for all stop location decisions on the route. This “rubber banding” concept is illustrated in Figure 3.3-4.

Figure 3.3-4 Example of "Rubber Banding" in Next Stop Location Choice



Stops closer to the current zone, such as Alt. Zone 1 in the example, might be farther from the route termination zone, whereas other alternatives such as Alt. Zone 3 in the example, are farther from the current zone and closer to the termination zone. The model is designed to be sensitive to both types of impedance. When the model is applied, it should balance the attraction potential at each candidate zone against the travel time and cost to make the most probable choices, given the current location and time of day. For time of day sensitivity, the model should include a variable representing the interaction between elapsed time on the tour and generalized travel time, which would make the model more sensitive to travel time as the elapsed clock time on the route increases. Details of the final model specifications may be found in the Model Estimation and Calibration technical report.

Inputs

The inputs to the Next Stop Location Choice model are the route-level attributes (see Table 3.2-5) and the in-memory variables related to the current stop purpose, simulated clock time, elapsed time since the route began, and the current location (MGRA).

Outputs

The Next Stop Location Choice model has two primary outputs, the choice of the next stop location type and zone (MGRA). All trip level outputs are recorded in the file **final_trips.csv**.

3.3.3 Stop Duration Simulation

The Stop Duration model simulates the amount of time that the commercial vehicle dwells at a stop for any purpose. Assigning a stop duration advances the clock time in the simulation together with the travel time from stop to stop. The model is applied by a random draw from a fitted distribution.

Method

Several different approaches to modeling stop duration were considered and tested with Commercial Vehicle Survey data. The most promising approach proved to be fitting the data to a Beta Distribution, because of its flexibility. The Beta distribution includes two shape parameters and can predict any probability on a unit

interval. The unit interval can be scaled by a maximum duration time. Based on extensive data exploration, two distinct model segmentation schemes were developed for regular establishments and TNCs. Figure 3.3-5 and Figure 3.3-6 show the observed distributions for stop dwell times for regular establishments and for TNCs, respectively. While the vast majority of observations are skewed toward the shorter-duration stops, there are some much longer duration stop observations scattered throughout each distribution.

Figure 3.3-5 Observed Stop Durations (15-minute intervals) Regular Establishments

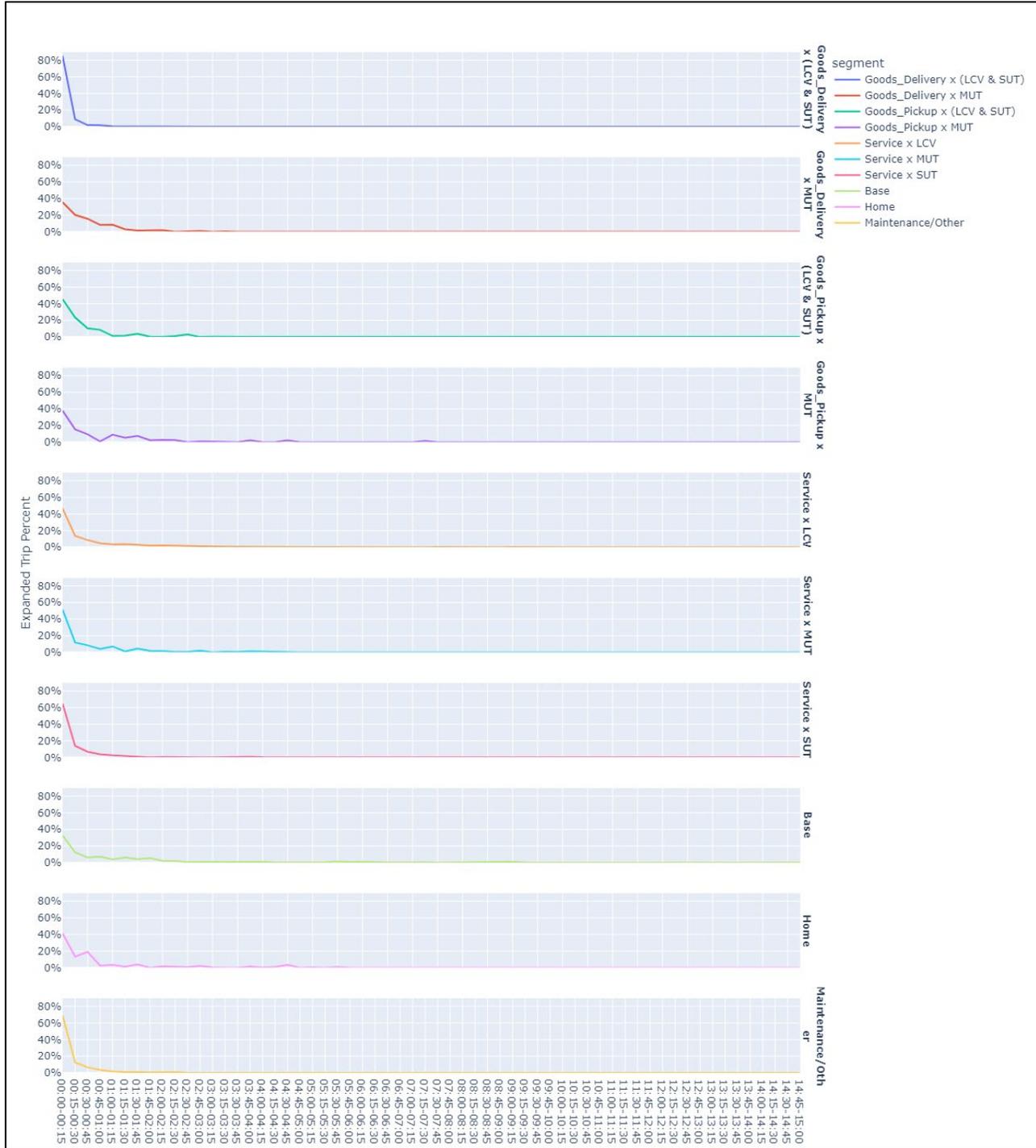
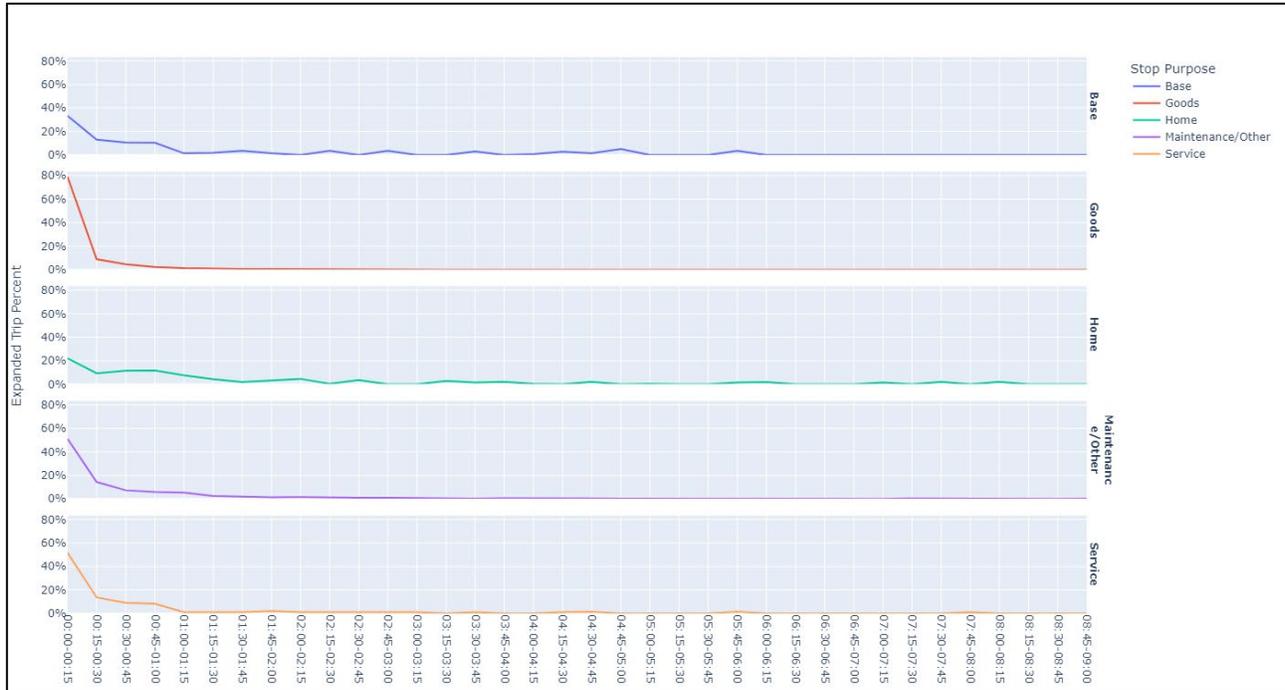


Figure 3.3-6 Observed Stop Durations (15-minute intervals) TNCs



Regular Establishments: the stop duration observations were segmented by purpose and vehicle type and by whether the stop began before or after 12 p.m. The time of day split recognized that stops beginning before 12 noon are more likely to last multiple hours, whereas stops which began after 12 noon were likely to be curtailed by the end of the workday. Table 3.3-1 shows the segments developed for the final model.

Table 3.3-1 Stop Duration Model Segmentation for Regular Establishments

Purpose x Vehicle	Before 12pm	After 12pm
Goods Delivery x [LCV or SUT]		
Goods Delivery x MUT		
Goods Pickup x [LCV or SUT]		
Goods Pickup x MUT		
Service x LCV		
Service x SUT		
Service x MUT		
Base		
Home		
Maintenance / Other		

Beta distribution shape parameters were estimated for each market segment and maximum duration assumed based on observed durations. Details of the final model specifications may be found in the Model Estimation and Calibration technical report.

TNCs: the stop duration observations were segmented by purpose and by three land use density variables. The thought behind using land use density variables that serve as a proxy for how long a TNC driver might take to pick up or deliver in higher density urban areas due to parking conditions and walking within buildings. The land use variables from the MGRA land use file for population and employment were summed to create a total density variable. Through statistical analysis break points labeled as low-, medium-, and high-density were determined. Table 3.3-2 shows the segments developed for the final model.

Table 3.3-2 Stop Duration Model Segmentation for TNCs

Purpose	Low Density (<15)	Medium Density (15 to 35)	High Density (> 35)
Goods Delivery			
Goods Pickup			
Base			
Home			
Maintenance / Other			

Inputs

The inputs to the Stop Duration model are the route-level attributes (see Table 3.2-5) and the in-memory variables related to the current stop purpose, simulated clock time, and the chosen stop location (MGRA). For TNCs, the MGRA land use data file was used to provide a variable for total density = population + employment.

Outputs

The Stop Duration model produces a single output, the simulated stop dwell time in minutes. The simulation draws a random value from the fitted distribution, described above. Once the dwell time is known, it is added to the travel time for the trip to advance the simulation clock, and control returns to the Next Stop Purpose mode. All trip level outputs are recorded in the file **final_trips.csv**.

3.3.4 Stops and Trips Complete Outputs

Table 3.3-3 lists these attributes (vertically) with an example of a set of trips for a route which begins and ends at its establishment location.

Table 3.3-3 CVM Stop and Trip Attributes

route_id	1209006	1209006	1209006	1209006	1209006
route_trip_num	1	2	3	4	5
trip_origin	11362	7114	8872	13144	5343
trip_destination	7114	8872	13144	5343	11362
trip_origin_purpose	originate	service	service	service	maintenance
trip_destination_purpose	service	service	service	maintenance	terminate
trip_destination_type	commercial	commercial	commercial	commercial	base
trip_start_time	9	11	16	19	20
trip_travel_time	4.9	8.7	9.8	13.8	14.4
dwelt_time	61.3	143.0	101.2	4.4	0.0
route_elapsed_time	66.2	217.8	328.8	347.0	361.4
cv_trip_id	1209006001	1209006002	1209006003	1209006004	1209006005
taz_origin	1357	1449	1447	1698	1255
taz_destination	1449	1447	1698	1255	1357
vehicle_type	MHDT	MHDT	MHDT	MHDT	MHDT
tod	AM	AM	MD	MD	MD
distanceDrive	1.31	2.82	3.74	5.98	9.86
costTollDrive	0	0	0	0	0
costOperatingDrive	133.66	288.09	382.70	611.80	1008.80

4.0 Heavy Truck Model

This section discusses the Heavy Duty Truck Model (HTM). This is the last component of the 2024 SANDAG Commercial Vehicle Model and focuses on long-distance heavy duty truck flows.

More specifically, the HTM covers long-distance truck movements *into, out of, or through* San Diego County. The key input demand source driving the HTM is the set of commodity flows between shippers and receivers throughout North America that focus on:

- Flows with **one trip end** (shipper or receiver) in San Diego County which are also referred to as internal-to-external or external-to-internal flows such as flows between Chicago and San Diego; and
- Flows which **pass through** San Diego County which are also referred to as “through trips.” An example of such flows would be freight flows between Mexico and Los Angeles.

Commodity flows are derived from the Federal Highway Administration (FHWA) Freight Analysis Framework version 5 (FAF5). FAF also includes movements to/from port facilities in San Diego that have an external trip end. As discussed in detail under Sections 2.0 and 3.0, the CVM accounts for freight and non-freight truck trips between establishments within San Diego County. It also explicitly accounts for truck movements involving warehouse and distribution centers and port facilities within San Diego County.

4.1 HTM Workflow

The 2024 HTM design also builds on and improves the ABM2 HTM model created in 2014 to account for long-distance freight truck flows. The new HTM design differs from the previous model in several important ways that include:

- Better accuracy by leveraging the most current fifth version of the Freight Analysis Framework to reflect more recent freight flow data;
- FHWA has developed FAF5 forecasts to better representing commodity trading and supply chain trends following the COVID-19 pandemic; and
- The HTM update includes improvements to process efficiency and accuracy and it provides greater flexibility to the user to evaluate different scenarios.

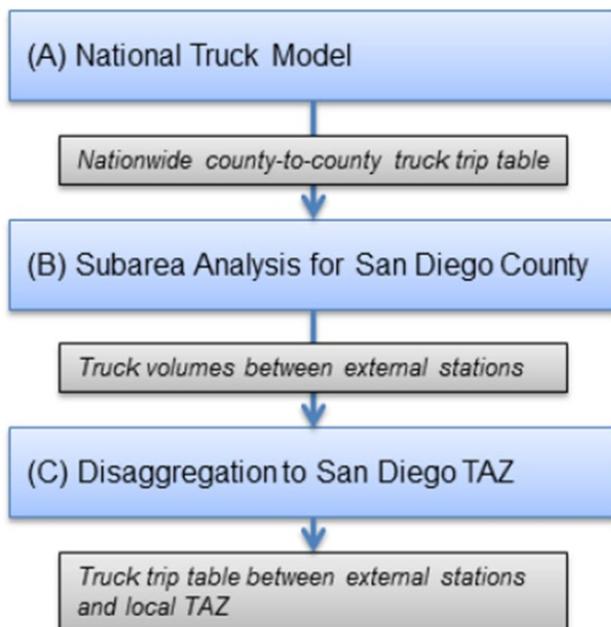
To highlight the different features between the earlier HTM and the 2024 version of the HTM model, we first provide the outline of the workflow in ABM2 using the flow chart in Figure 4.1-1.

The earlier HTM disaggregated FAF at the national level (to county-to-county flows) which was unnecessary. For example, it disaggregated flows between Florida and Maine which has nothing to do with San Diego County. The earlier model also provided truck volumes between external stations.

The last step of the earlier HTM model included the disaggregation of these station-to-station flows to the traffic analysis zones (TAZ) within San Diego county to provide a TAZ level of detail for estimates of heavy truck flows that originate in, are destined to, or are passing through the San Diego area.

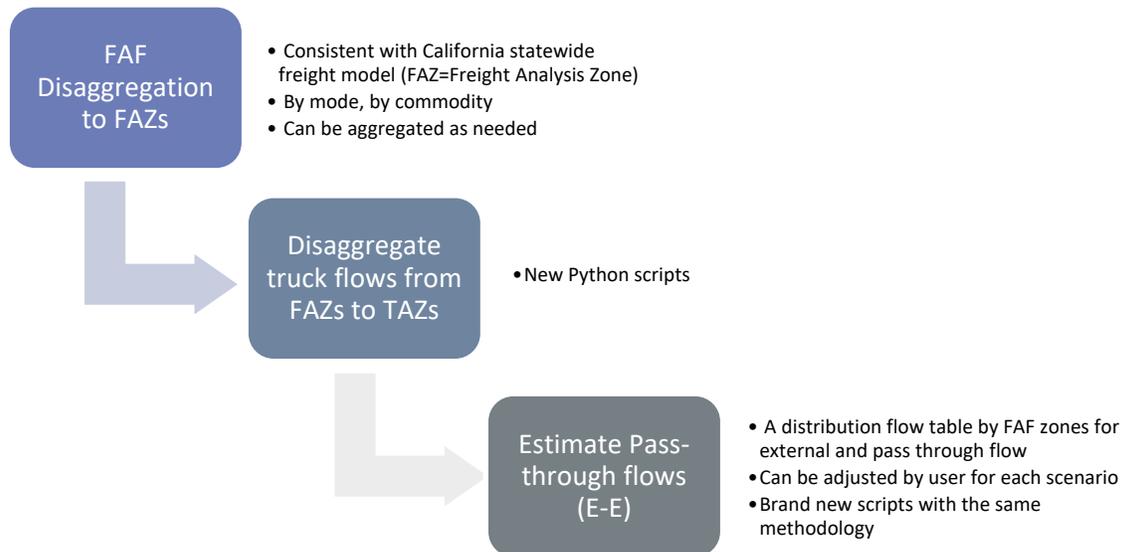
The workflow in ABM2 was described in flow chart below:

Figure 4.1-1 Previous HTM Workflow in ABM2+



The 2024 HTM model was updated and its workflow is shown in Figure 4.1-2. The new HTM version pivots off the newest version of FAF to provide more accurate up-to-date long-distance heavy-duty truck flow estimates, improves process efficiency, and provides greater flexibility to the user.

Figure 4.1-2 Updated HTM Workflow in ABM3



The first key enhancement of the 2024 HTM model is the use of the FAF disaggregation process to California Freight Analysis Zones (FAZ) instead of the national county-to-county truck trip table used in earlier databases. This approach is consistent with the California statewide freight model which provides a better starting point for the HTM model for the San Diego area. The detail available includes mode and commodity information and the user has greater flexibility to aggregate geographies of interest.

The second step of the process involves the disaggregation from FAZ to TAZ using a set of newly developed Python scripts. This step includes the assessment of updated “internal-to-external” and “external-to-internal” long-distance truck flows.

The third step focuses on the “pass through” long-distance truck flows (also referred to as “external-to-external” long distance truck flows) with scripts that allow the user to customize the analysis for different scenarios under study.

4.2 Major HTM Updates

The major updates to the 2024 HTM model can be summarized as follows:

- The inefficient subarea extraction step has been removed. Instead, the HTM uses only FAF zone interchanges that involve flows into, out of, or through San Diego County.
- All inputs have been consolidated in a single Excel file for better quality assurance / quality control and to facilitate the consistent evaluation across different scenarios.

- The new HTM provides an array of new features and allows the user the flexibility to add new gateways and change the distribution by gateway. The user can also adjust parameters such as average payloads and time-of-day distribution in the input Excel workbook.
- The daily traffic volumes at 12 gateways around San Diego County are set to be the control totals for the “internal to external”, “external to internal”, and “external to external” truck flows to provide a consistent analysis framework.
- Scripting for the updates used Python code for consistency with the ABM2+ and ABM3 code bases.

4.3 Detailed Steps in Running the Updated HTM

A series of seven steps is needed to run the new HTM. This summary is meant to provide a brief “how to” guide.

1. Aggregate by Commodity Group

- a. The objective of this step is to determine the level of aggregation of commodities that is appropriate for the analyses and scenarios of interest.
- b. A lookup table is available that contains the 40+ commodity groups of the Standard Classification of Transported Goods (SCTG) and their corresponding 15 aggregated commodity categories.
- c. The user can aggregate the annual tonnage by mapping each commodity to its respective aggregated commodity category.

2. Disaggregate Tonnage to TAZs

- a. The objective of this step is to distribute the FAF5 tonnage from the FAZ level to a more refined geography of interest.
- b. The user can implement a lookup table that links employee categories to the corresponding commodities since employment is used as the allocation criterion.
- c. The aggregated tonnage can be disaggregated and allocated to each TAZ based on the employment in each TAZ’s employment categories that were chosen.

3. Estimate Daily Tonnage

- a. An annualization factor that translates annual tonnage to daily tonnage is used to obtain average daily tonnage values for the Origin-Destination table. This step provides a more granular view for further analysis.

4. Truck Type Assignment

- a. Truck allocation is determined based on the distance between the origin and destination of each tonnage movement.
- b. Conversion rates that are specific to truck types (e.g., Light, Medium1, Medium2, Heavy Trucks) are used to distribute the daily tonnage to the number of different truck types. This classification is consistent with the California Statewide Freight Forecasting Model.
- c. Medium1 and Medium2 trucks are later combined to form the total truck and passenger car flows as an input to network assignment.

5. Tonnage to Truck Conversion

The number of trucks is calculated based on the average payload of trucks by commodity.

6. Balancing Trips

- a. Long-distance truck trips are balanced using 12 external gateways around San Diego County.
- b. The internal-to-external, external-to-external, and external-to-internal trips will be balanced to account for non-FAF external flows. Non-FAF flows may include truck flows not covered in FAF, e.g., service trucks, garbage trucks, empty back-haul, etc.
- c. The FAF-related flows will be subtracted from the truck counts at each gateway. The remainder will be distributed in a manner that is similar to the distribution of aggregate FAF flows to each TAZ based on employment. This step will account for empty trucks and any other non-FAF External trips.

7. Distribute Trucks by Time of Day

4.4 HTM Outputs

The Python script prepares a series of interim and final outputs. The interim outputs are mainly used for model calibration. The final outputs are integrated with the Commercial Vehicle Model and the Passenger Model for a multi-class network assignment. The list of final outputs from the Heavy Truck Model includes:

- Commodity Flows by TAZ for trips that have one end within the SANDAG region
- Truck Origin-Destination trip tables by Truck Type and by Time of day