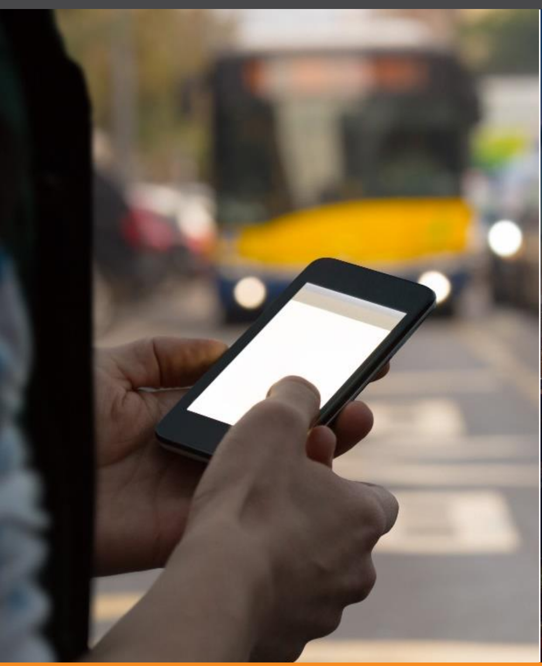
Oregon Department of Transportation

MODEL DESIGN AND ESTIMATION APPROACH

April 24, 2024



55 Railroad Row
White River Junction, VT 05001
802.295.4999
www.rsginc.com

With Data Perspectives Consulting

CONTENTS

1.0 INTRODUCTION 1

2.0 KEY QUESTIONS AND NEEDS 4

2.1 MODEL REQUIREMENTS..4

FEDERAL METROPOLITAN PLANNING ORGANIZATION REQUIREMENTS 4

STATE AND LOCAL PLANNING REQUIREMENTS 6

LAND USE PLANNING AND ANALYSIS...7

EQUITY ANALYSIS 8

GREENHOUSE GAS ANALYSIS 10

BICYCLE AND PEDESTRIAN PLANNING AND ANALYSIS................12

CHANGING DEMOGRAPHICS 15

PRICING 17

RELIABILITY ANALYSIS AND REPORTING 18

SPECIAL MARKET TRAVEL...23

DEMAND RESPONSE TO SUPPLY 25

2.2 KEY QUESTIONS...26

2.3 DESIRED PRINCIPLES AND FEATURES 27

GUIDING PRINCIPLES 27

BALANCING ANALYTIC CAPABILITIES AGAINST RUN TIME AND COMPUTING RESOURCES. 28

PARTIAL MODEL RUN MODE(S) 28

STANDARD DIAGNOSTIC TOOLS 28

POST-PROCESSOR LINKAGES 29

FTA GUIDANCE ON TRAVEL MODEL PARAMETERS AND USE 29
FLEXIBLE AND EXPANDABLE DESIGN 31
2.4 FEATURES NECESSARY TO SUPPORT NEEDS AND REQUIREMENTS 31
3.0 MODEL SPECIFICATION AND DESIGN 51
3.1 ACTIVITYSIM GENERAL MODEL FORM 51
3.2 TRAVEL DEMAND MODEL SEGMENTATION 51
3.3 ACTIVITY TYPES 54
3.4 TREATMENT OF TIME 56
3.5 TREATMENT OF SPACE 57
3.6 OVERALL MODEL DESIGN 57
3.7 POPULATION SYNTHESIS 58
GROUP QUARTERS DEVELOPMENT 61
3.8 ACTIVITYSIM ABM DESIGN 61
3.9 ACTIVITYSIM DETAILED DESCRIPTION 65
ACCESSIBILITIES 65
MOBILITY MODELS 69
DAILY TRAVEL MODELS 82
GENERATION OF JOINT HOUSEHOLD TOURS 91
3.10 MODELING OVERNIGHT VISITOR POPULATION 111
VISITOR MODEL DESCRIPTION 111
3.11 EXTERNAL TRAVEL 115
3.12 FREIGHT\COMMERCIAL VEHICLE TRAVEL 115
3.13 UNIVERSITY TRAVEL 117
3.14 LAND-USE MODEL CONNECTIONS 118
3.15 PARKING CONSTRAINTS 119
SHADOW PRICING APPROACHES 119
SIMULATION-BASED APPROACHES 120
4.0 ANTICIPATED INPUT AND OUTPUT SPECIFICATION 122
4.1 NETWORK ATTRIBUTES 122
HIGHWAY NETWORK ATTRIBUTES 122
TRANSIT NETWORK ATTRIBUTES 123
NON-MOTORIZED NETWORK ATTRIBUTES 124
4.2 LEVEL-OF-SERVICE SKIMS 126
PASSENGER AUTO SKIMS 126
TRUCK SKIMS 127
TRANSIT SKIMS 128
NON-MOTORIZED SKIMS 129
4.3 LAND-USE DATA INPUTS 129
4.4 SYNTHETIC POPULATION INPUTS 136
4.5 SYNTHETIC POPULATION OUTPUTS 136

4.6 ACTIVITYSIM SETTINGS....140
ACTIVITYSIM GLOBAL PARAMETERS....140
HOUSEHOLD, PERSON, AND LAND USE PREPROCESSORS....145
ACTIVITYSIM MODEL-RELATED FILES....147
4.7 ACTIVITYSIM OUTPUTS....148
5.0 DESIGN IMPLEMENTATION SPECIFICATIONS....165
5.1 SKIMMING AND ASSIGNMENT....165
TIME PERIODS....165
HIGHWAY ASSIGNMENT....167
TRANSIT ASSIGNMENT....169
NON-MOTORIZED ASSIGNMENT....171
DYNAMIC TRAFFIC ASSIGNMENT....178
ASSIGNMENT DIMENSIONS AND REQUIRED RESOURCES....180
5.2 DONOR MODEL IMPLEMENTATION....181
5.3 DATA PROCESSING....182
HOUSEHOLD SURVEY DATA....182
TRANSIT ON-BOARD SURVEY DATA....184
CENSUS DATA....186
TRAFFIC COUNT DATA....186
PASSIVE DATA....187
5.4 MODEL ESTIMATION....188
INTRODUCTION AND OVERVIEW....188
DATA PREPARATION....190
CREATING ESTIMATION DATA BUNDLES....197
MODEL ESTIMATION IN LARCH....199
HANDLING MULTIPLE SURVEY DAYS....199
JOINT ESTIMATION ACROSS MULTIPLE REGIONS....200
5.5 MODEL CALIBRATION....201
5.6 MODEL APPLICATION SCENARIOS....208
CONGESTION PRICING SCENARIO....208
ALL-ROAD USE FEE....210
HIGH OCCUPANCY TOLL LANES....212
CAR FREE/CAR LIMITED DOWNTOWNS....214
MICROMOBILITY (ELECTRIC BIKES\SCOOTERS)....215
5.7 SCENARIO MANAGER....216
5.8 POTENTIAL SIMPLIFICATIONS FOR SPECIFIC IMPLEMENTATIONS....218
6.0 NEXT STEPS....220
6.1 SURVEY DATA PROCESSING AND AUGMENTATION....220
6.2 PROPOSED DONOR MODEL, DEPLOYMENT, AND REQUIRED ENHANCEMENTS....221

6.3 LIMITATIONS OF THE PROPOSED DESIGN AND POTENTIAL ENHANCEMENTS 233
6.4 COORDINATION WITH ACTIVITYSIM CONSORTIUM 236
6.5 SAMPLE LIST OF MODEL OUTPUTS AND METRICS 237
7.0 CONCLUDING REMARKS 240

LIST OF FIGURES

FIGURE 1: JUSTICE 40 SCREENING TOOL 8
FIGURE 2: TRAVEL TIME RELIABILITY AS A FUNCTION OF TRAVEL TIME 20
FIGURE 3: STANDARD DEVIATION OF TRAVEL TIME 21
FIGURE 4: TRAVEL TIME DATA FOR A NETWORK LINK IN SAN DIEGO 23
FIGURE 5: SANDAG VOLUME-DELAY FUNCTION WITH UNRELIABILITY COMPONENT 23
FIGURE 6: OVERALL MODEL DESIGN 58
FIGURE 7: MOBILITY MODELS 63
FIGURE 8: DAY PATTERN AND PRIMARY ACTIVITY/TOUR MODELS 64
FIGURE 9: STOP AND TRIP LEVEL MODELS 65
FIGURE 10: DAILY ACTIVITY PATTERN CHOICES FOR A THREE PERSON HOUSEHOLD 84
FIGURE 11: SCHOOL DROPOFF/PICKUP MODEL EXAMPLE OF BUNDLING CHILDREN BY HALF-TOUR 88
FIGURE 12: CHAPERONE ASSIGNMENT AND TOUR TYPE 89
FIGURE 13: JOINT TOUR PARTICIPATION MODEL 94
FIGURE 14: PROPOSED MODE CHOICE STRUCTURE 101
FIGURE 15: SANDAG VISITOR MODEL DESIGN 114
FIGURE 16: OREGON METRO PARK AND RIDE LOT CHOICE MODEL 120
FIGURE 17: PARKING COST GRAPHIC 135
FIGURE 18: VOLUME DELAY FUNCTION WITH INTERSECTION DELAY 168
FIGURE 19: MODEL ESTIMATION WORKFLOW 189
FIGURE 1920 AN EXAMPLE OF AN ACTIVITYSIM YAML FILE CONTAINING MODEL PARAMETERS 217
FIGURE 21: MODEL DEVELOPMENT PROCESS FLOWCHART 222
FIGURE 22: MODEL DEVELOPMENT TASK LIST AND OVERALL SCHEDULE 224

LIST OF TABLES

TABLE 1: MODEL SENSITIVITIES AND FEATURES 32
TABLE 2: ESTIMATED LEVEL OF EFFORT FOR MODEL FEATURES AND OREGON MODELING STEERING COMMITTEE RESPONSE 41
TABLE 3: MODEL FEATURES BY OREGON MODELING STEERING COMMITTEE RANKING 50
TABLE 4: HOUSEHOLD INCOME CATEGORIES (YEAR $2023) 52
TABLE 5: AUTO PARITY CATEGORIES 53
TABLE 6: PERSON TYPES 53
TABLE 7:OUT-OF-HOME ACTIVITY TYPES 55
TABLE 8: MAZ LEVEL CONTROLS 59
TABLE 9: BLOCK GROUP CONTROLS 60
TABLE 10: TRACT LEVEL CONTROLS 60
TABLE 11: URBAN GROWTH BOUNDARY CONTROLS 61
TABLE 12: PROTOTYPE POPULATION FOR DISAGGREGATE ACCESSIBILITIES - HOUSEHOLD VARIABLES 67
TABLE 13: PROTOTYPE POPULATION FOR DISAGGREGATE ACCESSIBILITIES – PERSON VARIABLES 68
TABLE 14: BICYCLE COMFORT LEVELS AND POTENTIAL CLASSIFICATION BASED ON HOUSEHOLD SURVEY RESPONSES 72
TABLE 15: ALTERNATIVES IN COORDINATED DAILY ACTIVITY PATTERN MODEL 83
TABLE 16: ALTERNATIVES FOR ASSIGNMENT OF TWO TASKS TO CHAPERONES 89
TABLE 17: ALTERNATIVES FOR ASSIGNMENT OF THREE TASKS TO CHAPER ONES 89
TABLE 18: MODE CHOICE ALTERNATIVES AND NETWORK LEVEL-OF-SERVICE VARIABLES 102
TABLE 19: NON-MOTORIZED NETWORK NODE ATTRIBUTES 125

TABLE 20: NON-MOTORIZED NETWORK LINK ATTRIBUTES 126
TABLE 21: AUTO SKIM TABLES 127
TABLE 22: TRUCK SKIM TABLES....127
TABLE 23: TRANSIT SKIM TABLES....128
TABLE 24: MAZ_MAZ_WALK.CSV FIELDS....129
TABLE 25: MAZ_MAZ_BIKE.CSV FIELDS....129
TABLE 26: MAZ LAND-USE DATA FILE....130
TABLE 27: SYNTHETIC HOUSEHOLDS (SYNTHETIC_HHS.CSV) FILE FIELDS 137
TABLE 28: SYNTHETIC PERSONS (SYNTHETIC_PERSONS.CSV) FILE FIELDS 138
TABLE 29: ACTIVITYSIM GLOBAL PARAMETERS 142
TABLE 30: COMMON HOUSEHOLD PREPROCESSOR COLUMNS....145
TABLE 31: COMMON PERSON PREPROCESSOR COLUMNS 145
TABLE 32: COMMON LANDUSE PREPROCESSOR COLUMNS 146
TABLE 33: VEHICLE ALTERNATIVES FILE 148
TABLE 34: ACTIVITYSIM OUTPUT HOUSEHOLD FILE FIELDS....150
TABLE 35: ACTIVITYSIM OUTPUT PERSON FILE FIELDS 152
TABLE 36: ACTIVITYSIM OUTPUT VEHICLE FILE FIELDS....157
TABLE 37: ACTIVITYSIM OUTPUT TOUR FILE FIELDS 158
TABLE 38: ACTIVITYSIM OUTPUT TRIP FILE FIELDS 161
TABLE 39: MAPPING BETWEEN HALF-HOURLY PERIODS AND SKIMS 166
TABLE 40: TIME PERIODS USED FOR AUTO ASSIGNMENT 166
TABLE 41: TIME PERIODS USED FOR TRANSIT ASSIGNMENT....166
TABLE 42: VOLUME-DELAY FUNCTION TERMS USED IN THE SOUTHERN OREGON ACTIVITY-BASED MODEL....168
TABLE 43: OREGON METRO SIMPLIFIED BIKE PATH COST UTILITY FUNCTION PARAMETERS....172
TABLE 44: OREGON METRO BIKE PATH CHOICE UTILITY FUNCTION PARAMETERS....174
TABLE 45: SANDAG BIKE ROUTE CHOICE COEFFICIENTS....177
TABLE 46: DIMENSIONS OF EACH REGION AND LIKELY REQUIRED DISK SPACE AND RAM....180
TABLE 47: AVAILABLE ON-BOARD SURVEYS IN OREGON....185
TABLE 48: ACTIVITYSIM CALIBRATION SUMMARIES....205
TABLE 49 AN EXAMPLE OF A SCENARIO MANAGER INPUT FILE....216
TABLE 50: MODEL COMPONENTS, ESTIMATION DATA IN DONOR MODEL, AND RELATED INFORMATION FOR ESTIMATION....226
TABLE 51: PURPOSE AND PRIMARY LINKAGES FOR NON-CORE MODEL COMPONENTS....231
TABLE 52: METRICS FROM TRAVEL MODEL AND CALCULATIONS REQUIRED TO SUPPORT THEM 238

1.0 INTRODUCTION

The Oregon Department of Transportation (ODOT), under the Oregon Modeling Improvement Program (OMIP), embarked on a project in 2023 to create a common platform for the next generation of travel models in Oregon. The project is coordinated with the Oregon Modeling Steering Collaborative (OMSC) and includes participation from a number of member agencies including Portland Metro, Mid-Willamette Valley Council of Governments, Lane Council of Governments, Bend Metropolitan Planning Organization, Rogue Valley Council of Governments, and Oregon Cascades West Council of Governments.

The new model system will be jointly estimated[1] using a household travel survey currently being collected at a statewide level. This new travel model will be an Activity Based Model (ABM), which is a transition from the current trip-based models largely used within Oregon and is anticipated to utilize the ActivitySim code base (https://github.com/ActivitySim). A primary driver to transition to ABM is to better respond to policy questions around equity, emerging technologies, and active transportation modes. After estimation this work also includes further phases of deploying and calibrating this next generation of travel demand models for the MPO regions within Oregon. This report describes the design of the new model system.

Regional travel demand models such as trip-based and activity-based models are typically used for ***Tactical*** level analysis. Their primary purpose is to analyze projects and programs for Metropolitan Planning Organization (MPO) regional transportation plans (RTPs), as well as transportation system plans (TSPs) for cities and counties in or near MPO areas. The history of travel demand modeling in the United States can be traced back to the 1950's, when the regions of Detroit and Chicago developed computerized methods for predicting traffic demand for each region's freeway systems. As technology progressed, and travel demand modeling tools were more widely adopted, the capabilities of travel demand models improved. The level of spatial detail increased; transportation analysis zones grew from a few dozen to hundreds and even thousands. Travel markets were segmented by explanatory variables to incorporate important differences in travel behavior, and the functionality of travel demand models expanded from a focus on auto demand to consider transit and eventually non-motorized modes. These enhancements were made possible by improvements in the mathematical tools available to represent travel behavior and the increasing availability of observed data. Diversion curves were replaced by utility-maximizing choice models. Iterative capacity-constraint traffic assignment methods were replaced by Frank-Wolfe equilibrium solutions, and new methods of transit pathfinding were made available.

[1] Model estimation refers to the process of using statistical models to fit parameters to observed data given an assumed model form.

Oregon has a strong history of travel demand modeling. Portland Metro is a pioneering agency for transit forecasting and integrated land-use/transportation modeling. Oregon DOT was an early adopter and innovator in statewide integrated modeling. In 1996, the OMSC was formed as a peer review group for ODOT's Oregon Modeling Improvement Program. When the OMSC was first convened, participating agencies saw a need for collaboration as modeling tools for analysis of statewide and regional systems were developed by different transportation and planning organizations around the state. In addition, the OMSC has served as a valuable sounding board as ODOT and other modeling agencies developed modeling tools and associated data and research. The OMSC was instrumental during development of Oregon's first Statewide Integrated Model (SWIM), helping to ensure consistency between SWIM and other travel models managed by ODOT and Oregon's three largest metropolitan planning organizations. The forum has also allowed Oregon to leverage the collective strengths and experiences of multiple agencies in developing new analysis tools.

The model system described in this document will eventually be used as a replacement for the traditional "four-step" travel demand model currently used by Oregon's MPOs, sometimes refered to as JEMnR or "Joint Estimation Model in R code". The "four-step" process refers to the traditional methodology used for travel demand forecasting: 1. Trip generation, 2. Trip distribution, 3. Mode choice and 4. Trip assignment. These models are often enhanced with additional detail on commercial travel (long haul and local delivery). The JEMnR model system was estimated with household survey data collected at a statewide level, but each region's implementation is dependent upon local data and calibrated to match travel conditions in each unique location. Models are further customized for specific geographic areas. For example, the Corvallis-Albany-Lebanon model has a special university model overlay to deal with travel behaviors for Oregon State University and uses information from the region's on-board transit survey to refine transit representation. The Portland Metro model has more sophisticated treatment of bicycle route choice and transit parking constraints.

This design follows a similar approach in the specification of the new activity-based model system; model parameters will be estimated using pooled data from each participating region, but the data used for model implementation (land-use, networks, synthetic population) will be unique to each region, and each implementation will be calibrated to local travel behavior as indicated by that region's sample of survey data, as well as traffic counts, transit boardings, and any other relevant datasets. Each implementation may be further customized as necessary, though each feature can be implemented in multiple regions depending on need, through a shared software implementation. In addition to shared software code, it is expected that agencies will also take advantage of shared technical resources such as training and staff support.

The remainder of this document is divided into six further sections.

- Section 2.0 identifies key questions and needs to be addressed by the model and lists some alternative methods for addressing those key questions and needs.
- Section 3.0 describes the general design of the model system, including population synthesis and input requirements. The section describes the model form, provides an overview of the various dimensions of travel behavior that the model covers, describes each model component, and discusses options for modeling special travel markets and situations.
- Section 4.0 provides a blueprint for model inputs and outputs, and describes data created by ActivitySim software.
- Section 5.0 describes model implementation details such as network skimming procedures, the preparation of estimation and calibration data, the process of estimating, calibrating, and validating the models, and the application of the models to a few key types of scenarios.
- Section 6.0 describes next steps in the development of the model system; survey data processing, the implementation of a donor model as a starting point for model development, and all auxiliary processes required prior to model estimation. It also provides a schedule for model development and lists a few weaknesses of the model system that may be relevant for future work.
- Section 7.0 offers concluding remarks and summarization.

2.0 KEY QUESTIONS AND NEEDS

2.1 MODEL REQUIREMENTS

The following section lists the analytical requirements that the model must be able to address. These include federal and state planning requirements, equity, greenhouse gas emissions, multi-modal planning, land-use, demographics, reliability, induced demand, and other issues. Section 5 (Next Steps) will discuss these issues again, by detailing how the model addresses each issue. The rest of this section itemizes the purposes or “use cases” to which the model will be applied. Note that the design group proposed to assess the frequency in which these use cases arise as a means of prioritizing model features in the design; at this writing themploye subsection *Flexible and Expandable Design*

ActivitySim is generally a flexible model system, in which it is relatively straightforward to plug in different model components or extend the choice set of existing components to consider new alternatives such as modes or destinations. One must be careful, of course, to ensure that downstream models and summaries of outputs are consistent with whatever changes are made. Nonetheless, the design should anticipate to the extent possible, which components will be made available in the initial deployment, and which features would be available in subsequent components once additional data and/or resources are available. See the section below for more details.

Features Necessary to Support Needs and Requirements in *Section 2.3* contains a table that may be used for this assessment.

Federal Metropolitan Planning Organization Requirements

The model design must meet all federal requirements for Metropolitan Planning Organizations (MPOs). These include:

- Metropolitan Transportation Plan/Long-Range Transportation Plan development and updates, which in turn include:
  - Analysis of existing conditions
  - Forecast of future conditions
  - Systems deficiency analysis
  - Alternatives development & analysis
  - Development and assessment of system performance

- Development and update of the Transportation Improvement Program
- Air quality conformity support (including providing data to MOVES)
- Modal plan support including pedestrian, bicycle, freight and transit
- Direct model support (as opposed to simply providing data for off-model analysis) for active transportation modal planning is desirable, *subject to further discussion about the requisite model features.*
- Equity and environmental justice analysis support
- Congestion Management Process support
- Carbon Reduction Plans
- Safety planning support is desirable, *pending further discussion of the necessary features.*

It is important to note that discussion to date has indicated that internal features to support resiliency planning are low in prioritization for this joint modeling effort.

We also assume that the model must be capable of addressing other federal requirements not specifically assigned to MPOs, including:

- Corridor studies (road and transit)
- Project development analysis (road and transit)
- National Environmental Policy Act (NEPA) studies including:
    - Planning and Environmental Linkages (PEL) studies
    - Environmental Impact Statement (EIS and EA) support
- Bicycle and pedestrian plan modeling

The above discussion lists the federally inspired use cases to which MPO models may be applied. MPOs conduct these activities under general USDOT guidance that asks for due attention to several Planning Emphasis Areas (PEA).[2] The PEAs are overarching goals or topics, several of which encompass the model-related use cases already listed:

- **Tackling the Climate Crisis**
- Embedding **Equity and Justice40** in the planning process

[2] USDOT (memo). "2021 Planning Emphasis Areas for use in the development of Metropolitan and Statewide Planning and Research Work programs." Accessed 7/10/23 at https://www.transit.dot.gov/regulations-and-programs/transportation-planning/2021-planning-emphasis-areas

- Realizing **Complete Streets** outcomes

The model use cases Equity Analysis, Greenhouse Gas Analysis, and Safety Planning Support cover any technical requirements the PEAs would place upon the model design.

USDOT promulgates additional PEAs that are more process oriented. Productive **coordination with the US Department of Defense and Federal Land Management Agencies**, conducting planning activities through a **Planning and Environmental Linkages (PEL)** lens, and **supporting data sharing** are all ways of conducting the planning processes that travel models support and may be relevant to certain model design features (e.g. data storage and visualization).

State and Local Planning Requirements

We anticipate the model to be used for the following types of state and local applications:

- Growth Management Act (GMA) support for Urban Growth Boundary (UGB) decisions
- Transportation System Plan (TSP) update support (e.g., alternatives analysis, system performance measures)
- Property (land use) development assessment and permitting
- Traffic impact studies

These federal, state, and local requirements imply a set of standard metrics that the model must be capable of producing. These include:

- Travel & traveler metrics
    - Accessibility metrics (logsum) by trip purpose and mode
    - Total personal travel (like PMT) and level of time spent traveling (PHT)
    - Total household travel
    - Total vehicle travel (VMT, VHT)
    - Road reliability
    - Cost of travel, both direct (fares, operating costs) and indirect (vehicle ownership, etc.)
    - Mode share
- Health, Safety, and Environment Metrics (from model)
    - Accessibility in terms of the safety or stress of active mode options
    - Physical activity (hours spent being active, hours per capita, hours by traveler)

  - Air Quality (GHGs and criteria pollutants) at the regional level
- Input and output attributes that support off-model impact assessment
  - Noise (e.g. buffering around facilities, speeds, volumes)
  - Water runoff (lane-miles, etc.)
  - Crash numbers and severity (speeds, volumes)
  - Health metrics computed from travel, emissions: DALYS – Disability Adjusted Life Years

Land Use Planning and Analysis

In addition to the GMA and UGB decision support mentioned in the previous section, regional agencies may wish to test the transportation impacts of different land use scenarios for their own planning purposes, and local agencies—cities and counties—will need to test such scenarios for their comprehensive planning processes.

Also, Oregon land use regulations are undergoing changes due to the limits placed on single family zoning restrictions in urban areas and due to new rulemaking related to Climate Friendly and Equitable Communities ("CFEC").[3] These changes, once put in place, will likely increase densities in urban areas. Specifically, climate friendly areas (CFAs) will be designed to accommodate at least 30 percent of existing and future housing units and support greater mixed-use neighborhoods. In regulatory terms, CFEC rules create added responsibilities for GHG and VMT mitigation at the regional and local levels.

Currently the model design assumes that the feature requirements to the GMA, regional, local, and CFEC use cases are uniform. Land use scenario analysis is primarily conducted by varying the *inputs* of household and employment location and characteristics. The model must provide features to facilitate varying these inputs for the modeling region and zones within the region, which implies functionality for the population synthesis tool to be integrated into the modeling process. Additionally, the model must be appropriately sensitive to these inputs. Changing retail employment, for example, should result in a greater change in shopping tours and stops than in work tours. Land use density and mixed use effects should be included in appropriate models (e.g. mode choice) to ensure that densification scenarios are reflected in modal analysis so that the transportation system response to land use scenarios accounts for the built environment, the effects of the distribution and density of residents and jobs, and explicit transportation investments or policies.

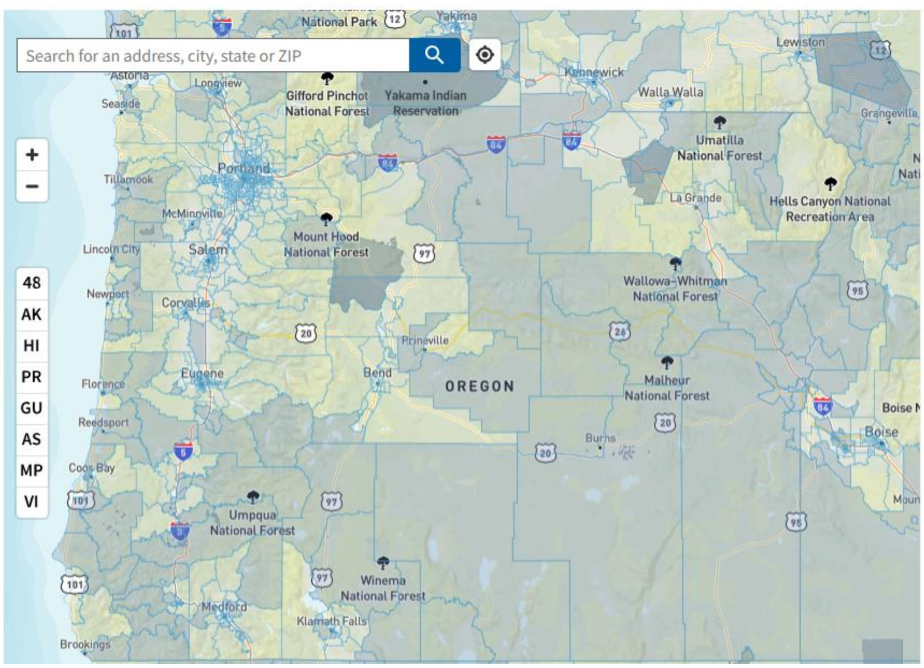
---

[3] CFEC Transportation Planning rules.
https://secure.sos.state.or.us/oard/displayDivisionRules.action?selectedDivision=3062

Equity Analysis

The model must provide the ability to inform equity discussions across as many metrics and demographic segments as possible. Equity analysis considers historical burdens, future benefits, and impacts. The Oregon DOT Office of Social Equity[4] has a roadmap to improve the considerations of equity throughout the day-to-day activities and specifically within the transportation planning and project development process which travel models support and inform. The Oregon efforts are reinforced by actions taken at the federal level, most specifically the Justice40 initiative which aims to direct 40% of the benefits of federal investment to specific disadvantaged communities. The federal government created a GIS based screening tool that uses a variety of data sources to identify disadvantaged communities as shown for Oregon in FigureFigure 1.

FIGURE 1: JUSTICE 40 SCREENING TOOL



Source: Climate and Economic Justice Screening Tool[5]

The Office of Social Equity has prepared an Oregon specific Equity Index aimed at providing a locally relevant and locally derived map on where there are higher relative magnitudes of

[4] Oregon Department of Transportation : What is Social Equity? : Social Equity : State of Oregon (https://www.oregon.gov/odot/equity/pages/about.aspx)
[5] Climate and Economic Justice Screening Tool. https://screeningtool.geoplatform.gov/en/#6.01/43.619/-120.46

burdened populations. The index is based on available census data at the tract level comparing the share of the population with specific characteristics. The index is informed by OAR 660-012-0125 that defines the following groups as underserved populations (but does not limit the analysis to only these groups).

- Black and African American people;
- Indigenous people (including Tribes, American Indian/Alaska Native and Hawaii Native);
- People of Color (including but not limited to Hispanic, Latina/o/x, Asian, Arabic or North African, Middle Eastern, Pacific Islander, and mixed-race or mixed-ethnicity populations);
- Immigrants, including undocumented immigrants and refugees;
- People with limited English proficiency;
- People with disabilities;
- People experiencing homelessness;
- Low-income and low-wealth community members;
- Low- and moderate-income renters and homeowners;
- Single parents;
- Lesbian, gay, bisexual, transgender, queer, intersex, asexual, or two-spirit community members; and
- Youth and seniors.

Equity analysis is helpful in evaluating how certain populations benefit or are impacted by current conditions, including: access to jobs or other opportunities, travel times, or the externalities from air pollution or noise. Equity analysis in the context of future planning considers benefits and burdens from possible future investments or policies. The model will allow specific measures (travel time and costs, mode choice, vehicle ownership, etc.) to be summarized across several socio-economic dimensions. Equity focused analyses can be designed to leverage the strength of the ABM by assessing where disbenefits accrue as well as what is causing those impacts. For example, noise and air pollution are often experienced by individuals who live near busy highways, but a large share of the sources of these impacts are often trips by people traveling through the location, not residents. These are the types of insights the ABM can start to unlock. The use of the ABM will require assessing these questions and determining whether prepared scripts or analyses can be utilized, or whether they are custom to each specific application.

Limitations include the ability to accurately forecast the demographic composition related to residential choices, and the ability to assert characteristics onto the population when data is sparse, for example sexual orientation is not often known at the resolution required for synthetic population. While some of these characteristics exist in today's population, forecasting them in future populations may lead to unintended consequences; as example, by carrying base year demographic assumptions into the future.

Most population synthesis procedures include age, gender, and income as control variables. Some population synthesis procedures also include race and ethnicity, but tend to exclude these variables from the activity-based model; instead they are only used to analyze policy and infrastructure impacts post-model. Note that sub-regional controls for race/ethnicity is desirable since most transportation project and policy impacts are spatially varied. In cases where race and/or ethnicity is used as a control variable, often the distribution of households by race and ethnicity is held constant into the future due to difficulties in predicting these distributions at anything lower than a regional level.

Methods for evaluating equity should include indices by which results can be aggregated to specific polygons or to households and individuals with specific characteristics. Each method has benefits and drawbacks which will need to be considered in each application. The initial design of the model will account for evaluating and segmenting results across the explanatory variables that can be forecast as well as summarizing results by specific polygons derived by third parties (i.e., The Office of Social Equity). The model will be flexible, allowing the user to join spatial layers with the embedded synthetic population data, which will produce a rich set of data to enable later segmentation. The application of the model can develop standard scripts to analyze measures across these segmentation criteria.

Greenhouse Gas Analysis

In 2009, the Oregon Legislature set goals to reduce greenhouse gas (GHG) emissions 10 percent below 1990 levels by 2020 and at least 75 percent below 1990 levels by 2050.[6] More recently, Executive Order 20-04 set new emissions reduction goals that call for the State of Oregon to reduce its GHG emissions at least 45 percent below 1990 emissions levels by 2035 and at least 80 percent below 1990 levels by 2050.[7] These updated goals are consistent with the reductions that climate scientists now believe are necessary to avoid catastrophic climate change impacts. Beginning in 2012 and updated as recently as July 2022, the State has set GHG reduction targets for metro regions across the state. These targets have changed in definition over the years, with the most notable change occurring in 2017 that specified the targets apply to light duty passenger vehicles including household vehicles, lightweight commercial vehicles, and public transit vans and demand response. Further, the targets focus on reducing vehicle miles traveled. As the State has primary responsibility over vehicles and

---

[6] Oregon Department of Environmental Quality, Oregon Greenhouse Gas Emissions, https://www.oregon.gov/deq/aq/programs/Pages/GHG-Oregon-Emissions.aspx
[7] https://www.oregon.gov/gov/Documents/executive_orders/eo_20-04.pdf

fuels sold in Oregon, the ability to reduce VMT is more centered at the regional level. The GHG targets then focus on VMT reduction per capita as the primary measure, however, because overall GHG reduction is the fundamental goal, it is necessary to also account for tailpipe emissions.

Note also Oregon's recent new transportation planning rules (TPRs) related to GHGs discussed above under the *Land Use Planning and Analysis* heading.

In order for the model to be useful for modeling GHG emissions and the effects of policies to reduce them, it must be sensitive to inputs and policies that affect transport demand and be capable of summarizing GHG emissions and energy consumption at a household level. Required sensitivities include policies that affect auto ownership, vehicle technology and fuel type, Travel Demand Management (TDM) policies, pricing (including parking price and supply, managed lanes, tolling, and congestion pricing), policies that encourage transit, the availability of emerging modes such as micromobility, rideshare, and other emerging modes, and potential effects of Intelligent Transportation System (ITS) and safety improvement investments on congestion.

The ActivitySim model software requires input level-of-service matrices (also known as skims) from traffic assignment software, and is agnostic regarding the type of software used to generate those skims. Typically, static equilibrium processes are used to generate auto skims. There are several implementations of ActivitySim (mostly in a research context) that integrate ActivitySim with Dynamic Traffic Assignment (DTA) software. Although the initial version of the Oregon model will utilize skims from static assignment software, the Oregon modeling community should continue to monitor developments in integrating the tool with DTA and consider adopting an integrated model when performance and technology has evolved to a 'best practices' level.

The model design will consider the number of time periods to include in the Oregon deployment of ActivitySim to ensure a reasonable representation of changes in congestion throughout the day. We suggest a minimum of four time periods (A.M. peak, midday, P.M. peak, and evening). Several implementations split the night period into early A.M., which covers the period from 3 AM to the start of the A.M. peak, and evening, which covers the period from the end of the P.M. peak through 2:59 A.M. The Chicago model further splits the A.M. and P.M. peak periods into even smaller one-hour periods.

The Oregon model will produce outputs at the household level which will enable VMT summaries required by the GHG target rules and consistent with VisionEval summaries. VMT generated by commercial vehicles should also be summarized. Inclusion of the ActivitySim vehicle type model would enable VMT, operating cost, and energy consumption levels that consider private vehicle fleet composition. The model will also produce link-level volumes by class and link-level congested speeds which can be used in emissions models such as MOVES, or used in combination with Oregon DEQ emissions rates by county to generate GHG emissions without resorting to MOVES.

GHG emissions can be aggregated by zone, link type, or other network characteristic if assignment classes are stratified by vehicle type or can be post-processed using path information from a previous assignment. Multiple years or interpolation of specific years can provide the opportunity to calculate cumulative emissions of GHG between scenarios. This type of post-process can be developed after the core model design phase.

Bicycle and Pedestrian Planning and Analysis

Bike and pedestrian travel have been important in urbanized areas such as Portland and Eugene, and are increasingly a focus in less urbanized areas as well. Increased ownership of electric bikes and provision of (electric) bike share systems are likely to lead to increased bicycle mode share in the coming years.

At a minimum, model users in Oregon will want to be able to look at scenarios defined in terms of the bicycle infrastructure network—the type of bicycle facilities that are provided on each link of the road network, and the impacts of sharing space with motor vehicle traffic on their propensity to bike. Other network attributes could be coded as well. These attributes could include:

- Other types of infrastructure configurations such as conventional and protected bike lanes, bike boulevards, or multi-use paths
- Road classification in terms of the typical intensity of vehicle traffic flow
- Slope and/or elevation gain along each link
- Indicators of surface type and/or quality
- Intersection treatments, including signals and mode-specific crossing aids such as special turn pockets for cyclists
- Allowance for cyclists to travel the wrong way against traffic (“contraflow lanes”)
- Identification of specific locations that cyclists may find particularly dangerous, such as near freeway entrances and exits or bridges that do not have adequate bike facilities.

The modular nature of the ActivitySim model framework can provide a number of options for the network assignment of bicycle traffic, and the treatment of bicycle as a mode in mode choice. Some options include:

- Using the estimated utility function from an existing bicycle route choice model (such as the utility function estimated by Broach, et al. at PSU) to use a “perceived” distance or time function rather than simple distance or time. This can be used to determine and skim the best bicycle path in the network.

- Fully applying an existing bike route choice model (including path size functions for overlapping paths). This option, which was carried out by Oregon Metro, for example, is more computation intensive than the preceding option, and requires custom programming outside of a network package such as EMME or VISUM. Note that this functionality already exists in the Oregon Metro trip-based model.
- Specifying a bike network at a finer level of spatial detail than the auto network. Several ActivitySim models use microzones (typically the size of Census blocks), as well as the typical TAZ system used to define the auto network. A bike network can be defined at the microzone-to-microzone level, and can include some local streets that are not in the modeled auto network. ActivitySim will read more detailed level-of-service information between close-together microzones and can optionally blend microzone level of service with TAZ level of service to help eliminate any potential cliff effects associated with mixing data from different geographic scales. This method, when coupled with a bike route logsum, is one of the most advanced methods for representation of bike lane investments on mode and route choice. It is currently used by SANDAG.
- Treating electric bikes and standard bikes (and possibly bike share) as separate modes.
- Using land use data and street connectivity data at the microzone level in the mode choice models.

One of the key considerations for model design is how to treat heterogeneity in the population in terms of bicycle comfort level. Mineta Transportation Institute proposed a scheme for classifying road segments by one of four levels of traffic stress[8]. Level of traffic stress 1 (LTS 1) is meant to be a level that most children can tolerate; LTS 2, the level that will be tolerated by the mainstream adult population; LTS 3, the level tolerated by American cyclists who are “enthused and confident” but still prefer having their own dedicated space for riding; and LTS 4, a level tolerated only by those characterized as “strong and fearless.” They propose a set of criteria by which roads segments are classified by each LTS dependent on traffic characteristics and presence of bike lanes.

That criteria has been significantly expanded upon in Version 2 of the Oregon DOT Applications Procedures Manual[9]. The manual provides lookup tables for BLTS based upon the class of bicycle facility: physically separated paths and lanes, standard bike lanes, and without bike lanes (mixed traffic). Physically separate lanes typically have a BLTS of 1 (best). Marked bike

[8] Maaza C. Mekuria, Peter G. Furth, Hilary Nixon, Report 11-19
Low-Stress Bicycling And Network Connectivity, Mineta Transportation Institute, May 2012.
[9] Analysis Procedures Manual Version 2, Policy, Data & Analysis Division, Planning Section, Transportation Planning Analysis Unit, Salem, Oregon, April 2023.

lanes have different criteria depending on whether or not they are adjacent to a parking lane and level of stress also depends on bike lane width and prevailing traffic speed. BLTS for mixed flow conditions depends upon prevailing speed, average daily traffic, and number of lanes. Many of the variables can be derived from a combination of coded bicycle facilities on an all-streets network and travel demand model results (prevailing speed and ADT); however, some variables proposed in the APM, such as lane width and pavement condition (used as a modifier on BLTS calculations from other variables), are less likely to be available.,

Roger Geller[10] categorizes bicycle riders into three categories: “The Strong and the Fearless,” “The Enthused and the Confident,” “The Interested but Concerned,” with a fourth group of non-riders, the “No Way No How” group. These categorizations were a useful tool for the City of Portland to understand the target market for bicycle infrastructure improvements and what their concerns might be. Initial estimates of the population size of each market segment were the guesswork of a city planner; since then, numerous studies and reports have sought to quantify market size in many urban areas in the U.S., and recent research attempts to statistically estimate class membership. There are opportunities for the Oregon ActivitySim model to push state of the practice in bicycle modeling forward by attempting to model bicycle comfort level class membership of each individual in the synthetic population and associate the persons LTS level with a path consistent with their class. This would provide the ability to test changes in attitudes towards bicycling in future scenarios. It would require a targeted survey to identify class membership and provide a foundation for model development.

Other, more typical segmentation variables for bicycle route choice logsums used in travel models include purpose (Oregon Metro’s logsums are segmented by commute/non-commute), and person attributes such as age and/or gender.

Most of the options above, such as using a perceived distance function and using a more detailed microzone-to-microzone network, could be used for the walk mode as well as the bike model. Land use and street connectivity measures are very important for both the bike and walk modes, particularly in mode choice models. Such measures can include:

- The density of street intersections, possibly with different weight given to 4+ way and 3-way intersections, and a negative weight given to dead ends/cul-de-sacs
- Residential and employment density measures
- Mixed-use measures, such as using an entropy-type formulation
- The existence and extent of streets that are closed to motor vehicle traffic
- The existence and extent of traffic calming and/or pedestrian safety infrastructure

[10] https://nacto.org/references/four-types-of-cyclists/, references August 23, 2023.

These types of variables along specific links could also be used in path choice functions, but they are more typically used in mode choice models.

The types of outputs from the model will include, at a minimum, the number of bike and walk trips and the distance traveled. The latter are particularly important for measuring physical activity from walking and biking, which can be an input to health impacts models such as ITHIM. The flows on the links in the bicycle network are typically a desired model output as well. This requires assignment of flows based on the same or very similar path-finding method to that used to generate skims for mode choice. These assigned flows can be validated against bicycle count data, where available.

Congested (multi-iteration) bicycle assignment is not necessary, as there are currently no US cities where bike lanes or bike paths are so congested that cyclists will choose paths with fewer bikes. In fact, the “safety in numbers” phenomenon is likely to attract cyclists to use paths that other cyclists are using. Bicycle path choice can be influenced by the level of auto traffic on the roads, but classification of roads based on typical AADT can be used, so that the bicycle assignment does not need to be repeated each time the auto network assignment is carried out.

A detailed walk network (including off-street paths) would need to be coded and maintained if walk trips are to be assigned to the network, and they should be validated against a set of pedestrian counts with significant coverage. An alternative approach would be to measure walk demand at the zone or microzone level and couple that data with other data such as VMT and average speeds to measure potential safety impacts.

We should also point out that existing model implementations are restricted to representing utilitarian travel; that is, travel to activities and not for general exercise and sightseeing, which are often referred to as ‘loop’ trips in household surveys. That has implications for comparison of model assignment results for active modes to bike and pedestrian counts. If there is interest in modeling these types of trips, we would need to add an activity purpose to ActivitySim so that they can be represented.

Changing Demographics

Oregon’s demographic profile is constantly evolving due to changes in birth rates, mortality, and migration. Oregon’s estimated population reached 4.195 million on July 1, 2018 and is expected to grow to approximately 4.744 million by 2030[11], though this forecast may need to be discounted somewhat due to very slow population growth in recent years[12]. While certain age groups are projected to increase very little (18-24 and 45-64 year olds), other age groups are

[11] Oregon’s Demographic Trends, Office of Economic Analysis, Department of Administrative Services, State of Oregon, July 2019

[12] According to census estimates, Oregon’s July 2022 population was 4,240,137, which amounts to 0.1% growth since 2020 (https://www.census.gov/quickfacts/fact/table/OR/PST045222, accessed July 27, 2023).

expected to increase more significantly (25-44, and 65+). These changes will result in an average increase in the median population age, from 39.7 to 41.6, with more population particularly in the 75+ age category. These changes are likely to increase travel for medical appointments and recreation compared to the current population profile, and also have implications for household income (and effects on travel), auto ownership, demand for transit, willingness to pay for priced infrastructure, and other transport policies and investments. Oregon's future models must be sensitive to demographics in order to reflect these changes.

Although Oregon is one of the least racially diverse states in the country, Oregon's population is becoming increasingly diverse in terms of race and ethnicity, especially due to in-migration. Economic indicators are less clear; the recently released Oregon Economic and Revenue Forecast[13] forecasts shrinking labor force participation rates (due to increasing population age) but higher productivity per worker. Whether this translates into real increases in wages remains to be seen.

These trends have implications for population synthesis procedures and travel model variables and coefficients. First, the synthetic population should include standard controls by age, household size, workers per household, household income, and other important variables. Second, population controls and explanatory variables should be sufficiently detailed to represent variation in age, particularly in the 65 and older category, as we expect an increase in the share of octogenarians specifically as the boomer population moves through the population pyramid. Age effects in the model should consider splitting the 65+ age category into 65-80 and 80+ in order to better capture age-related differences in mobility. And third, we should be careful about age-related effects in the model, as technology adoption rates by age will likely not be constant into the future, and increases in medical technology and services may result in increased mobility for senior citizens in the future.

We also note that with an aging population, household income is a less reliable measure of household wealth. The Census bureau notes[14] that respondents report income from wages and salary much more reliably than other sources, and that household income is exclusive of capital gains. Furthermore, as persons age, they may be more likely to have paid off their mortgage, leaving more money for other expenses. Income is an important explanatory variable, so we expect to include it, but we may want to apply a model to predict adjusted gross income from variables that are available in the synthetic population.

Current implementations of ActivitySim assume that anyone age 16 or over can choose to 'drive alone'. However, the actual share of license holding is lower than 100%, perhaps much less so for persons over age 80. And in the case of autonomous vehicles, holding a license may not be

[13] Oregon Economic and Revenue Forecasting, May 2023, Oregon Office of Economic Analysis, Volume XLIII, No. 2 Release Date: May 17th, 2023.

[14] See https://www.census.gov/topics/income-poverty/income/about.html, accessed July 10, 2023.

a requirement for being a single occupant vehicle. We may want to explicitly predict mobility status and/or driver's license holding to more accurately model driving behavior differences by age. The Census asks a series of disability status questions that could be used to estimate such a model; however it does not ask license status. The household survey does ask these questions, so a model would need to be estimated or a probability distribution derived and applied to the synthetic population. It could be calibrated to published data from the Oregon Department of Motor Vehicles.

Pricing

Priced infrastructure is a particularly important issue in Oregon, particularly in the Portland region in which tolls are being considered on a number of facilities. However, the treatment of prices and costs in travel models extends to parking costs, auto operating (and ownership) costs, transit passes and fares, shared micro-mobility modes (e-bikes and scooters) and ride-hail modes.

The treatment of pricing in travel demand models has been the focus of many research projects and practical efforts over the past 20 years. One of the most seminal recent works in this area is the work funded under the Strategic Highway Research Program (SHRP) C04 track on pricing and reliability[15]. The final report recommended a number of key features:

- Travel time heterogeneity: Sensitivity to travel time should be represented as a distribution reflecting personal preference and contextual conditions
- Continuous representation of income: Sensitivity to cost should ideally be represented as a continuous function of income rather than a global average
- Vehicle occupancy effects: Sensitivity to cost is also dependent upon the occupancy of the vehicle, but it is not a linear relationship.

[15] Parsons Brinkerhoff, Northwestern University, Mark Bradley, University of California at Irvine, RSG, University of Texas at Austin, Frank Koppelman, GeoStats, SHRP2 C04 - Improving Our Understanding of How Highway Congestion and Pricing Affect Travel Demand, Transportation Research Board of the National Academies (TRB), The Second Strategic Highway Research Program (SHRP2) Capacity Research, April 2013

If agencies wish to invest research funds into specifying a different time and cost function, additional variables (such as presence of children) could be used to influence travel time sensitivity and willingness to pay.

Recommendations in this report were implemented in the SANDAG model system and calibrated against observed demand in both the SR-125 toll road and the I-15 high occupancy toll lane[16]. The key features of the model include:

- Each traveler has their travel time sensitivity sampled from a log-normal distribution based on an analysis of several stated preference surveys conducted for toll road pricing studies
- The cost function recommended in the SHRP C04 report was implemented in the activity-based model mode choice structure. All cost variables regardless of mode are treated the same in utility equations.
- Elimination of toll/non-toll choice as separate vehicle classes in assignment; instead, value of time segmentation (three classes) is used in auto skimming and assignment to represent different path options based on the value of time of the traveler. The auto skims are 'mapped' to each traveler given their value-of-time, which is derived from their travel time coefficient, travel time sensitivity factor, and cost coefficient, as described above.

The model implementation will need to explicitly track and write out all costs paid as part of a trip, including auto operating costs, parking costs, tolls, transit fare, ride-hail costs, micromobility costs, and any other relevant cost information. Partial and/or full subsidies and rates also must be written to outputs.

Reliability Analysis and Reporting

Travel time reliability reflects the amount of variation in travel time, which can affect various dimensions of travel behavior, including mode, destination, time-of-day, and route. Therefore, travel time reliability (referred to throughout this section as reliability) is increasingly being used in travel models as an explanatory variable affecting model outcomes as well as a system performance measure.

There are number of ways to measure reliability, but there are significant limitations in its measurement within travel demand models, particularly ones which rely upon static equilibrium

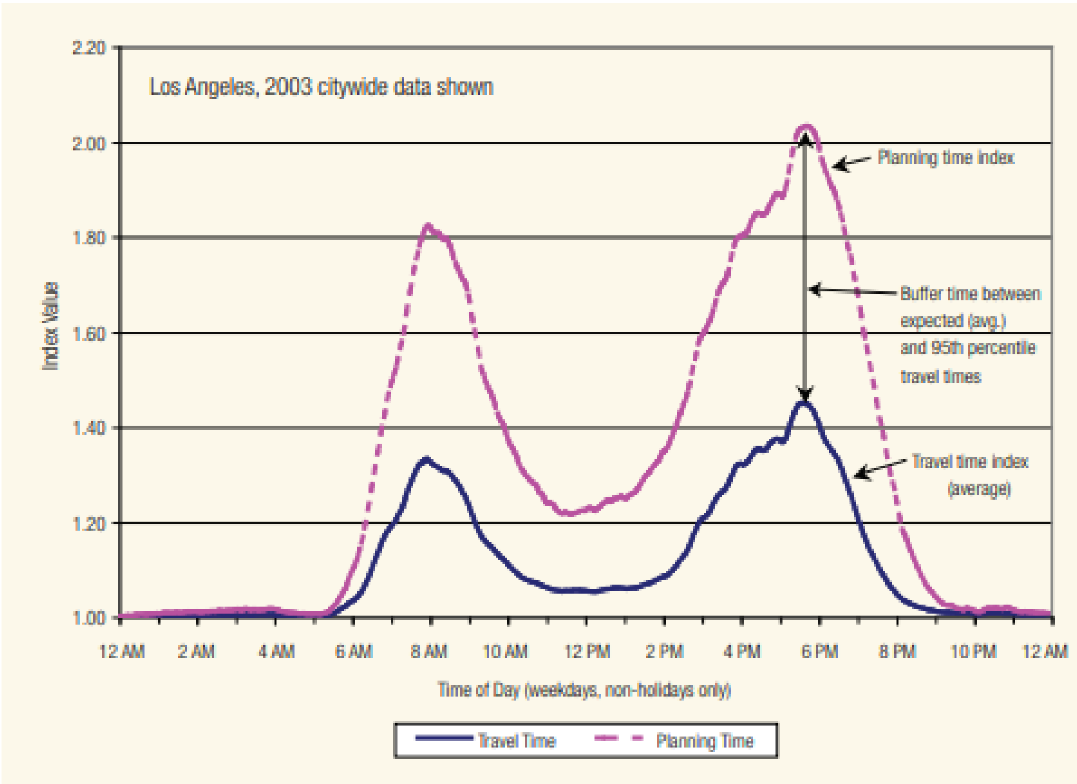
[16] RSG, Pricing and Travel Time Reliability Enhancements in the SANDAG Activity-Based Travel Model: Final Report, San Diego Association of Governments, June 2016.

highway assignment methods. Below is a brief list of reliability measures, which are shown graphically in Figure 2.

- *90th or 95th percentile travel times* for specific travel routes or trips, which indicates how bad delay will be on the heaviest travel days.
- *Buffer index*: the extra buffer time (or time cushion) that most travelers add to their average travel time when planning trips to ensure on-time arrival. This extra time is added to account for any unexpected delay.
- *Planning time index*: Total travel time that should be planned when an adequate buffer time is included that includes typical delay and unexpected delay.
- *Travel time index*: The ratio of the travel time during the peak period to the time required to make the same trip at free-flow speeds. This measure is consistent with other methods Oregon plans to use and easy to compute from travel models. However, although travel time reliability is related to congestion, the travel time index measures only congested travel time rather than the variability of travel time.
- *Standard deviation of travel time*: Measures the distribution or spread of travel time (see Figure 3)

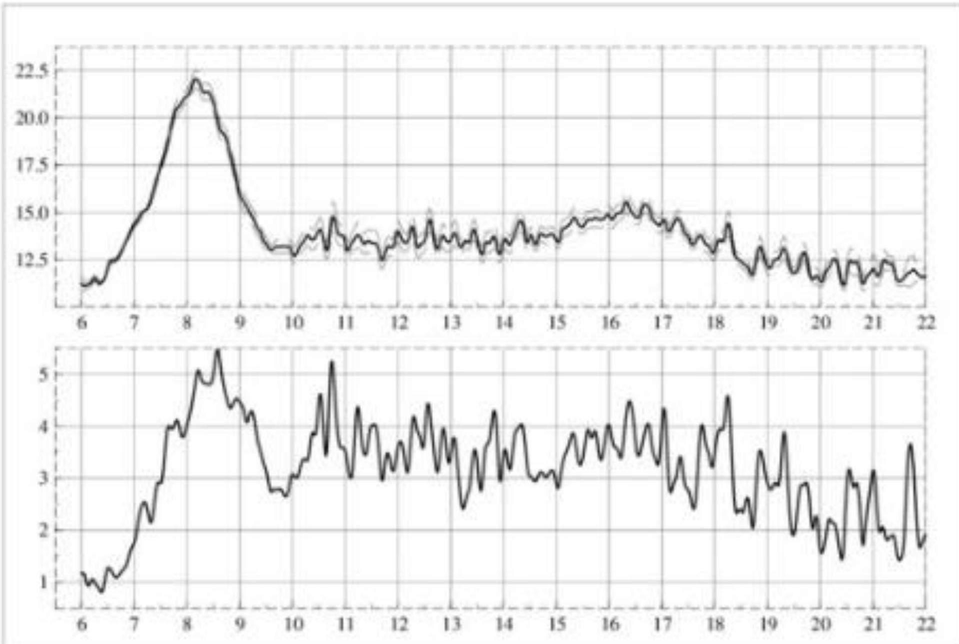
ODOT's metric evaluation that occurred alongside the model design work identified that using the Travel Time Index (TTI) would best align with methods and metrics that the state uses in other planning and monitoring context.

FIGURE 2: TRAVEL TIME RELIABILITY AS A FUNCTION OF TRAVEL TIME



Source: Mobility Monitoring Program, http://mobility.tamu.edu/mmp/.

FIGURE 3: STANDARD DEVIATION OF TRAVEL TIME



Fosgerau, M. and Karlström, A., The Value of Reliability, Transportation Research Part B Methodological, December 2007

There are many potential causes of travel time unreliability. Typically, highway congestion is classified as 'recurring' and 'non-recurring'. Recurring congestion occurs because of capacity constraints coupled with high demand. Non-recurring congestion occurs due to vehicle crashes, breakdowns, construction, and other random events. These give rise to travel time variation or unreliability. It should be noted, however, that travel time reliability is related to trip length, because the probability of random events such as accidents occurring increases with respect to distance. The line between recurring and non-recurring congestion is therefore somewhat fuzzy.

Transit travel time can also be unreliable. This is due to unreliability in roadway network times (for non-grade separated transit), fluctuations in demand which can cause increased dwell times, and variations in schedule adherence. Aging infrastructure, particularly for rail transit, can cause maintenance failures and schedule unreliability. Transit routes with very high demand can also result in travel time variation; for example, when buses or trains arrive to a stop full, the rider must wait for the next vehicle with capacity, adding travel time equal to the vehicle headway.

Travel models have a number of limitations which lead to challenges in the representation of auto or transit reliability:

- The network representation is typically limited to average conditions. In other words, they do not typically represent accidents or road construction.

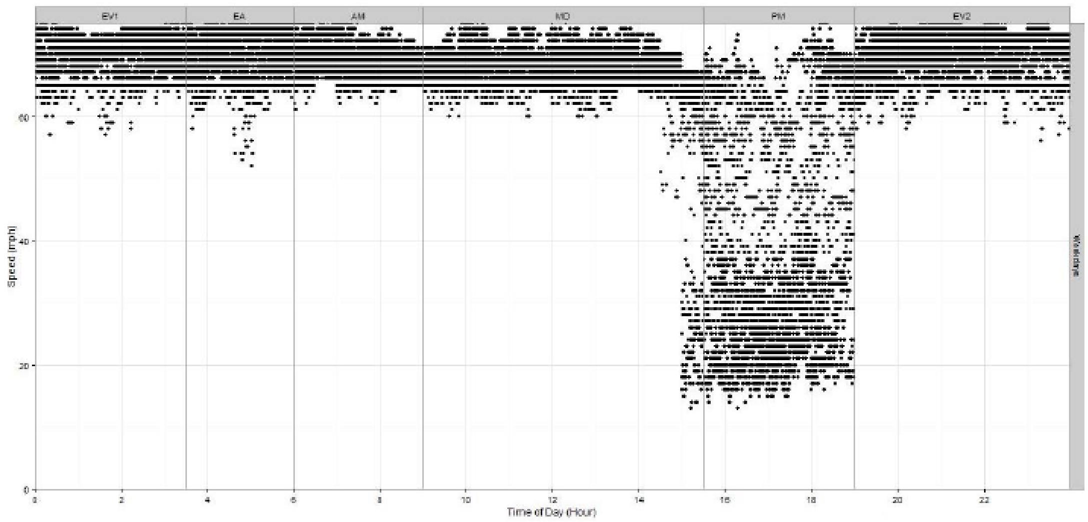
- Static equilibrium models do not represent queuing and therefore cannot address unreliability due to complicated weaving movements or the increased likelihood of accidents due to those conditions.
- Static equilibrium models assume that link characteristics are independent of each other. The variance of travel time is not independent across links. Other measures, such as the standard deviation of travel time, are not additive across links, which is inconsistent with the way in which assignment models calculate time and cost along a path through the network.
- Transit assignment models are typically not set up to handle transit vehicle capacity or the interaction of demand and supply in transit assignments. These models also require iteration which can increase runtime.

Due to these challenges, many models ignore auto and/or transit reliability. Some models (SANDAG, MAG, MTC, Toronto) represent reliability explicitly. Some other models post process results. We suggest explicit representation if measurement of reliability is important to the Oregon modeling community. Real world travel time information (including both highway travel time and transit schedule adherence data) is increasingly available at sufficient resolution to explore travel time variability for road segments and transit routes across Oregon. That data can be used to estimate equations that predict travel time reliability based upon modeled attributes, the results of which can be incorporated into travel demand models.

An example of such an approach is the SANDAG SHRP2 C04 implementation project[17]. In this project, RSG analyzed INRIX travel time data (see Figure 4) and developed regression equations to represent the relationships between known network attributes and modeled volume/capacity ratios and the standard deviation of travel time. These regressions were then implemented in the travel demand model. The relationship between volume/capacity ratio and deviation in time was represented explicitly in the model volume-delay function, which includes both mid-link and intersection delay terms. The variance in travel time is calculated for each link and summed across links for each OD pair. The standard deviation is calculated from the variance and used as an explanatory variable in mode choice, which also affects upstream models.

[17] RSG, Pricing and Travel Time Reliability Enhancements in the SANDAG Activity-Based Travel Model: Final Report, for San Diego Association of Governments, June 30, 2016

FIGURE 4: TRAVEL TIME DATA FOR A NETWORK LINK IN SAN DIEGO



1(b) – Speed Profile without Outliers

Source: RSG, Pricing and Travel Time Reliability Enhancements in the SANDAG Activity-Based Travel Model: Final Report, for San Diego Association of Governments, June 30, 2016

FIGURE 5: SANDAG VOLUME-DELAY FUNCTION WITH UNRELIABILITY COMPONENT

$$T_{f+r} = T_f + T_f \ * \Big[\underbrace{\sum_{t=1,n} (\gamma_t * \frac{v}{c} - t + 0.01)}_{\text{LOS (un)reliability}} + \underbrace{R}_{\text{Static (un)reliability}}\Big]$$

Where:

T_{f+r} = Travel time with (un)reliability

T_f = Travel time without (un)reliability

t = v/c thresholds (C, D, E, F-low, F-high)

γ_t = Coefficients for v\c thresholds

R = non-v\c link (un) reliability

Special Market Travel

Most travel demand models represent an average weekday. Weekend and seasonal travel variations, though common in many urban areas as well as state facilities, are typically not explicitly addressed due to increased costs of developing and maintaining such models. These costs include collection of seasonal and weekend travel behavior data, unique special generators that drive demand, and validation data such as traffic counts and transit boardings. Furthermore, there is often significant variation in weekend travel days; for example, Saturday morning travel patterns can differ significantly from Sunday afternoons in terms of travel purpose, origin-destination, and peaking characteristics. A common approach to represent non-typical conditions is to pivot forecasts off of average weekday conditions using adjustment factors developed from base-year data. Average weekday conditions can be annualized in much the same way.

This approach suffers from obvious shortcomings. Weekend and seasonal conditions cannot be represented explicitly, and therefore models are not well-suited to test policies and projects whose impacts disproportionately affect weekend or off-season travel periods. However, some models do explicitly represent weekend conditions (New Jersey, San Francisco County Transportation Authority); therefore, it is certainly possible to do so given the resources. Active mode use can also be seasonal with more walk and bike trips typically occurring during more milder weather. One positive note is that the Oregon household travel survey will have some weekend travel days represented in the rMove smartphone mobile app data. However, summer months will be avoided in the survey.

Overnight visitors are a key travel market in certain regions, parts of regions, and/or facilities. An overnight visitor model would explicitly address travel for visitors. The model would generate a synthetic population for visitors, attribute the population with key characteristics (lodging zone, party size, purpose of visit, availability of a rental car, etc.) and model their travel patterns. We recommend a tour-based model so that visitor trip lists would contain similar information as residents and have similar responsiveness to input land-use data and skims. RSG recently implemented the SANDAG overnight visitor in ActivitySim which could be transferred and/or modified for use in Oregon. Alternatively, a trip-based model treatment can be used to address overnight visitor travel.

Students of major universities often have unique travel patterns and regions with major universities often need to analyze unique transport projects and policies such as inter-campus shuttles and parking demand. Although ActivitySim models university student travel, there are some simplifications in the core software which limits the ability of ActivitySim to model intra-campus trips when the campus spans multiple zones, and has implications for the ability of the model to represent walking, biking and transit trips. The ActivitySim model developed for Southeast Michigan Council of Governments (SEMCOG) enhanced ActivitySim to represent multiple potential school zones for students of major universities and introduced a new parking lot model for university travel which improves the ability of the model to better account for student travel patterns. We note that university students living in group quarters would be explicitly represented in the group quarters synthetic population, but modeling their travel patterns accurately is dependent on a travel survey which includes this population.

Regions with major airports may need a model that can address transport investments that provide access to the airport, airport parking and rental car policies, and analysis of scenarios involving ride-hailing services and other related investments. Airports also have unique travel characteristics with respect to origin/destination patterns and trips by time of day that can make model validation on facilities that serve airports challenging without a special treatment. An airport ground access model can be used to address these needs. The model would generate airport visitor parties, attribute them with important variables (party size, length of stay, visitor/resident status, purpose of travel, etc.) and model their origin (for departing passengers) or destination (for arriving passengers), time of travel, and mode. The airport model can address parking location and modes that are specific to airport travel that are typically not options for other travel, such as rental cars and hotel shuttles. RSG recently implemented the SANDAG airport ground access model in ActivitySim which could be transferred and/or modified for use in Oregon. Alternatively, a trip-based model treatment can be used to address airport travel.

Demand Response to Supply

The term “Latent Demand” is often used to describe demand that exists but has been suppressed due to capacity constraints, congestion, or lack of service or infrastructure to meet the potential demand. The term “Induced Demand” typically refers to demand that is met (in terms of travel) due to improvements in accessibility. However, both terms are referring to the same phenomenon, which is (as research shows) that demand increases with respect to supply; in other words, there is a net increase in travel (trip destinations, walk and bike trips, transit riders, traffic counts, vehicle miles of travel, etc.) depending on the specific changes in infrastructure, land use, or other factors. In a sense, latent demand becomes induced demand when travel conditions improve and causes an increase in travel using a specific mode at a specific time and place. A common example is increased peak hour vehicle flow on a (previously) congested highway link after a lane is added. Induced demand can be revealed by several different types of shifts in travel choices:

- Changes in route choice
- Changes in departure time
- Changes in destination choice
- Changes in mode choice
- Changes in tour frequency and trip-chaining behavior
- Changes in the frequency of performing various out-of-home activities

“Suppressed demand” (leading to an increase in latent demand), is the opposite of induced demand, and can also be caused by all of the types of changes listed above.

Activity-based modeling platforms such as ActivitySim are designed to simulate all of the types of changes in travel choices above, and sensitivity tests are often carried out to compare the relative magnitude of each of the changes when comparing two scenarios—typically a “build” scenario to a base or “no-build” scenario. The table below shows an example of elasticities derived from the SACSIM model in Sacramento, which show slight shifts in total person trips, as well as more substantial shifts in mode choice and distance traveled. Such tests can also isolate specific periods of the day and/or network locations.

Example of sensitivity tests for various types of behavioral responses

(Source: Table 4 in “*SACSIM: An applied activity-based model system with fine-level spatial and temporal resolution*”, Bradley, Bowman and Griesenbeck. 2014. Journal of Choice Modelling, 3(1), pp. 5-31.)

Test variable	Transit fare	Auto fuel cost	Highway capacity	Household income
Response variable:	*Elasticity*	*Elasticity*	*Elasticity*	*Elasticity*
Total person trips	-0.001	-0.010	+0.012	+0.119
Vehicle trips	+0.004	-0.036	+0.021	+0.151
Vehicle miles traveled	+0.006	-0.126	+0.144	+0.090
Transit trips	-0.226	0.151	-0.035	-0.415
Walk and bike trips	+0.005	0.067	-0.055	-0.091

Activity-based modeling platforms such as ActivitySim also include random number synchronization, which ensures that the same random number sequence is applied for specific synthetic households and persons for specific choice models under two scenarios (as long as the synthetic population remains the same for the two scenarios. The key benefit from this synchronization is that it greatly reduces the effect of random simulation variation when comparing the results across two scenarios.

2.2 KEY QUESTIONS

This section contains a list of model application scenarios that will be used to illustrate how the model will be used in application; what inputs to change, and what outputs to summarize, for various model applications. These will be referred to in subsequent sections on model inputs and outputs and include:

- Testing a congestion pricing scenario on key facilities where all lanes are tolled to manage congestion. These facilities include bridges such that the tolls are not avoidable, meaning all travelers are expected to pay the toll. These tolls will likely also toll different travelers different rates based on household characteristics. In previous versions of CT-RAMP and ActivitySim, tolling all facilities that provide access to a specific zone could result in a model crash depending on how the toll is coded. The model needs to be able to find valid paths in all cases where toll policies are levied. In addition, the model should be flexible in terms of varying the representation of the toll paid by different traveler segments. This can be challenging to define in advance of the specific policies and segments to be tolled. The model design should address potential ways to modify configuration files so that these policies can be represented.

- Testing an all-road use fee (VMT tax) by time-of-day pricing scenario. With the advent of autonomous vehicles, it may be possible to vary mileage taxes by time of day. Such policies could be modeled in ActivitySim by changes to the mode choice configuration files where fees are represented.
- Testing the ability to represent high occupancy toll ("HOT") lanes. This implies the ability to skim and assign auto classes by vehicle occupancy and to represent differential pricing by occupancy in network procedures.
- Testing car free downtown zones or car limited downtowns. No vehicles are allowed to enter a central business district or employment campus that is several zones in size. Or potentially parking capacity constrained areas where zoning laws are changed for certain districts such that buildings no longer need to supply enough parking for residence and customers. This implies the need for a parking location choice model and ensuring that parking cost and quantity are considered in mode choice decisions and influence upstream decisions through mode choice logsums.
- Testing how light weight electric vehicles (bikes and scooters) might increase bike and walk speeds for users who choose electric assist. This implies the need to represent micromobility ownership and/or shared micromobility mode availability and cost explicitly in the model system.

We note that some of these key questions imply specific network features that may be outside the purview of the joint estimation project but the design document should identify these needs so that agency partners can consider their implementation. The key questions also imply the development of a scenario builder tool where users can more easily represent inputs to the model system which would then be used to modify and replace parameters contained in multiple configuration files without needing to resort to changing each individual parameter manually. This is addressed in Section 5: Design Implementation Specifications.

2.3 DESIRED PRINCIPLES AND FEATURES

Guiding Principles

A guiding set of considerations for model design include the need to balance complexity with computational requirements, the need for flexibility, and integration with other analytical processes. This section will address details of these guiding principles.

Balancing analytic capabilities against run time and computing resources.

The need to keep the model runtimes fast and computational resources reasonable will be a key aspect of the model design. Models designed without regard to complexity, ease-of-use, or runtime, ultimately become abandoned. The new models (and the network and other data processing procedures required to support them) must run in a time that is commensurate with existing trip-based models, while taking advantage of the improved processing power and memory availability of current hardware. Moreover, the model must be maintainable – that is, the model design must take into account current and future data availability. Features will not be added to the system unless there is data upon which to estimate, calibrate and/or validate them.

Partial Model Run Mode(s)

Certain policies and scenarios may only be expected to change a subset of travel choices; for example, a single intersection improvement may affect route choice but is unlikely to affect mode choice or any other upstream models. A small bus service frequency may only be expected to change mode choice and route choice, but is unlikely to affect tour generation. Therefore, we anticipate the analyst's need for flexibility in turning on or off various model components. At a minimum, we expect the design will accommodate running on traffic and/or transit assignment with fixed trip tables. Beyond that, the ActivitySim is very flexible in terms of turning on or off various model components, by changing the 'pipeline' used to run the model. In fact, RSG recently took advantage of this functionality when implementing the disaggregate accessibility calculations, where only tour mode and destination choice models are run[18]. For the Central Subway Project, a version of the San Francisco County Transportation Authority's activity-based model was used to analyze project benefits for the Federal Transit Administration's New Starts (now known as Capital Grants) application. In this version, all travel dimensions were held constant from the baseline alternative except tour mode choice and downstream models, to ensure that only benefits accruing to improvements in mode choice utility were being taken into account[19]. Another example of selectively re-running alternatives is the MTC Travel Model Two system, in which mode and time-of-day choice models are re-run if a traveler chooses an over-capacity parking lot, after removing the lot from the choice set[20].

Standard Diagnostic Tools

The model implementation should include standard diagnostic tools that can be executed with every model run. The diagnostic tools should include both a step that checks the quality of the

[18] Link to or cite disaggregate accessibilities paper when available
[19] Freedman, J., Castiglione, J., & Charlton, B. (2006). Analysis of New Starts Project by Using Tour-Based Model of San Francisco, California. Transportation Research Record, 1981(1), 24–33. https://doi.org/10.1177/0361198106198100105
[20] Link to or cite disaggregate accessibilities paper when available

model inputs and a check that assesses the reasonableness of the model outputs; flagging and warning the model user of any issues. An input checker is currently under development as part of the ActivitySim consortium Phase 8 scope of work. However, we expect customization of input checks to be required for specific implementations given differences in land-use, synthetic population, networks, and level-of-service data between regions. Currently an output checker is not available in the standard ActivitySim code base. Examples of output checks include very long tour and trip distances or travel times, very long link travel times, and/or very significant changes in model outcomes compared to a baseline set of indicators. Other bespoke checks may be added based on specific inputs and outputs, such as average tour and trip rates, mode shares, etc. We have found that automating a visualizer tool which always runs, and compares the outputs from a model run to either survey data or a baseline scenario, is an invaluable tool for performing diagnostic checks and troubleshooting results.

Post-Processor Linkages

Travel models used by Oregon's planning agencies are used to inform a number of other modeling and post-processing tools. These include:

- MOVES, used for air quality analysis
- The Transportation Modeling Improvement Program's Exploratory Modeling Analysis Tool ("TMIP-EMAT")
- Dynamic traffic assignment (DTA) software such as Dynameq and TransModeler
- Traffic microsimulation software such as Paramics and VISSIM

The model must be able to continue to provide linkages to these other tools, and if possible improve the fidelity of the information being provided. For example, a more refined temporal resolution or more explicit ridesharing data may provide better inputs to DTA software. The design will consider the consistency and quality of data interchange between these platforms.

FTA Guidance on Travel Model Parameters and Use

The Federal Transit Administration's capital grants program (also known as the New Starts and Small Starts program) relies on travel forecasts prepared by sponsors of proposed projects. In its reviews to ensure their usefulness in project evaluation, FTA considers five aspects of the forecasts[21]:

[21] https://www.transit.dot.gov/funding/grant-programs/capital-investments/travel-forecasts, accessed June 20, 2023

1. **The properties of the forecasting methods.** In the past FTA conducted training on multimodal forecasting and shared some lessons learned from their review of travel demand models. A few key takeaways from their findings:
    a. Mode choice parameters should be in reasonable ranges. For example, out-of-vehicle time parameters should be between 1.5 and 3.0 times in-vehicle time for reasonable model elasticities to changes in transit accessibility and frequency of service.
    b. Transit in-vehicle time should be reasonably close to auto in-vehicle time, or generic coefficients should be used.
    c. Alternative-specific constants, particularly for different transit technologies, should be reasonably related to the non-included attributes of those technologies.
    d. Finally, one should be aware of potential ‘cliff effects’ in the model, such as constants that only apply to certain ranges of variables, as these can cause very high elasticities to the attributes of certain alternatives depending on how they are configured.
2. **The adequacy of current ridership data to support useful tests of the methods.** This is a general rule that can be applied to any model component or forecast; there should be adequate observed data to estimate, calibrate, and or validate the model component and/or check the reasonableness of model elasticities to its inputs.
3. **The successful testing of the methods to demonstrate their grasp of current ridership.** To put it another way, the credibility of the model is established by demonstrating that it reasonably depicts baseline travel conditions. This is known as the process of validation, which often leads back to further calibration or even estimation of certain model components. Model development schedules and resources need to account for sufficient effort spent on this important step.
4. **The reasonableness of inputs (demographics, service changes) used in the forecasts.** See above discussion on ‘Input Checker’.
5. **The plausibility of the forecasts for the proposed project.** This can be achieved by analyzing demand for similar projects in the same or other regions in terms of total demand, and distributions of demand by purpose, geography, time of day, and demographic segments.

Flexible and Expandable Design

ActivitySim is generally a flexible model system, in which it is relatively straightforward to plug in different model components or extend the choice set of existing components to consider new alternatives such as modes or destinations. One must be careful, of course, to ensure that downstream models and summaries of outputs are consistent with whatever changes are made. Nonetheless, the design should anticipate to the extent possible, which components will be made available in the initial deployment, and which features would be available in subsequent components once additional data and/or resources are available. See the section below for more details.

2.4 FEATURES NECESSARY TO SUPPORT NEEDS AND REQUIREMENTS

Several features were implied by discussion to date under the *Model Requirements* section above. These appear in Table 1. Each feature is described by its benefit, how it is addressed in ActivitySim and/or commercial transport software, the data requirements to support development and/or implementation of the feature, and whether the feature must be addressed prior to model estimation. An anticipated level of effort was assigned to each feature, as shown in Table 2. The feature was assigned a low level of effort if no additional data or software development would be required to implement the feature. A medium level of effort was assigned if either additional data or software development would be required. A high level of effort was assigned if both additional data and additional software development would be required. The features were the basis of discussion in a series of virtual sessions including discussions within the Oregon Modeling Steering Committee, which eventually rated each feature by desirability for inclusion in the model (low, medium, and high), also shown in Table 2. Several features were identified for further discussion, as indicated by yellow shaded boxes in the table. The tabulation of OMSC priority rankings by feature rating is shown in Table 3.

TABLE 1: MODEL SENSITIVITIES AND FEATURES

Model feature\sensitivity	Benefit	ActivitySim representation	Commercial transport software representation	Data requirements	Required for joint estimation?
Number of households by zone (including GQ)	Core feature	Synthetic population inputs		Typical land-use data; base year from inventory, future year data estimated	NA
Employment by type and zone	Core feature	Land-use data inputs		Typical land-use data; base year from inventory, future year data estimated	The employment types should be common across regions
Enrollment by type and zone	Core feature	Land-use data inputs		Typical land-use data; base year from inventory, future year data estimated	The types of enrollment should be common across regions
Active park space by zone	Core feature	Land-use data inputs		Typical land-use data; base year from inventory, future year data estimated	NA
Household demographic variables: household size, number of workers, number of children, household income, etc.	Core feature	Standard synthetic population inputs		Typical land-use data; base year data from Census, future year data estimated	The exact variables should be determined
Person demographic variables: work status, student status, age, etc.	Core feature	Standard synthetic population inputs		Typical land-use data; base year data from Census, future year data estimated	The exact variables should be determined
Roadway capacity projects (adding/subtracting lanes of travel, changes in free-flow speed, and facility type)	Core feature	Mode choice via travel time skims by mode and time of day, and upper-level models via logsums		Typical network data - number of lanes, free-flow speed, facility type	NA

Model feature\sensitivity	Benefit	ActivitySim representation	Commercial transport software representation	Data requirements	Required for joint estimation?
Toll roads and managed lanes	Core feature	Mode choice via auto cost skims by occupancy and time-of-day, and upper-level models via logsums	Toll/non-toll path in assignment	Typical network data - cost of travel on each link by time period and occupancy	NA
Transit service frequency	Core feature	Mode choice via travel time skims by mode and time of day, and upper-level models via logsums		Typical network data - route alignment, stop locations, route frequency by time period	NA
Transit technology (local versus BRT, LRT, CR, etc.)	Core feature	Mode choice via travel time skims by mode and time of day, and upper-level models via logsums		Typical network data - route alignment, stop locations, route frequency by time period	The mode choice model structure and skims should be determined
Transit fare	Core feature	Mode choice via transit fare skims by mode and time of day, and upper-level models via logsums		Typical network data - fare by route or operator	NA
Auto operating costs	Core feature	Mode choice utilities and upper-level models via logsums		Either average AOC or by vehicle fuel type, body type, and age from vehicle type choice model (see below)	Whether to use average AOC or vehicle type model must be determined
Auto parking costs	Core feature	Mode choice utilities and upper-		Parking cost (hourly, daily, monthly) by zone	NA

Model feature\sensitivity	Benefit	ActivitySim representation	Commercial transport software representation	Data requirements	Required for joint estimation?
		level models via logsums			
Built environment variables (density, intersections, mixed use, place type(?), etc.)	Core feature	Variables in mode choice and possibly other models		Typical land-use data; benefits from all-streets network and small zones	The exact variables should be determined
Intersection improvement projects; turn lanes, through lanes, stop signs, traffic controls, roundabouts	Measure benefits of intersection improvements as part of plan updates	Mode choice via travel time skims by mode and time of day, and upper-level models via logsums	Intersection-level VDF in assignment required	Additional network attributes - turn pocket presence and directionality at intersections, access control via medians and limits on driveways for road segments, control type and character (stop vs. signal control, presence of signal coordination and progression, g/c ratio, cycle length, etc.).	No, will come through travel time skims. So long as times are close to observed should be OK.
Heterogeneity in travel time sensitivity and willingness to pay	Represent variation in individual willingness to pay on route choice and mode choice	Draw sensitivity factor from log-normal distribution; compute person-specific value-of-time; match to auto (and transit) skims used in mode choice	Skim and assign auto (and transit) by VOT segment	No additional data	We typically estimate average travel time in model and apply heterogeneity based on separate estimations using SP data

Model feature\sensitivity	Benefit	ActivitySim representation	Commercial transport software representation	Data requirements	Required for joint estimation?
Auto travel reliability	Represent auto network reliability on route and mode choice	Mode choice via travel time reliability skims by TOD; upper level models via logsums	Skim auto reliability; possibly include in VDF/path cost	Travel time reliability data or function	Can attempt to estimate parameter if included; if not can apply/assert a parameter in application
Transit fare subsidies	Represent who receives subsidized transit fares from employers\schools, and measure effects of changes in subsidy programs on mode choice	Transit subsidy model and variables in mode choice		No additional data	Yes, model needs to be estimated
Transit pass ownership	Model who owns transit passes, and measure effects of changes in pass ownership on mode choice	Transit pass ownership model and variables in mode choice		Average costs and values of transit passes compared to base transit fare.	Yes, model needs to be estimated
License holding status	Better represent age impacts and measure potential changes in license holding on auto demand	New license holding model and variables in mode choice		No additional data	Yes, model needs to be estimated
Mobility impairment/disability	Better represent age impacts and disabilities on mode choice	New mobility impairment model and variables in mode choice		No additional data	Yes, model needs to be estimated

Model feature\sensitivity	Benefit	ActivitySim representation	Commercial transport software representation	Data requirements	Required for joint estimation?
Transit vehicle capacity	Represent crowding effects and transit capacity improvements (more service, bigger cars) on transit demand	Mode choice via crowded time skims (IVT and wait) by mode and time of day, and upper-level models via logsums	Capacity-constrained transit assignment required	Transit vehicle type for each route and capacity of each vehicle	Can attempt to estimate parameter if included; if not can apply/assert a parameter in application
Transit park-and-ride lot capacity	Represent parking supply at park-and-ride lots on transit demand	Mode choice via skims indicating crowding levels and/or other mechanism	Capacity-constrained transit assignment or other method required	Capacity for park-and-ride lots	Can attempt to estimate parameter if included; if not can apply/assert a parameter in application
Transit reliability	Represent transit reliability on route and mode choice	Mode choice via modified transit time skims by TOD; upper-level models via logsums	Modify travel times and/or wait times according to reliability	Data and analysis of transit reliability	Can attempt to estimate parameter if included; if not can apply/assert a parameter in application
Subsidized (employer-provided) parking. Cash out. Employer provided cars.	Represent employee subsidies for parking and employer-provided parking and potential future changes in subsidies.	Employer parking subsidy model, mode choice variables			Yes, model needs to be estimated
Auto parking capacity. EVs may be a separate issue.	Represent parking supply at parking lots in urban areas on auto demand	Include supply to weight expected parking costs by quantity, implement parking constraint mechanism		Parking supply (number of spaces)	Can attempt to estimate parameter if included; if not can apply/assert a parameter in application

Model feature\sensitivity	Benefit	ActivitySim representation	Commercial transport software representation	Data requirements	Required for joint estimation?
Bicycle facilities, slope, traffic exposure/stress, quality of facilities, etc.	Represent bicycle infrastructure and quality of bike route options on mode choice	Mode choice via weighted skims or logsum from bike route choice model	All-streets network processing, possibly stochastic route choice with path overlap compensation	All-streets network, smaller zones/microzones, bicycle facility representation, bicycle counts for validation	Need to determine exact measures to be included in mode choice
Willingness to bicycle	Represent variation in individual willingness to bike; change assumptions about bicycle comfort levels for future year scenarios and measure effects on mode choice	New latent class model predicting bicycle comfort level; additional bike skims or logsums (by class), and mode choice variables	All-streets network processing, possibly stochastic route choice with path overlap compensation	All-streets network, smaller zones/microzones, bicycle facility representation	Yes, model needs to be estimated
Privately owned e-bikes	Represent privately owned e-bike ownership and changes in e-bike ownership on mode choice	New e-bike ownership model or share and fixed percentages and mode choice extension		Average e-bike speed	Yes, model needs to be estimated
Shared e-bikes	Represent e-bike fleet and availability	Mode choice extension		Shared e-bike access time by zone, cost, average speed	The mode choice model structure and skims should be determined
Privately owned e-scooters	Represent privately owned e-scooter ownership and changes in e-scooter ownership on mode choice	New e-scooter ownership model or share and fixed percentages and mode choice extension		Average e-scooter speed	Yes, model needs to be estimated

Model feature\sensitivity	Benefit	ActivitySim representation	Commercial transport software representation	Data requirements	Required for joint estimation?
Shared e-scooters	Represent e-scooter fleet and availability			Shared e-scooter access time by zone, cost, average speed	The mode choice model structure and skims should be determined
Sidewalks, walking trails	Represent presence of sidewalks and walking trails on mode choice	Mode choice via weighted skims or logsum from walk route choice model	All-streets network processing, possibly stochastic route choice with path overlap compensation	All-streets network, smaller zones/microzones, walk facility representation, pedestrian counts for validation	The mode choice model structure and skims should be determined
Ride-hail services	Represent taxi, Transportation Network Company (TNC) in model	Mode choice extension		Average wait times (typically varies by origin zone) and costs	The mode choice model structure and skims should be determined
Car share	Represent proximity of car share in travel models	TBD; possibly car share availability model, mode choice variable affecting car mode use		Car share availability by zone, cost	Yes, model needs to be estimated
Worker occupation and/or industry	Represent linkage between worker characteristics and jobs; represent propensity to telecommute and other mobility effects by industry or occupation	Synthetic population variable used in work location choice and other models		Base year data from Census, future year data estimated	The exact variables should be determined

Model feature\sensitivity	Benefit	ActivitySim representation	Commercial transport software representation	Data requirements	Required for joint estimation?
Person level race\ethnicity	Can be used to calculate disaggregate benefits and impacts for disadvantaged populations	Synthetic population variable used for post-processing results		Base year data from Census, future year data estimated	No, these variables are typically used to summarize results as opposed to be used as explanatory variables
Electric vehicles	Represent EV fleet on auto operating cost, mode choice and use model to forecast changes in energy consumption and emissions	Implement vehicle type choice model and mode choice variables. Post-process results			This would be handled by the vehicle type model; needs to be determined
Autonomous vehicles	Represent AV fleet on auto operating cost, mode choice, and VMT due to vehicle repositioning	Implement vehicle type choice model and mode choice variables. Post-process results			No
Internal-external travel	Model travel to/from external stations made by residents more accurately. Prevent double-counting IE trips.	Implement internal-external identification and external station choice models		Calibration data	If included, models need to be estimated
Airport ground-access travel	Model travel to/from airports made by residents and visitors more accurately.	Implement airport model		Application and calibration data	If included, models need to be estimated

Model feature\sensitivity	Benefit	ActivitySim representation	Commercial transport software representation	Data requirements	Required for joint estimation?
University campus travel	Model intra-campus trips between zones, trips between remote parking and campus	Implement university enhancements		Land-use data: university space by zone, locations of remote lots, restrictions on use	If included, models need to be estimated
Overnight visitor travel	Model trips made by overnight visitors	Implement overnight visitor model		Total regional visitors, lodging/rooms by zone	If included, models need to be estimated
Seasonal/weekend travel	Model travel made during winter\summer peak and/or weekends	Estimate parameters representing such travel		Behavioral data and counts on weekends and/or by season	If included, models need to be estimated

TABLE 2: ESTIMATED LEVEL OF EFFORT FOR MODEL FEATURES AND OREGON MODELING STEERING COMMITTEE RESPONSE

Model feature\sensitivity	Level of effort\runtime impacts	OMSC Response	ODOT	Metro	SKATS	LCOG
Number of households by zone (including GQ)	Low	Core	Core	Core	High	High
Employment by type and zone	Low	Core	Core	Core	High	High
Enrollment by type and zone	Low	Core	Core	Core	High	High
Active park space by zone	Low	Core	Core - needs to be better than in the past	Core	High	High
Household demographic variables: household size, number of workers, number of children, household income, etc.	Low	Core	Core	Core	High	High
Person demographic variables: work status, student status, age, etc.	Low	Core	Core	Core	High	High
Roadway capacity projects (adding/subtracting lanes of travel, changes in free-flow speed, and facility type)	Low	Core	Core	Core	High	High
Toll roads and managed lanes	Low	Core	Core	Core	Low - But pricing VMT/mile High	Toll roads low. Managed lanes maybe high.

Model feature\sensitivity	Level of effort\runtime impacts	OMSC Response	ODOT	Metro	SKATS	LCOG
Transit service frequency	Low	Core	Core	Core	High	HIgh
Transit technology (local versus BRT, LRT, CR, etc.)	Low	Core	Core	Core	Medium	High
Transit fare	Low	Core	Core	Core	High	High
Auto operating costs	Low	Core	High - need vehicle level representation, agencies that don't require vehicle level can code all vehicles at the global average.	Core	High via average AOC Low for vehicle type model	High
Auto parking costs	Low	Core	Core	Core	High	High
Built environment variables (density, intersections, mixed use, place type(?), etc.)	Low	Core	Core - needs to be better than the past	Core	High	High

Model feature\sensitivity	Level of effort\runtime impacts	OMSC Response	ODOT	Metro	SKATS	LCOG
Intersection improvement projects; turn lanes, through lanes, stop signs, traffic controls, roundabouts	Medium (additional data requirements)	High, but assuming no impact to estimation beyond a relative skim consistency	High	High - But this will most likely be achieved through proxy variables (adjustments to capacity, speed, etc.)	High	High
Heterogeneity in travel time sensitivity and willingness to pay	Medium (increased runtime)	High	Medium	High	Medium	Medium
Auto travel reliability	Medium (increased data)	Low, come back to this in a future improvement phase. Making the assumption that a MPO could always add a reliability component to assignment if desired.	Medium, feels like there could be phases to the development and this is a good example of something that could be added in a later phase, perhaps a middle ground assignment level representation that isn't fully TDM integrated.	Medium - Difficult to see how to achieve this with current assignment tools	Low	Low - I think we would derive this from the RITIS or other similar platform, not the travel model

Model feature\sensitivity	Level of effort\runtime impacts	OMSC Response	ODOT	Metro	SKATS	LCOG
Transit fare subsidies	Low	High	High	High	High	High
Transit pass ownership	Low	High	High	High	High	High
License holding status	Low	Medium-Low, could the design team discuss how much effort and impacts of adding this feature (what data is needed).	High	Med	Low	Low
Mobility impairment/disability	Low	High	High	High	High	High
Transit vehicle capacity	High (additional data requirements, increased runtime)	Low, come back to this in a future improvement phase. Assuming that interested parties could potentially represent with assignment methods prior to a solution that impacts estimation	Low, feels like there could be phases to the development and this is a good example of something that could be added in a later phase. Could likely handle with just transit assignment as a middle-of-the-road approach, without fully integrating	Low	Low due to work load.	Seems of value but not a high priority. I like the phased approach in ODOT's comment

Model feature\sensitivity	Level of effort\runtime impacts	OMSC Response	ODOT	Metro	SKATS	LCOG
			vehicle capacity into the TDM			
Transit park-and-ride lot capacity	High (additional data requirements, increased runtime)	Low, but this is a core feature for Metro and the estimation needs to anticipate that a future enhancement will be adding this capability (we'd like to explore options with design team)	Low, perhaps traffic assignment techniques could be approached for lower level of effort.	Core - This should be paired with a PnR lot choice model	Low	Low
Transit reliability	Medium (additional data requirements)	High, but since we don't know the method - we will have to determine in the design work if the costs are too high to complete this feature	Medium	Medium - Difficult to see how to achieve this with current assignment tools	High	High

Model feature\sensitivity	Level of effort\runtime impacts	OMSC Response	ODOT	Metro	SKATS	LCOG
Subsidized (employer-provided) parking. Cash out. Employer provided cars.	Low	High	High	High	High for testing parking cash-out	High
Auto parking capacity. EVs may be a separate issue.	Medium (possibly increased data depending on availability of parking quantity data in region. Increased runtime)	Medium, but this is a core feature for Metro and the estimation needs to anticipate that a future enhancement will be adding this capability (we'd like to explore options with design team, Metro's current trip-based solution is linked to the PNR feature - can these be linked for the ABM too)	Medium - this could become really important as CFA's reduce the number of required spaces	High	Medium?	Medium-high. CFEC has parking reduction requirements that may be interesting to better understand.
Bicycle facilities, slope, traffic exposure/stress, quality of facilities, etc.	High (increased data, increased runtime)	High, OMSC will be having further discussions on learning more about Metro's bike model and how we might be able to build on that success	High	High	High	High
Willingness to bicycle	High (increased data, increased runtime)	High	High	High	High	High

Model feature\sensitivity	Level of effort\runtime impacts	OMSC Response	ODOT	Metro	SKATS	LCOG
Privately owned e-bikes	Low	High, just looking for a probably look-up - not expecting a choice model	High	High	Medium	High
Shared e-bikes	Medium (increased data)	Low	Low	Low	Low	Medium - Eugene may be quite interested in this
Privately owned e-scooters	Low	Medium, we would like this to be considered similar to e-bike, like a micro mobility treatment	Medium	Low	Medium	Medium
Shared e-scooters	Medium (increased data)	Low	Low	Low	Low	Medium
Sidewalks, walking trails	High (increased data, increased runtime)	Medium, interested in talking through options during the design - Can we discuss how the MAZ system complements this, how can we best represent walking in the ABM.	Medium, might need to learn more from RSG about the real costs of doing this - could this tie to built environment variable	Medium - Difficult to see how to achieve this with current assignment tools	Low	See ODOT's response
Ride-hail services	Low	High	High (given low effort)	High	High	High
Car share	Medium (increased data)	Low	Medium	Medium	Low	Low

Model feature\sensitivity	Level of effort\runtime impacts	OMSC Response	ODOT	Metro	SKATS	LCOG
Worker occupation and/or industry	Medium (increased data)	Medium	Medium, might need to learn more from RSG about the real costs of doing this	Medium	Medium	Medium
Person level race\ethnicity	Medium (increased data)	Low	N/A - should not be something that has any impact on the design	Unless this impacts traveler behavior, low	Low - Seems difficult in the future	Low
Electric vehicles	Low	High - don't plan on using the choice model, using "Option 2" the simpler model.	High	High	Low	High
Autonomous vehicles	Low	Low, perhaps attempt in a later enhancement phase, but can we learn more about how easy it is to put in - is this easy?	Low, feels like there could be phases to the development and this is a good example of something that could be added in a later phase.	Low	Low	See ODOT's response
Internal-external travel	Medium (increased data, model development)	High - interest in getting external travel integrated with SWIM	High - we expect this to use SWIM	Core	High	High
Airport ground-access travel	High (increased data, model development)	Low, Important (critical) to Metro, but they have a working option off-model (added to the trip table). Some future	Low	Medium	Low	Low

Model feature\sensitivity	Level of effort\runtime impacts	OMSC Response	ODOT	Metro	SKATS	LCOG
		effort will need to update the current airport representation				
University campus travel	Medium (increased data, model development)	Low, perhaps attempt in a later enhancement phase	Low - given data costs - perhaps a future phase and specific data collection	Low	Low	See ODOT's response
Overnight visitor travel	High (increased data, model development)	Low, perhaps attempt in a later enhancement phase	Low - given data costs - perhaps a future phase and specific data collection	Medium	Low	See ODOT's response
Seasonal/weekend travel	High (increased data, model development, runtime)	Low, perhaps attempt in a later enhancement phase	Low - given data costs - perhaps a future phase and specific data collection	Medium - Difficult to see how to achieve this with current assignment tools	Low	See ODOT's response

TABLE 3: MODEL FEATURES BY OREGON MODELING STEERING COMMITTEE RANKING

Level of Effort	OMSC Priority Rank			Total
	High	Low	Medium	
Low Effort	7	1	2	24
Medium Effort	4	6	2	12
High Effort	2	5	1	8
Total	**13**	**12**	**5**	**44**

3.0 MODEL SPECIFICATION AND DESIGN

3.1 ACTIVITYSIM GENERAL MODEL FORM

The ActivitySim model, which is fully described in the following sections, has the following characteristics:

- Utilizes tours (sequences of trips beginning and ending at an anchor location such as home or work) as an organizing principle for the generation of travel and to ensure consistency across trips within a tour.
- Utilizes micro-simulation for modeling travel choices, in which a synthetic population is generated, and explicit mobility and travel choices are made for each decision-maker in the population according to contextual probability distributions.
- Addresses both household-level and person-level travel choices including intra-household interactions between household members.
- Uses a half-hourly temporal resolution and schedules tours into time-windows to ensure there are no overlapping travel episodes.
- Reflects and responds to detailed demographic information including household structure, aging, changes in wealth, and other key attributes.
- Considers recent changes in work arrangements including working from home and telecommuting.
- Includes explicit representation of both shared and privately owned mobility including e-scooters, e-bikes, and ride-hail services.
- Includes explicit representation of electric vehicles and autonomous vehicles.

These aspects of the model system are described in greater detail below.

3.2 TRAVEL DEMAND MODEL SEGMENTATION

ActivitySim represents all internal travel made by residents of each region with the exception of travel to airports (typically handled via an air ground access model) and travel for purely recreational purposes, such as jogging or bike travel with no fixed out of home activity location. The decision-makers in the model system include both persons and households. These decision-makers are created (synthesized) for each simulation year and land-use scenario based on Census data and forecasted distributions of households and persons by key socio-economic categories. This synthetic population is created within PopulationSim. The decision-makers are used in the subsequent discrete-choice models in a microsimulation framework where a single alternative is selected from a list of available alternatives according to a probability distribution. The probability distribution is typically generated from a logit model which considers the attributes of the decision-maker and the attributes of the various alternatives. The application paradigm is referred to as Monte Carlo simulation, since a random number draw is used to select an alternative from the probability distribution. The decision-making unit is an important element of model estimation and implementation and is explicitly identified for each model specified in the following sections. In a few cases, the choice is drawn from an observed probability distribution that takes into account certain explanatory variables, rather than from a discrete choice model. Such cases are noted in the detailed description below.

The model system is implemented in a micro-simulation framework. A key advantage of using the micro-simulation approach is that there are essentially no computational constraints on the number of explanatory variables that can be included in a model specification. However, even with this flexibility, the model system will include some segmentation of decision-makers. Segmentation is a useful tool to both structure models (for example, each person type segment could have their own model for certain choices) and to characterize person roles within a household. Segments can be created for persons as well as households.

Household Segmentation Variables

There are several variables that convert continuous variables into categorical variables for more efficient model estimation. These include household income and auto "parity" (sometimes referred to as "sufficiency" though that term implies a positive connotation for owning more cars). These variables will be coded in the household preprocessor so that they will be consistently defined throughout the model system.

The household travel survey has nine categories of household income as shown in Table 4 (plus refused, not shown, which will require imputation). The table proposes a universal variable naming convention for each range. It also suggests possible collapsed categories if necessary for model estimation in certain models to ensure reasonable model responses with respect to income.

Auto parity compares autos owned to licensed drivers in the household. Those variables are shown in Table 5. Both will be modeled explicitly in ActivitySim.

TABLE 4: HOUSEHOLD INCOME CATEGORIES (YEAR $2023)

Income range	Variable Name	Possible Collapsed Name
Income < $15k	inc_lt_15	inc_lt_25
$15k - $24.9k	inc_15_24	inc_lt_25
$25k - $34.9k	inc_25_34	inc_25_49
$35k - $49.9k	inc_35_49	inc_25_49
$50k - $74.9k	inc_50_74	inc_50_99
$75k - $99.9k	inc_75_99	inc_50_99

$100k - $149.9k	inc_100_149	inc_100_199
$150-$199.9k,	inc_150_199	inc_100_199
$200k+	inc_200_plus	inc_200_plus

TABLE 5: AUTO PARITY CATEGORIES

Autos to drivers comparison	Auto parity
Autos = 0	autos_0
Autos > 0 and autos < drivers	autos_lt_drivers
Autos > 0 and autos >= drivers	autos_ge_drivers

Person Types

A total of seven segments of person-types, shown in Table 6 are used for the model system. The person-types are dependent upon age, work status, and school status, and are coded according to a hierarchy of rules. The same methodology is used to code person type in travel behavior survey data, so that apples-to-apples comparisons can be made between model results and observed data. Note that previous ActivitySim implementations break out driving age student from other K-12 student. However this is typically the smallest person type category which makes estimation for this category difficult, and it implies driving status for this person type that may or may not apply depending upon changes in license holding status over time and emerging mobility modes such as autonomous vehicles.

TABLE 6: PERSON TYPES

NUMBER	PERSON-TYPE	AGE	WORK STATUS	SCHOOL STATUS
1	Full-time worker	16+	Full-time	None
2	Part-time worker	16+	Part-time	None
3	University student	16+	Any	College/University
4	Non-worker & non-student	16-64	Unemployed	None
5	Non-working senior	65+	Unemployed	None
6	K-12 student	6-18	None	Pre-college
7	Pre-school	0-5	None	None

Rules for coding person type consist of just calculating variables to be able fill out table 6 above. Synthetic populations are derived from PUMS data and that data is then used to determine the work status, school status, and ultimately person type. Work status is first determined by the ESR (employment status recode) column. If ESR denotes the person does not work, then they are unemployed. If ESR denotes the person is a civilian or in the armed forces and is at work, then it goes to the WKW (weeks worked in the past 12 months) and WKHP (usual hours worked per week in the past 12 months). If WKW is 27 weeks or more and WKHP is 35 hours or more, then the person is classified as a full-time worker. Employed people with WKW and WKHP less than those thresholds are part-time workers.

School status is determined by the SCHG (grade level attending) variable. If the person reports not being in university level school and they are under the age of 20, they are listed as grade or high school students. People aged 16 and older who report being in college are listed with a college status. Anyone aged 20 and older who reports attending school are also given a college school status.

Once work and school status are determined, people are just sent down the list in Table 6 and are provided with the first category they meet the criteria for. Each person can only have a single person type. That means that university students who work full time have a person type of full-time worker. Similarly, if a person over the age of 65 is working only part-time, they are classified as part-time workers.

The calculation of the person types based on PUMS data is performed in the ActivitySim annotate_persons.csv file that pre-processes the persons table at the beginning of ActivitySim runs. (See the subsequent section on Household and Person Preprocessors for more details.)

3.3 ACTIVITY TYPES

The activity types explicitly modeled in ActivitySim are shown in Table 7. The activity types are also grouped according to whether the activity is mandatory, maintenance, or discretionary, and eligibility requirements are assigned determining which person types can be used for generating each activity type. The classification scheme of each activity type reflects the relative importance or natural hierarchy of the activity, where work and school activities are typically the most inflexible in terms of generation, scheduling and location, whereas discretionary activities are typically the most flexible on each of these dimensions. Each out-of-home location that a person travels to in the simulation is assigned one of these activity types. Any activity with missing purpose is recoded as discretionary activity. Note that change mode trips are linked according to a set of trip-linking rules.

Also note that the current ActivitySim implementation does not have a separate purpose for work-related activities, such as meetings and service calls. Work activities are specifically work at fixed locations. Work at non-fixed locations and other work-related activities are grouped with 'Other Maintenance'. The share of workers without a regular out-of-home workplace location significantly increased post-COVID, which was accompanied by an increasing share of work-related travel as those workers visit clients and attend occasional meetings. The project team should evaluate the Oregon household survey and determine whether it is feasible to break out work-related as a new activity purpose. Adding a purpose would require re-designing certain aspects of the model such as the mandatory tour frequency models and adding time of day, mode, and destination choice models. It would be a significant effort that would add to both model development and software implementation costs.

Also, current ActivitySim implementations break out social/visiting from other discretionary activities. Survey respondents may be challenged to identify differences in activity purposes according to this categorization and social/visiting tends to have few observations; the project team may want to collapse these two purposes.

TABLE 7:OUT-OF-HOME ACTIVITY TYPES

TYPE	PURPOSE	DESCRIPTION	CLASSIFICATION	RELEVANT PURPOSES FROM SURVEY
1	Work	Working at regular workplace.	Mandatory	Went to work/work-related/volunteer-related at primary workplace
2	University	College	Mandatory	Attended school/class at primary school, college/university or vocational education
3	Grade/High School	Grades K-12	Mandatory	Attended school/class at primary school, K-12
4	Escorting	Pick-up/drop-off passengers	Maintenance	Pick someone up, drop someone off, or both
5	Shopping	Shopping away from home.	Maintenance	Grocery shopping Other routine shopping (e.g., pharmacy), Shopping for major item (e.g., furniture, car)

6	Other Maintenance	Personal business/services, and medical appointments.	Maintenance	Refueled/Recharged the vehicle, Medical visit (e.g., doctor, dentist), Errand without appointment (e.g., post office), Errand with appointment (e.g., haircut), other type of class or education related activity not at school location, work-related activity not at primary workplace
7	Eat Out	Eating outside of home.	Discretionary	Dined out/got coffee or take-out
8	Social/Recreational	Recreation, visiting friends/family.	Discretionary	Social activity (e.g., visit friends/relatives), go to someone else's home
9	Other Discretionary	Volunteer work, religious activities.	Discretionary	Family activity (e.g., watch child's game), Leisure/entertainment/cultural (e.g., cinema, museum), Religious/civic/volunteer activity, Went to temporary lodging, other

3.4 TREATMENT OF TIME

The Oregon ActivitySim ABM model system will utilize a half-hourly temporal resolution, beginning with 3 A.M. and ending with 3 A.M. the next day. There are a total of 48 half-hour periods in the model. Temporal integrity is ensured so that no activities are scheduled with conflicting time windows, except for short activities/tours that are completed within a half-hour increment. For example, a person may have a short tour that begins and ends within the 8:00 A.M. - 8:30 A.M. period, as well as a second longer tour that begins within this time period but ends later in the day.

A critical aspect of the model system is the relationship between the temporal resolution used for scheduling activities, and the temporal resolution of the network simulation periods. Although each activity generated by the model system is identified with a start time and end time in half-hour increments, level-of-service matrices are only created for five aggregate time periods – Night, Early A.M., A.M. Peak, Mid-Morning, Midday, Afternoon, P.M. Peak, and Evening. The trips occurring in each time period reference the appropriate transport network depending on their trip mode and trip departure period (for outbound trips) or arrival period (for inbound trips). The definitions of time periods for each skim period will vary for each implementation and will need to be defined based upon an analysis of traffic and congestion by time of day and the provision of transit service.

3.5 TREATMENT OF SPACE

The Oregon ActivitySim model relies on network level of service skims at the Transportation Analysis Zone (TAZ) level and land-use data at the Micro-Analysis Zone (MAZ) level. The development of Transportation Analysis Zones is described in the Travel Demand Model Development and Application Guidelines manual. MAZs are typically developed in one of two ways. One way is to start from Census Blocks. These can either be split out (for large Census blocks) or aggregated (for small Census Blocks) in order to better define non-motorized accessibilities and land-use variables. The other way is to start from an intersection of road links, after excluding freeways ramps and creating single boundaries for split facilities in the network such as freeways and possibly principal arterials. All household locations and trip origins and destinations are modeled at the lowest MAZ level geography. MAZs are used for representation of walk and bike times, as described in Section 4.3.

3.6 OVERALL MODEL DESIGN

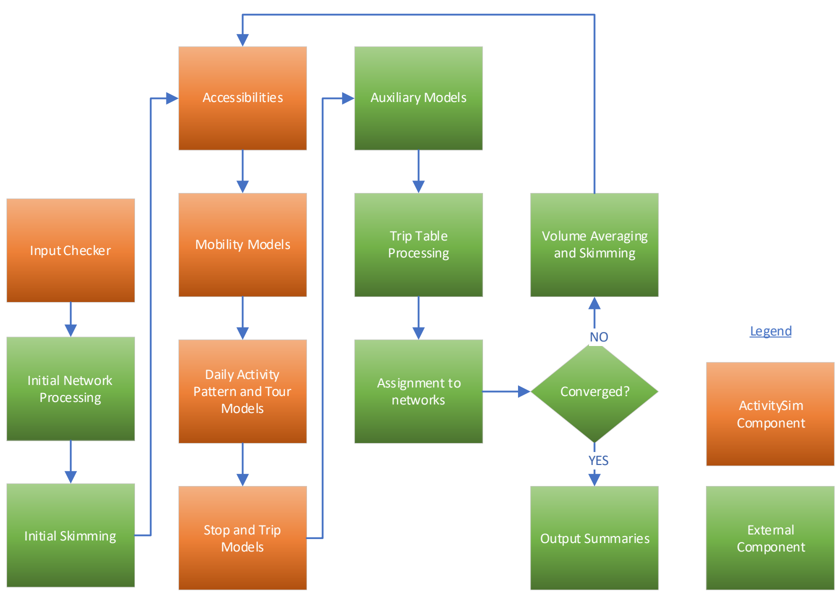
The overall model design is shown conceptually in Figure 6. ActivitySim steps are shown in orange, and steps calculated outside ActivitySim are shown in green. The model starts with population synthesis (not shown, as the OMSC has decided to not include population synthesis in the automated model system). The model starts with network processing procedures such as creating networks by time of day from input data, calculation of link and intersection capacities from input variables, and so on.

An initial assignment and set of skims will be created. There are typically two options for creating initial skims. The first option is to use speed lookup tables by time of day; these usually vary by facility type and area type. The advantage of this approach is that it does not require input trip tables. The disadvantage of this approach is that input speed calculations are typically not very accurate or consistent with assignment results, which can require more iterations to reach convergence. The other option is to assign input trip tables from a similar scenario to the network. The advantage of this approach is that it is typically consistent with output trip tables from a converged model run and therefore requires less model iterations to reach convergence. The disadvantage is that it requires trip tables consistent with the zone system(s) used by the model. We recommend including both approaches with a switch to enable the assignment of existing trip tables in the initial iteration if they are available, and to use a speed lookup table if they are not. Skims will be created for auto, transit, and non-motorized modes. Auto and transit skims are created using the planning network by origin-destination TAZ for each time period, mode, and market segment. Non-motorized skims are created from the all-streets network by MAZ pair.

ActivitySim is run next; there are four components shown: accessibility calculations, mobility models, daily activity pattern and tour level models, and stop/trip level models. Each component is explained in more detail below.

Next, auxiliary models are run. This includes commercial vehicles, heavy trucks, external-internal and external-external trip tables, and any other special market models not otherwise addressed by ActivitySim (airport ground access models, overnight visitor models, etc.). All trip tables are processed for input to assignment and assigned to transport networks. Next, the model system checks for convergence. The convergence criteria is typically defined as less than 5% RMSE change on peak period trip tables between the current iteration and the last iteration, and less than 5% RMSE change on assigned volumes in any time period. If convergence criteria is not met, the model iterates. This involves calculating the average of the current iteration assigned volume and previous iteration's averaged assigned volume on each link in the network[22], and reskimming the network for the next iteration. If convergence criteria is met, the model stops iterating and model outputs are summarized for performance metrics.

FIGURE 6: OVERALL MODEL DESIGN



3.7 POPULATION SYNTHESIS

Activity-based models are applied to a synthetic population. The PopulationSim software[23] will be used to generate that synthetic population. PopulationSim is highly flexible; it allows controls in as many geographies as desired, and any categorical variables can be listed as controls as long as they are specified in the seed data. Household and person files from the American Community Survey (ACS) Public Use Microdata Sample (PUMS) will be used as seed data. The only fully required control in PopulationSim is that the number of households needs to be specified at the lowest level of geography (MAZ). All other controls are fully up to the user.

---

22 This is referred to as the method of successive averages (MSA).
23 https://activitysim.github.io/populationsim/

However, the user should be aware that while PopulationSim will allow any control from the seed data, that does not mean that any set of controls will result in a meaningful output/result. A handful of incorrect uses with PopulationSim will result in a meaningless (error prone) resulting synthetic population:

- In addition to ensure that all geographies nest within one another, the user must also ensure that all controls sum properly. Meaning that if there are 10 blocks in a block group, and those 10 blocks sum to 150 total households. Then any controls at the block group level must sum to 150 households. If the totals are off at different levels of geography, the result will be an inconsistent and poorly matched set of controls.
- If a user tries to control two different PUMS fields that are highly correlated, it may result in conflicting controls because the two highly related fields may conflict with one another at some levels of geography. A good example of this might be controlling persons by age and also trying to control households by age of head of household. If both sets of controls are not perfectly consistent, the population synthesis procedure may have more variance in the match to controls due to that inconsistency.
- Certain controls, for example most of the household summaries from ACS, are specific to non-group quarters households. Therefore group quarters population needs to be synthesized separately from non-GQ households, and then appended to the non-GQ households prior to running ActivitySim.
- Lastly, it should also be noted that PUMS data is a 5% sample of households over a five-year period. Specification of highly disaggregate controls can result in household replicates which must exceed the maximum expansion factor in order to match controls, or for combinations of controls for which no seed data can be found. So users must be cognizant of striking a balance between specificity in controls versus data coverage.

With those cautions being considered, the following section describes the proposed controls for the general population (not group quarters) to use at the different levels of geography (in order from smallest to largest).

MAZ level controls are specified in Table 8. Total households will be matched perfectly by PopulationSim. For future years, demographic controls such as household size, household income, and workers per household are difficult to project without a land-use model. Households by dwelling type are recommended since this is typically relatively easy to forecast in the future based on zoning and comprehensive plan data and helps control for population size at a small geographic level. If the area does have a good forecast of households by dwelling type and a poor forecast on demographics, then the future year forecasts can be constrained at a higher geographic level or base-year distributions can be applied and varied only for geographies that are likely to undergo significant change from the base year.

TABLE 8: MAZ LEVEL CONTROLS

FIELD NAME	DESCRIPTION	CATEGORIES	SOURCE
TOTNGQHHS	Total Non-GQ households	NA	Census or household inventory

TYPE_[CATEGORY]	Optional – Dwelling Type if available at this level	SF, duplex, MF<4, MF 4-25, MF 25-50, MF>50, Mobile home	Census or household inventory

Block group controls are specified in Table 9. These controls are derived from census data available at the block group level. It is important to highlight that age will likely be difficult to estimate at the block group level for future years. In Oregon these can be estimated or obtained from the Portland State University Population Research Center (PRC) but are only available at larger geographies (UGB).

TABLE 9: BLOCK GROUP CONTROLS

FIELD NAME	DESCRIPTION	CATEGORIES	SOURCE
HHSIZE_[1…7]	Household Size	1,2,3,4,5,6,7+	2022 ACS 5-year, Table B11016
HHINC_[RANGE]	Household Income	<$15k, $15-$25k, $25-$35k, $35-$50k, $50-$75k, $75-$100k, $100-$150k, $150-$200k, $200k+	2022 ACS 5-year, Table B19001
HASCHILD	Presence of children	Yes/No	2022 ACS 5-year, Table B11005
AGE_[RANGE]	Person age	0-5, 5-9, 10-14, 15-17,18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85 or older	2022 ACS 5-year, Table B01001

Two types of recommended information are only available at the tract level from Census: number of workers and occupation (Table 10). Similar to the block group level, for future years these categories might need to be aggregated up to larger geographies. Both workers and occupations are estimates that can be provided from sources like the Oregon Statewide Integrated Model (SWIM). SWIM could be considered to provide these targets at the PUMA or UGB level for future years.

TABLE 10: TRACT LEVEL CONTROLS

FIELD NAME	DESCRIPTION	CATEGORIES	SOURCE
HHWRK_[0…3]	Number of workers	0, 1, 2, 3+	2022 ACS 5-year, Table B08202
OCC_[VALUE]	Person occupation	1) Management, business, science, and arts; 2) Service; 3) Sales and office; 4) Natural resources, construction, and maintenance; 5) Production, transportation, and material moving; 6) Military specific occupations	2022 ACS 5-year, Table B08124

As context, it's required that all regions outside of Portland-Metro align with the PRC population estimates and forecasts. Those are provided at a UGB level, although some of their other information, such as age projections may only be available at a county level. Table 11 shows UGB level controls for the base year, but as is noted above, for the future year, age and occupation might be moved from block group to this level, since the data is not available or reliable at lower levels.

TABLE 11: URBAN GROWTH BOUNDARY CONTROLS

FIELD NAME	DESCRIPTION	CATEGORIES	SOURCE
POP_[MALE/FEMALE]	Population by Gender	M/F (potentially additional groupings as they become available)	PRC

Group Quarters Development

Non-institutional group quarters populations generate travel and as such must also be represented in the synthetic population. However, much less information is known about group quarters (GQ) population, making group quarter synthesis much easier. We only use one control for GQ households, which identifies whether the household is a military household, a university student, or some other type of non-institutional group quarter household. GQ households are created separately from non-GQ households, which results in two sets of output household and person files which must be appended together to create the full population prior to running ActivitySim. Note that GQ households have one person per household. The seed PUMS data must be pre-processed to just include group quarters households.
Institutional GQ are not included, under the assumption that they are not able to leave the facility for travel. If institutionalized populations are large in one's region; an extra check should be done to determine if the institutionalized population is in the total PRC population for the region, and if so, that population total should be removed from the PRC controls at the UGB level.

GQ households for the year 2020 can be obtained from the 2020 Decennial Census, Table P18. Future year data will need to be forecast, likely by understanding how GQ buildings / areas are forecast to grow into the future.

3.8 ACTIVITYSIM ABM DESIGN

The overall design of ActivitySim is shown in Figure 6. Each component is described in more detail below.

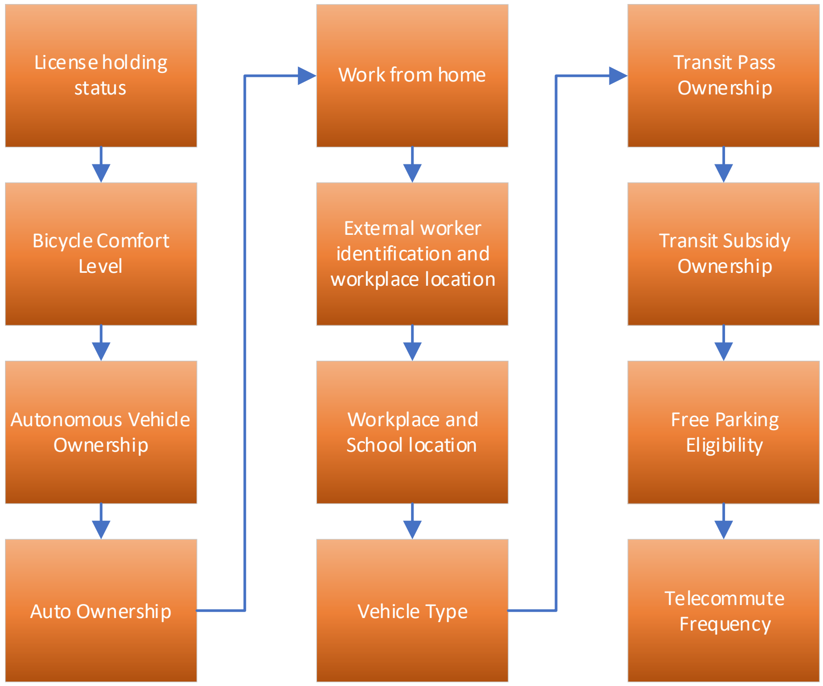
Accessibilities

The first step in the sequence is accessibility generation. There are two types of origin-based accessibilities generated in this step; aggregate mode-specific accessibilities and disaggregate multi-modal origin-based accessibilities. The aggregate origin-based accessibilities are useful because they differentiate accessibility to total employment, retail employment, and total households by each of several modes: auto, walk-transit, walking, and biking. Aggregate accessibility calculations are created using matrix calculations. Disaggregate origin-based accessibilities are tour destination choice logsums from the ActivitySim model system. These are calculated by applying the ActivitySim destination and mode choice models to a tour list created for a prototypical synthetic population that the user defines. The outputs of this step are appended to the actual synthetic population. Each set of accessibilities can be used as explanatory variables to influence mobility models such as auto ownership and models that generate travel such as the coordinated daily activity pattern model and tour frequency models.

Mobility Models

Long term choice models and mobility models, shown in Figure 7, are run next. These include license holding status of each person in the household, bicycle comfort level, whether a household owns an autonomous vehicle, the number of autos owned, whether a worker regularly works from home, and mandatory location choice. Other mobility models are run next, including transit subsidy and transit pass models, free parking eligibility, and telecommute models. The figure does not show bike or micro-mobility ownership; these could be formal models or the model system can simply sample from an ownership probability for each type of mobility option. The latter approach offers the advantage that it is easy to implement and change the ownership probabilities for scenario analysis; the former option would require more resources to estimate and implement but would better account for variables that affect the likelihood of ownership such as socio-economic variables and accessibilities. *Note that the sequence of running these models is illustrative; they can be re-ordered, but the sequence does need to be defined prior to estimation so that the outcomes of previous models can be used as explanatory variables in subsequent models.*

FIGURE 7: MOBILITY MODELS



Day Pattern and Tour Level Models

Next, day-level and tour-level models are run, as shown in Figure 8. The daily activity pattern model predicts the general activity pattern type for every household member. Then mandatory tours[24] are generated for workers and students, and the tours are scheduled (their location is already predicted by the work/school location choice model). A school pickup/dropoff model matches children with school activities with chaperones responsible for escorting them to and from school. Fully joint tours are generated at a household level, their composition is predicted (adults, children, or both), and participants are determined. The primary destination of fully joint tours is predicted, and the tours are scheduled. Next, non-mandatory tours are generated, their primary destination is chosen, and they are scheduled. At-work sub-tours are generated for workers with a work tour, its primary destination is chosen, and the tour is scheduled.

At this point all tours are generated, each tour's primary destination is known, as well as the tour departure time and arrival time period. Next a vehicle availability model is run for each tour, which predicts which household vehicle would be used for each vehicle-related tour mode. A tour mode is chosen for each tour, then 0 or more intermediate stops are chosen.

[24] A tour is a sequence of trips that start and end at an anchor location. Home-based tours start and end at home and Work-based subtours start and end at work. Each tour has a primary destination, at an activity that is the same as the tour purpose, and zero or more intermediate stops, which are activity locations between the anchor location and the primary destination.

Intermediate Stop and Trip Level Models

The flow of stop and trip level models is shown in Figure 9. Each intermediate stop purpose is selected from a distribution. The location of each stop is determined, along with the stop departure (for outbound stops) or arrival (for inbound stops) time period, and a mode is chosen for each trip on the tour. The parking location for auto trips to parking-constrained areas is modeled. Trip lists are aggregated to matrices by period, mode, and market segment for assignment.

FIGURE 8: DAY PATTERN AND PRIMARY ACTIVITY/TOUR MODELS

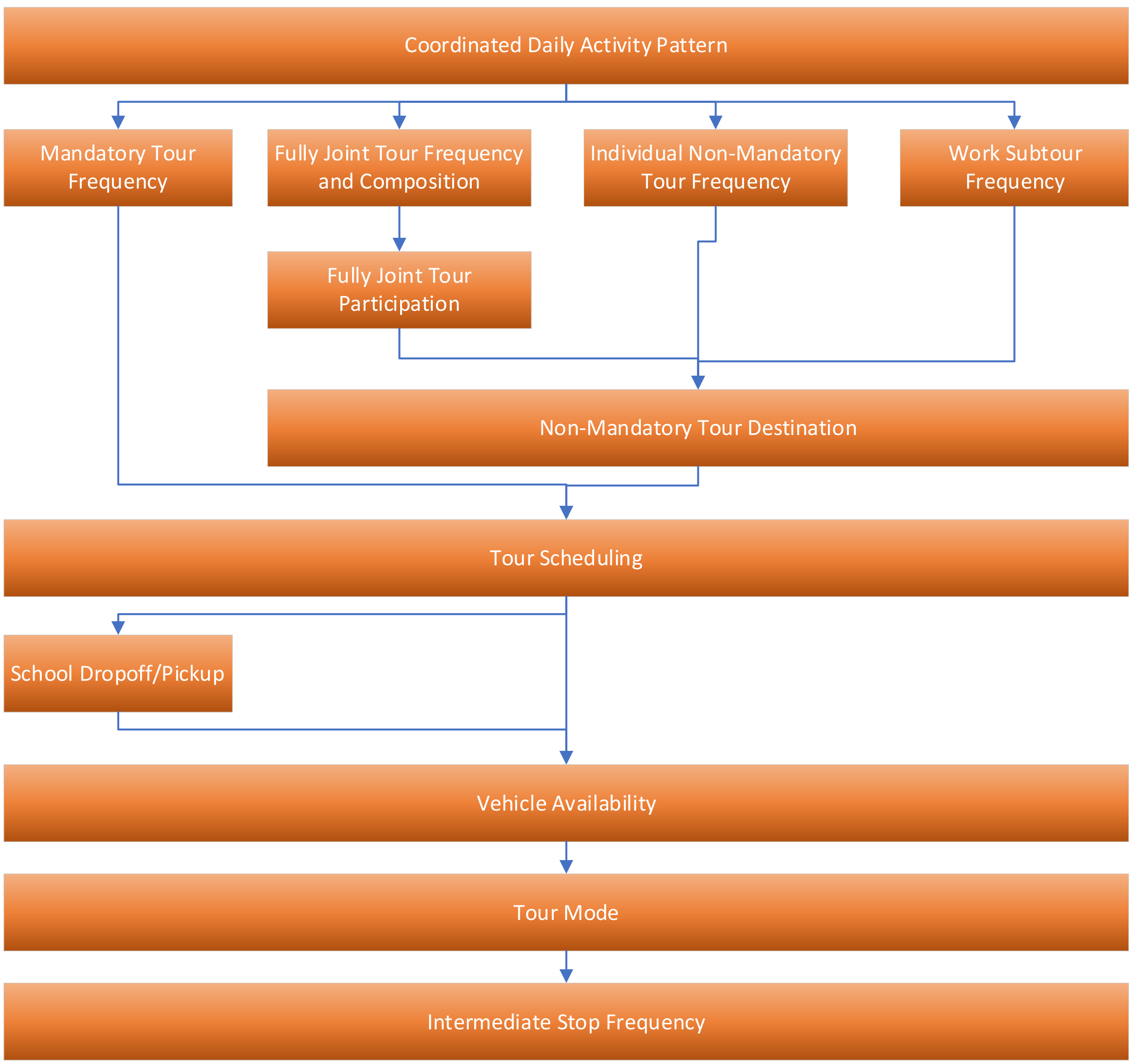
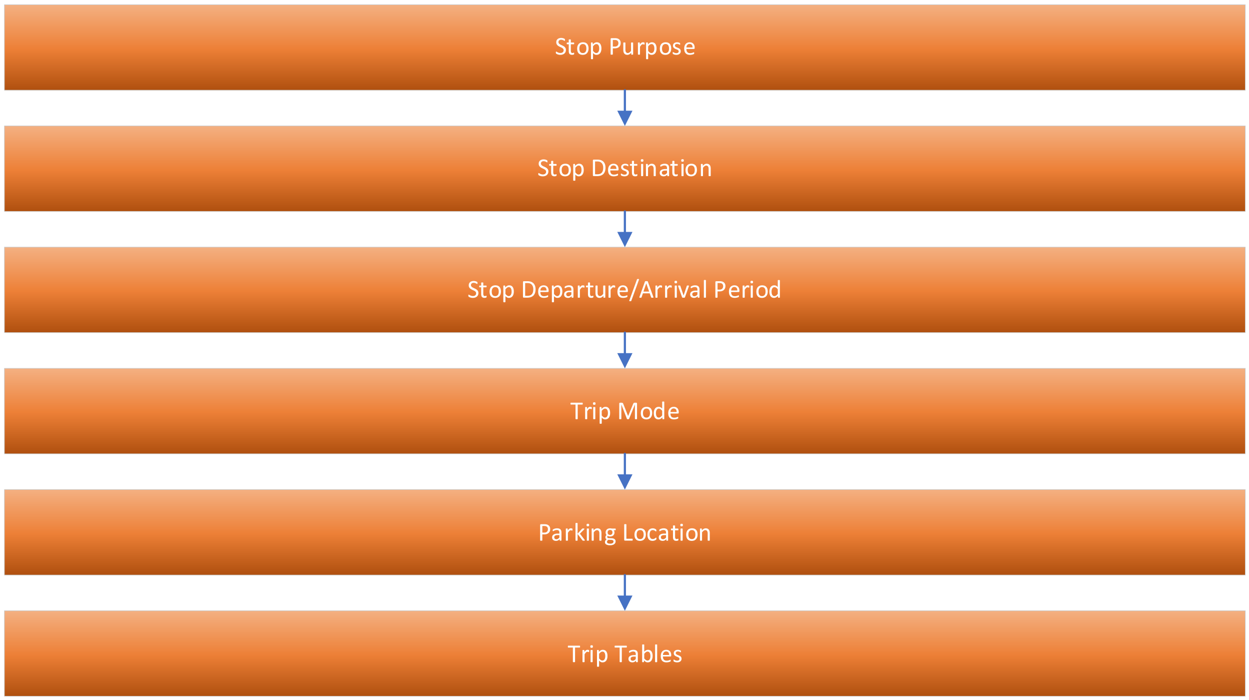


FIGURE 9: STOP AND TRIP LEVEL MODELS



3.9 ACTIVITYSIM DETAILED DESCRIPTION

Below we describe each ActivitySim component in more detail. We list the alternatives to be considered by the model, the model form, and the typical variables used in each model. We note areas for potential refinement of the model where relevant. Note that the outcomes from each model are recorded in the relevant file (household level decisions are recorded to the household file, tour level decisions are recorded to the tour file, and so on) and can be used as explanatory variables in subsequent models.

Accessibilities

Aggregate Accessibilities

The accessibilities model computes aggregate (zonal) measures of accessibility used by the downstream models. The accessibility measure is the equivalent of a destination choice logsum where the level-of-service variable is restricted to a certain mode (or set of modes) and the size term is restricted to a certain employment variable. The equation is shown below.

$$A_i = \ln\left[\sum_{j=1}^{I} S_j \times \exp\left(-\gamma c_{ij}\right)\right]$$

where:

A_i is the accessibility for origin TAZ i

S_j is the size term for destination TAZ j

C_{ij} is the cost of travel from origin TAZ i to destination TAZ j, expressed as a generalized cost including time and cost.

γ is a parameter indicating the sensitivity to the cost of travel

Accessibilities are typically used in mobility models, the daily activity pattern model, and tour frequency models. Because they are mode-specific, they are useful for explaining the influence of changes in level-of-service for certain modes, such as transit, that may not affect the disaggregate logsum as much because it weights the transit utility by the probability of use.

We propose to create the following output accessibility measures by zone:

- autoPeakRetail – the accessibility by auto during peak conditions to retail employment
- autoPeakTotal – the accessibility by auto during peak conditions to all employment
- autoOffPeakRetail – the accessibility by auto during off-peak conditions to retail employment
- autoOffPeakTotal – the accessibility by auto during off-peak conditions to all employment
- transitPeakRetail – the accessibility by transit during peak conditions to retail employment
- transitPeakTotal – the accessibility by transit during peak conditions to all employment
- transitPeakHouseholds – the accessibility by transit during peak conditions to all households
- transitOffPeakRetail – the accessibility by transit during off-peak conditions to retail employment
- transitOffPeakTotal – the accessibility by transit during off-peak conditions to all employment
- railOffPeakRetail – the accessibility by rail during off-peak conditions to retail employment
- railPeakTotal– the accessibility by rail during peak conditions to retail employment
- walkRetail – the accessibility by walking during all time periods to retail employment
- walkTotal – the accessibility by walking during all time periods to all employment
- bikeRetail – the accessibility by biking during all time periods to retail employment
- bikeTotal – the accessibility by biking during all time periods to all employment
- totalPeakAcc – Multimodal total accessibility in the peak period calculated by a weighted average of peak auto, transit, and non-motorized accessibilities to all employment

Bike accessibilities would utilize bicycle logsums as the measure of accessibility. If bike logsums are built by bike comfort level, we suggest using ‘cautious’ accessibilities in order to weight more heavily bike infrastructure availability in the logsum.

Aggregate accessibilities are created in the accessibilities.csv configuration file. The file can be easily modified to create different or more aggregate accessibilities depending upon the needs of downstream models.

Disaggregate Accessibilities

Disaggregate accessibilities are calculated by creating a ‘prototype’ synthetic population (and their tours) through ActivitySim’s destination choice models and recording the destination choice logsum for each tour. The logsums from the prototype sample are then merged with the actual synthetic population and used as explanatory variables in downstream models. The advantage of disaggregate logsums over aggregate logsums is that the disaggregate logsums include all alternatives and explanatory variables considered by the full ActivitySim model. Therefore they will respond to many more variables than the aggregate accessibilities consider and are closer to an actual logsum compared to the aggregate accessibility.

In order to create disaggregate logsums, one must define the attributes of the prototypical synthetic population, the location of those households, and the types of tours to generate logsums for. In order to keep the calculations parsimonious, only a limited number of household and person types should be considered, and zonal sampling is used in the case of regions with many (>5k) microzones. The disaggregate_accessibility.yaml file is used to specify the household types used for the prototype population, the tour types, and the spatial sampling method used. The values used to describe the households for the SANDAG application of the disaggregate accessibilities as an example are shown in Table 12 and the person variables are shown in Table 13. As shown in Table 12, households are specified for four income categories and three auto ownership categories (12 different households per zone). Each household has two persons, so the values for each person are listed separately in Table 13.. At the bottom of this table the tours specified for each person are listed. In total, three logsums are generated for each household – one for work tours (generated for person 1), one for shopping tours (generated for person 2), and one for other discretionary tours (also generated for person 2).

The exact household segments and household and person variables, and exactly what tour purposes to build disaggregate logsums for, to be used for the Oregon application can be borrowed from SANDAG or refined. Variables and their values will need to be refined to be consistent with the Oregon model specific inputs, values created in pre-processors, and models applied. After running the disaggregate accessibility component, the results are joined with actual synthetic households based on the location of the household and the income segment. Because the logsums are used in auto ownership models, it is necessary for each household to have access to logsums for each of the auto ownership levels in the prototype population. So each actual synthetic household will get logsums for each of the three tour purposes and each of the three auto ownership levels appended to it. The logsums are named as *purpose*_accessibilities_*autos* where purpose is one of [workplace_location, othdiscr, shopping] and autos is one of [0, 1, 2].

TABLE 12: PROTOTYPE POPULATION FOR DISAGGREGATE ACCESSIBILITIES - HOUSEHOLD VARIABLES

VARIABLE	DESCRIPTION	VALUE(S)

hinccat1	Household income category	1,2,3,4 (4 categories, corresponding to income variable below)
income	Household income in dollars	15000, 45000, 67000, 120000
hworkers	Number of workers	1
auto_ownership	Auto ownership level	0, 1, 2 (3 categories)
persons	Persons in household	2
HHT	Household type	1 – Married couple family household
Bldgsz	Building size	2 - Single family detached

TABLE 13: PROTOTYPE POPULATION FOR DISAGGREGATE ACCESSIBILITIES – PERSON VARIABLES

VARIABLE	DESCRIPTION	PERSON 1 VALUE	PERSON 2 VALUE
pnum	Person number	1	2
age	Person age	35	55
sex	Person Gender	2 - Female	1 – Male
ptype	Person type	1 – Full time worker	4 – Non-working adult
pemploy	Employment status	1 – Full time	3 – unemployed
weeks	Weeks worked per year	1 – 50-52 weeks	0 - none
hours	Hours worked per week	35	0
soc2	Standard occupation classification	11 - Management	0
DAP	Daily activity pattern	“M” - mandatory	“N” – non-mandatory

military	Military classification	4 – not military	4 – not military
pstudent	School attending	3 – not attending	3 – not attending
educ	Education level attained	13 – Bachelors degree	13 – Bachelors degree
grade	Grade level	0 – not attending	0 – not attending
timeFactorWork	Distributed travel time sensitivity for work purpose	1.0	1.0
timeFactorNonWork	Distributed travel time sensitivity for non-work purpose	1.0	1.0
Tour purposes	Purposes of tours for each household member (note this is not a person variable)	work	Shopping, other discretionary

Mobility Models

License Holding Status

Number of Models: 1
Decision-Making Unit: Person
Model Form: Multinomial Logit
Alternatives: 2 (0: does not hold license, 1: holds license)

This model would predict whether each person in the population holds a driver's license. We suggest a logit model for this variable rather than a lookup table so that the model can consider explanatory variables other than age. For example, mobility status could be included if it is used as an input variable in the synthetic population. In addition, area type, transit accessibility, non-motorized accessibility, and/or density measures could be tested to reflect the influence of non-auto accessibility on the likelihood of holding a driver's license. The alternative-specific constant in the model can be adjusted to test scenarios in which driver license holding decreases in the future due to forces external to the travel model.

The following types of variables should be tested for significance:

- Household attributes (income)

- Person attributes (age, gender)
- Density terms
- Aggregate accessibility terms
- Disaggregate non-work location choice model logsums
- Region-specific variables

The outcome from the model is an input to auto parity (see above) which influences mode choice. It is also an explanatory variable for auto ownership, influences the school pickup/dropoff model, and directly impacts mode choice (non-license holders cannot drive alone). If driver versus passenger is modeled directly, license holding can be used to restrict who can be selected as a driver.

An agency may wish to modify the age-specific constants in the model, to test for potential changes in preferences for license holding by age. These can be introduced as part of sensitivity testing, by creating a set of scenario-specific constants that reflect lower or greater levels of license holding by age than the base-year estimated constants. These constant factors can then be turned on or turned off by setting a property in the constants.yaml file, described in Section 4.6.

Bicycle Comfort Level

Number of Models:	1
Decision-Making Unit:	Person
Model Form:	Multinomial Logit
Alternatives:	Four (1: Does not bike, 2: Cautious, 3: Confident, 4: Fearless)

This model predicts the bicycle comfort level for each person in the population. We recommend a logit model so that bike comfort level can be influenced by socio-economic variables such as age and gender and take into account accessibility effects – particularly with respect to bicycle infrastructure. The model will have four alternatives, relating to levels of bicycle comfort as outlined by Roger Geller[25] and as shown in Table 14 along with how each bicycle comfort level relates to bike path choice and how bicycle comfort level can be identified by responses to household travel survey questions. The initial classification is based on the reported frequency of biking. The other variables are used to recode the initial classification. Most of those reclassifications are meant to downgrade to the previous classification of a subsequent response is inconsistent with the current classification. For example, someone might report that they are a frequent biker (4+ days per week), leading to an initial classification of "Strong and Fearless", but if we don't observe at least 2 bike trips per day, they would be downgraded to "Enthused and Confident". Similarly, they would have to report being comfortable biking under any conditions..Distance traveleds by bicycle could also be used to differentiate bike comfort level.Distance traveleds by bicycle could also be used to differentiate bike comfort level.

---

[25] https://nacto.org/references/four-types-of-cyclists/, references August 23, 2023.

Bike ownership is listed as a potential classification variable, but modeling bike comfort level precludes the need to explicitly model human-powered bike ownership, because a household with only No Way No How members would not own a bike, and the presence of any member in any other category would require at least one bike. So, modeling bike ownership is likely redundant, but we do recommend modeling e-bike and e-scooter ownership via sampling from a probability.

Explanatory variables would include household variables such as presence of children and household income, person variables such as age and gender, land-use variables, origin-based accessibilities, (including bike-only accessibility variables), etc. Network variables such as miles of bike lanes within a certain radius of the household could be tested. This would provide some feedback to bike investments in terms of class membership.

Each of the bike comfort levels other than "No Way No How" would be associated with a path logsum, whose path weights correspond to the broad indications in the table. Ideally, those weights would be estimated in a route choice model, using the GPS traces for each segment. However, that effort could go well beyond what would be needed for initial model development. An initial set of logsums could be specified by a Delphi approach, starting from the bike route choice path weights developed by Oregon Metro, with sensitivity parameters for each path attribute determined by a panel of experts.

TABLE 14: BICYCLE COMFORT LEVELS AND POTENTIAL CLASSIFICATION BASED ON HOUSEHOLD SURVEY RESPONSES

Segment	Path Implications	Initial Classification based on Reported Frequency	Reclassification based on Actual Frequency	Bike attitude	Bike Comfort	Bike Ownership
No way no how	No path – bike unavailable	Never	Reclassify to Interested but Concerned if at least one biking trip on any surveyed day			
Interested but concerned	Negative weights on mixed traffic, flow, etc.	Less than monthly to 1-3 days per month		Reclassify to No Way No How if respondent answers Not an option or not of interest		
Enthused and confident	Less negative weight on mixed traffic, flow	1 day per week-3 days per week	Reclassify to Interested but Concerned if surveyed 3 days and no observed bike trips.		Reclassify to Interested but Concerned if not Very Comfortable or Somewhat Comfortable biking on a major street with four lanes and no bike lane	Reclassify to Interested but Concerned if no bikes or e-bikes owned by household
Strong and fearless	Ride on virtually any facility, little difference in path weights for non-bicycle lane paths	4+ days per week	Reclassify to Enthused and Confident if less than 2 trips per day. Reclassify to Interested but Concerned if surveyed 3 days and no observed bike trips.		Reclassify to Enthused but Concerned if not "Very comfortable" biking in all categories	

Autonomous Vehicle Ownership Model

Number of Models:	1
Decision-Making Unit:	Households
Model Form:	Multinomial Logit or Household Pre-processor
Alternatives:	Two (0: Does not own, 1: Owns)

The autonomous vehicle ownership model determines whether the household owns one or more fully autonomous vehicles (AVs). It is used to precondition the auto ownership choice model, as it is likely that households that own AVs would be less likely to own multiple cars.

Because there is no observed data on AV ownership, this model is asserted. The specification used for the MWCOG ActivitySim model has just a few explanatory variables, including the following:

- Household income: higher income households are more likely to own AVs
- Age of head of household: generally younger households are more likely to own AVs. Since head of household is not well defined, we suggest instead using the age of the oldest household member.
- Total commute time of all workers in the household: The peak one-way auto time between home and the usual workplace location, summed across all workers in the household. This variable is positively correlated with AV ownership due to increased convenience and comfort of owning an AV for commuting.

The coefficients in the model can be asserted or based on stated preference survey data. A set of constants turn the AV ownership option off (for base-year model runs and future scenarios where AV ownership is asserted to be zero) or can be used to calibrate the AV share for a future scenario where the private AV ownership is greater than zero.

Alternatively, the household pre-processor can be used to sample from a fixed percentage of AV ownership specified in the constants.yaml file. The advantage of this approach is that it is easier to set the percentage than to calibrate a constant term. The disadvantage is that it is harder to correlate AV ownership with household and/or person attributes via specification of percentages.

If the household is predicted to own an AV, then the household’s auto ownership level can be reduced compared to a household that does not own an AV, and the 0-auto choice is turned off for the household. In the vehicle type choice model, the first vehicle selected by the household must be an AV.

Note that this model can be ‘turned off’ for the base-year or for agencies that do not wish to include autonomous vehicle functionality in their model system.

Auto Ownership Model

Number of Models:	1
Decision-Making Unit:	Households
Model Form:	Multinomial Logit
Alternatives:	Five (0, 1, 2, 3, 4+ autos)

The car ownership models predict the number of vehicles owned by each household. It is formulated as a choice model with five alternatives, including “no cars”, “one car”, “two cars”, “three cars", or "four or more cars”. The following variables should be tested in model estimation:

- Number licensed drivers in the household
- Number of persons in the households by age group and presence of children by age
- Number of workers in the household
- Household income
- Density variables
- Aggregate mode-specific accessibilities to employment
- Disaggregate non-mandatory accessibilities, by auto ownership (corresponding to auto ownership choices)
- Sum of disaggregate work location choice logsum across all workers in the household
- Region specific variables

Nesting structures should be tested in estimation (0 versus 1+ alternatives, etc.).

As noted above, constants can be asserted to reflect lower levels of auto ownership if a household owns an AV, to reflect the potential use of the AV for serving trips by multiple household members via vehicle repositioning. For example, in the MWCOG model, the 0-auto and 4+ alternative is turned off if a household owns an AV, and the 2 and 3 auto alternatives are made less likely by 50% and 80% respectively. These constants do not affect model estimation but should be applied in the coefficient file in application.

Work-from-Home Model

Number of Models:	1
Decision-Making Unit:	Person
Model Form:	Multinomial Logit
Alternatives:	2 (0: does not work from home, 1: works from home)

The work from home model determines whether a worker works from home or not. Workers who work from home either have no out-of-home workplace location, or they have an out-of-home workplace that is too far to commute to on a regular or semi-regular basis. For example, a self-employed worker who does freelance website design out of their home would qualify, as would a worker who is lives in Portland but is employed by a company with offices in San Francisco, and who would intermittently fly to San Francisco for occasional on-site visits.

The model is run before work location choice. The model can include the following types of variables:

- Household attributes (income, presence of children, presence of non-working adults)
- Person attributes (age, gender, full or part-time work status, industry category)
- Density terms
- Aggregate accessibility terms
- Disaggregate work location choice model logsum
- Region-specific variables

External Worker Identification Model

Number of Models: 1
Decision-Making Unit: Person
Model Form: Multinomial Logit
Alternatives: 2 (0: regular workplace is internal to region, 1: regular workplace is external)

This model predicts which resident workers have a regular workplace that is out of the region but within a reasonable commute distance from home. This distance needs to be defined based on analysis of survey data but should encompass the 99th percentile distance for persons who drive outside the region for work commutes. The model is only applied to workers with a regular out-of-home workplace within commute distance. Workers who have a workplace that is not within a reasonable commute distance will be modeled as work-from-home workers. The following explanatory variables should be tested in estimation:

- Household income
- Person variables (age, gender, license status, industry of worker)
- Distance to closest external station
- Size (base-year count of IE work tours) at external station from statewide model
- Region specific variables

If a worker is predicted to be an external worker, they will choose an external station from the external workplace location choice model to determine their work zone. The identification of the worker as an external worker should be tested as an explanatory variable in the Coordinated Daily Activity Pattern model, the Mandatory Tour Frequency model, the Mandatory Tour Scheduling Model, and the Intermediate Stop Frequency models at a minimum. We expect workers who work externally to the region to generate less work tours and have a longer work tour duration than workers whose workplace is internal to the region.

External Worker Workplace Location Choice Model

Number of Models:	1
Decision-Making Unit:	Person
Model Form:	Multinomial Logit (destination choice)
Alternatives:	Equal to number of external stations in model area

This destination choice model predicts the external station for each worker who works externally to the region. The choices are ‘dummy’ zones corresponding to external stations. For each zone there must be a TAZ and an MAZ. The size of each zone is equal to the number of resident workers who commute to a location outside the modeling region using that external zone from the Statewide Integrated Model (SWIM). The size terms should be compared carefully to traffic counts and other data sources, such as big data, prior to developing estimation data bundles for model estimation. The other key terms in the model are distance to each external station and the mode choice logsum to each external station. For a full description of the destination choice utility function, see Mandatory Activity Location Choice below.

The external zone is saved in the person table as the workplace for the worker. Each work tour generated by the worker is attracted to the external zone.

Mandatory (workplace/university/school) Activity Location Choice

Number of Models:	4 (Work, K-8, 9-12, University)
Decision-Making Unit:	Workers for Work Location Choice; Persons attending K-8 for K-8 students model, Persons attending High School for High School model; University Students for University Model
Model Form:	Multinomial Logit
Alternatives:	Microzones

A workplace location choice model assigns a workplace microzone (MAZ) for every out-of-home worker in the synthetic population according to a multinomial logit destination choice model. Since mode choice logsums are required for each destination, a two-stage procedure is used for all destination choice models in ActivitySim in order to reduce computational time (it would be computationally prohibitive to compute a mode choice logsum for each of approximately 3,000 zones and every worker in the synthetic population). In the first stage, a simplified destination choice model is applied in which all zones are alternatives. The variables in this model are the distance to the MAZ, distance interaction terms and the size term of the MAZ. The logsum term is not used in the simple model used to sample alternatives. This model creates a probability distribution for all possible alternatives (but zones with no employment are not sampled). A set of thirty alternatives are sampled from the probability distribution and these alternatives constitute the choice set in the full destination choice model. Mode choice logsums are computed for these alternatives and the destination choice model is applied. A discrete choice of MAZ is made for each out-of-home worker from this more limited set of alternatives. The same general structure is used for all destination choice models in ActivitySim.

Accessibility is measured by the tour mode choice logsum which considers outbound travel in the AM peak period travel and return travel in the P.M. period, the auto ownership of the household in the input synthetic population, and other person and household characteristics (see ***Tour Mode Choice Model***, below). The mode choice logsum represents the total ease of travel between two zones across all available modes. Non-linear distance terms are used to represent the portion of accessibility not represented by the mode choice logsum term and help to improve the goodness of fit of the home-work distance frequency distribution.

Work location choice size terms can be estimated by cross-tabulating expanded workers in the Census Public Use Microdata Sample (PUMS) by employment category (based on NAICS in the PUMS data) and worker occupation category. The largest occupation category is set as the base category with a size term of 1.0, and coefficients for other occupation categories are set to the ratio of workers in that category to the base category.

A university location choice model assigns a university (large college or small college) location for every university student in the synthetic population. The University school location choice model parameters include university mode choice logsum, distance, and a size term consisting of university enrollment.

A school location choice model assigns a school location for every person enrolled in K-12 school in the synthetic population. The size term in this model is enrollment by type (K-8 and high school), including both public and private grade schools. Grade school parameters include mode choice logsum, distance, and size term.

The application procedure utilizes an iterative simulation-based constraint mechanism in order to match workers to input employment totals. A previous shadow pricing procedure would calculate zone-specific constants by taking the ratio of the share of the total employment of each zone to the share of the total workers in each zone (see **EQUATION 1**) by market segment (household income, for work location choice). This procedure would require many iterations and was only guaranteed to converge to the relative size term of the zone rather than the relative total jobs in the zone.

In the new procedure, all workers are run through the model. Next, the total relative workers by work zone are compared to total relative jobs in the zone. If the share of workers is higher than a minimum threshold than the share of jobs, workers are randomly selected from the zone and re-simulated in the next iteration, after making all over-subscribed zones unavailable. This procedure converges much faster than shadow pricing and eliminates the need to maintain a shadow price file. The model iterates for a user-defined set of iterations, or until the pre-specified convergence criterion is reached. We find it sufficient to run the simulation mechanism for a maximum of 10 iterations, or if 99% of MAZs are within 5% relative share of workers to jobs.

EQUATION 1: SHADOW PRICE

$$Price_{inc,j} = \ln\left(\frac{Employment_{inc.j}}{\sum_{j=1}^{n} Employment_{inc.j}} \Big/ \frac{Workers_{inc.j}}{\sum_{j=1}^{n} Workers_{inc.j}}\right)$$

Vehicle Type Model

Number of Models:	1
Decision-Making Unit:	Household
Model Form:	Multinomial Logit
Alternatives:	Two phase model: First phase is multinomial logit whose alternatives are combinations of 9 body types and 20 age categories; not all categories available for each year. Second phase is Monte Carlo simulation for user-specified probabilities of 5 categories of fuel type for each combination of age and body type.

This model predicts body type, fuel type, and age of each vehicle owned by a household. There are five body types in the base year: Car, van, sport-utility vehicle (SUV), pickup, and motorcycle. In future year scenarios where AVs are modeled, the car, van, SUV, and pickup categories can be expanded into AVs and non-AVs, for a total of nine body types. There are five fuel types, as follows: Gas, diesel, hybrid, plug-in hybrid electric vehicle (PEV), and battery-electric vehicle (BEV). And there are a maximum of 20 age categories (capped by 20+ years old).

There are two version of the vehicle type choice model available in ActivitySim. The “Option 4” model is a simultaneous model whose alternatives are a combination of the above dimensions, for a maximum of (9*5*20) 900 alternatives. However, not all combinations are available; for example, it is widely assumed that AVs would only be BEVs, which trims the choice set to (5*5*20 + 5*20) 600 alternatives. The choice set is further trimmed in the base year due to lack of hybrid, PEV and BEVs in certain years, and potentially trimmed in future years by lack of gas, diesel, and hybrid options in certain years.

The other model ("Option 2") is a two-stage model where the first stage is a multinomial logit choice of vehicle body type and age. The second stage is a Monte Carlo simulation of fuel type from a probability distribution that is segmented by vehicle age and body type. This probability distribution must be provided by the user; a base-year probability distribution by five non-autonomous vehicle body types and 20 age categories is available from the initial estimation. This version of the vehicle type choice model will be implemented upon review and agreement by the OMSC.

This model was estimated using 2017 National Household Travel Survey (NHTS) data, a survey of roughly 130,000 households, 275,000 persons, and 247,878 vehicles. Explanatory variables from this dataset include household and person attributes such as household income, household size, number of children, and distance to work for workers. The data was supplemented with US Environmental Protection Agency's (EPA's) fuel economy testing database, which includes fuel economy rating, emissions, electric vehicle range, and time to charge for PEV and BEVs. These variables were averaged by body type, fuel type, and year. Vehicle prices for each combination were derived from California Energy Commission (CEC) and Bureau of Transportation Statistics (BTS) data, and total number of charging stations by metropolitan area was obtained from US Department of Energy, Alternative Fuels Data Center. All of this data was used to estimate a rich multinomial logit model with variables that include number of makes and models of each vehicle type available for sale (logged), fuel economy and range, round-trip commute distance in the household, vehicle price, and the number of public charging stations. The model includes a set of constants by age, body type, and fuel type, as well as constants for each NHTS region including Oregon. The model also considers the types of vehicles already owned by the household – interaction terms that account for the decreased probability of owning multiple vehicles of the same body type.

Because of the complexity of the model, the fact that the model was estimated from a survey that has orders of magnitude more households than will be collected in Oregon, and the dependence upon a number of supplemental datasets, we do not suggest that the model be re-estimated. Rather we suggest that the model be calibrated to vehicle ownership data from Oregon Department of Motor Vehicles and/or Oregon household travel survey data.

The results of the model are used in the Vehicle Availability Model, which predicts which vehicle would be used for each auto alternative in mode choice (if auto were chosen) and uses the auto operating cost for that vehicle in mode choice utility calculations. It can also be used to calculate energy consumption and greenhouse gas emissions after travel is generated by ActivitySim.

Transit Subsidy Model

Number of Models: 1

Decision-Making Unit: Workers and students
Model Form: Multinomial Logit
Alternatives: Two (0: No transit subsidy, 1: has subsidized transit)

The transit subsidy model predicts whether workers and students have transit subsidized by their employer, have transit paid for as part of their tuition or have free transit as a student. Explanatory variables to be tested include:

- Household variables (income, autos owned)
- Person variables (work/school status, age, industry of worker, license holding status)
- Difference between auto travel time and transit travel time to work (if worker) or school (if student)
- Transit accessibility from work location to all households (transitPeakHouseholds)
- Average daily parking cost at work or school location
- Density in work zone
- Region specific variables

Persons who have their transit partially or fully subsidized have a reduced or free transit fare in *work/school* tour and *work/school* trip mode choice.

Transit Pass Ownership Model

Number of Models: 1
Decision-Making Unit: All persons
Model Form: Multinomial Logit
Alternatives: Two (0: No transit pass, 1: has transit pass)

The transit pass model predicts whether persons own a monthly, semester, or yearly transit pass. Explanatory variables to be tested include:

- Household variables (income, autos owned)
- Person variables (work/school status, age, license holding status)
- Difference between auto travel time and transit travel time to work (if worker) or school (if student)
- Average daily parking cost at work or school location
- Whether person receives a transit subsidy
- Per trip equivalent value of transit pass (e.g. cost of pass divided by average number of transit rides in time period)

- Density in work zone
- Region specific variables

Persons who own a transit pass see no transit fare in tour or trip mode choice. This is because the cost of the pass is already accounted for in this model, and therefore including it in mode choice would be double-counting. Note that it is unlikely that the cost coefficient can be reliability estimated due to lack of variation in cost, though joint estimation may solve this problem. If not, the cost coefficient can be fixed to the coefficient used in trip mode choice models since the cost will be scaled to a per-trip cost (thus ensuring reasonable elasticity with respect to cost).

Free Parking Eligibility Model

Number of Models:	1
Decision-Making Unit:	Workers
Model Form:	Multinomial Logit
Alternatives:	Two (0: Worker pays for parking, 1: worker has free parking)

The free parking eligibility model determines whether each worker pays for parking or has free parking provided for them by their employer. Variables to be tested include:

- Household variables (income, autos owned)
- Person variables (age, gender, industry of worker)
- Expected cost of parking in work zone
- Density in work zone
- Region specific variables

Telecommute Frequency Model

Number of Models:	1
Decision-Making Unit:	Persons
Model Form:	Multinomial Logit
Alternatives:	Four (No telecommute/Telecommute less than one day per week, telecommute 1 day per week, telecommute 2 days per week, telecommute 3 days per week, telecommute 4 or more days per week)

The telecommute frequency model predicts, for each worker in the household with a regular out-of-home workplace, whether they participate in a telecommute program. The definition of telecommuting in the ActivitySim model is where a worker has a regular out-of-home workplace but works from home for one or more weekdays rather than commutes to work. The alternatives in the model include no telecommuting or telecommutes less than one day per week, telecommutes one day per week, telecommutes 2 days per week, telecommutes 3 days per week, or telecommutes 4 or more days per week.

Explanatory variables to be tested include:

- Household variables (Household income, number of adults, presence of children, autos owned)
- Person variables (work status, school status, industry of worker)
- Distance to work
- Expected parking cost at work

Nesting structures should be explored in estimation.

Telecommute frequency is used as an explanatory variable in the following models:

- Coordinated daily activity pattern model: workers who telecommute are less likely to have a mandatory day pattern and are about equally likely to either engage in a non-mandatory pattern or stay at home
- Work tour frequency: Workers who telecommute with a mandatory pattern are less likely to generate more than one work tour
- Non-mandatory tour frequency: Workers who telecommute are less likely to make multiple non-mandatory tours
- Stop frequency: Workers who telecommute are less likely to make multiple stops per tour

Daily Travel Models

Coordinated Daily Activity Pattern (DAP) Model

Number of Models: 1
Decision-Making Unit: Households
Model Form: Multinomial Logit
Alternatives: 691 total alternatives, but depends on household size (see Table 15)

The DAP model predicts the general daily activity pattern for each person in the household. The DAP is classified by three main pattern types:

- Mandatory pattern (M) that includes at least one of the three mandatory activities – work, university, or school. This constitutes either a workday or a university/school day, and may include additional non-mandatory activities such as separate home-based tours or intermediate stops on the mandatory tours.

- Non-mandatory pattern (N) that includes only individual and/or joint maintenance and discretionary tours. By virtue of the tour primary purpose definition, maintenance and discretionary tours cannot include travel for mandatory activities.
- At-home pattern (H) that includes only in-home activities. At-home patterns are not distinguished by any specific activity (e.g., work at home, take care of a child, being sick, etc.). Cases with complete absence from town (e.g., business travel) are also combined with this category.

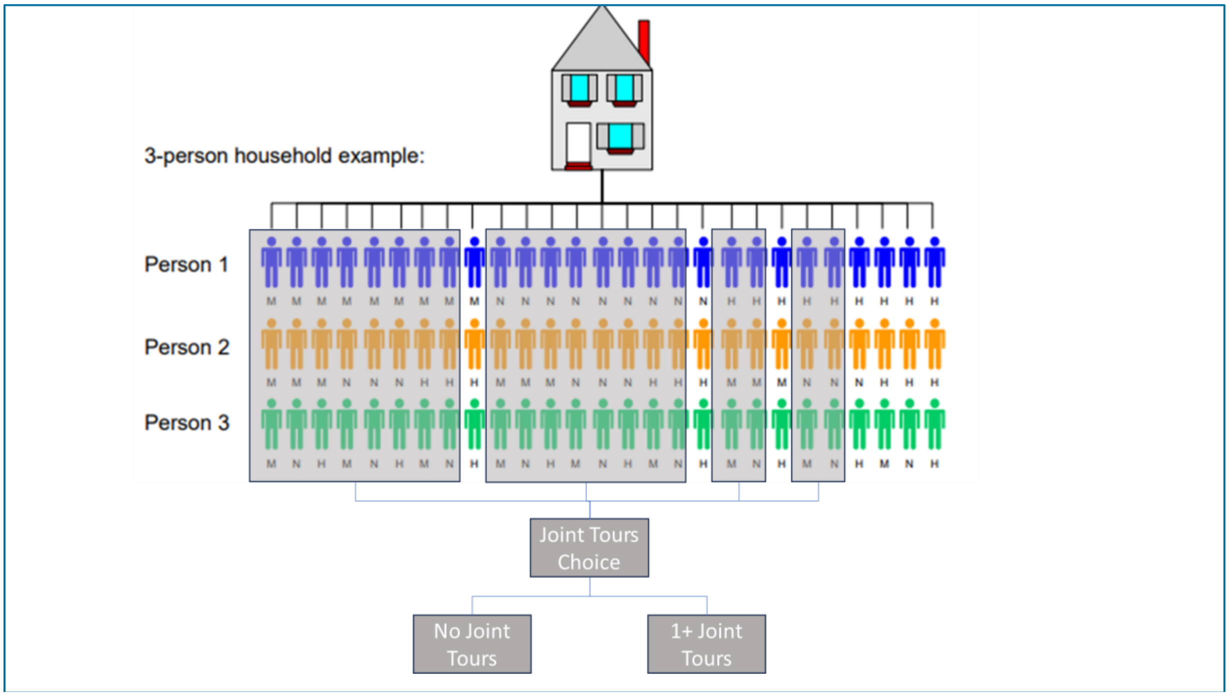
Statistical analysis of survey data across a number of regions has shown that there is an extremely strong correlation between DAP types of different household members, especially for joint N and H types. For this reason, the DAP types for different household members are not modeled independently but rather simultaneously across multiple household members. The total number of possible DAP type combinations is significant for large households.

The model also simultaneously predicts the presence of fully joint tours for the household. Fully joint tours are tours in which two or more household members travel together for all stops on the tour. Fully joint tours are only a possible alternative at the household level when two or more household members have an active (M or N) travel day. The joint tour indicator predicted by this model is then considered when generating and scheduling mandatory tours, in order to reflect the likelihood of returning home from work earlier in order to participate in a joint tour with other household members. The choice structure includes 363 alternatives with no joint travel and 328 alternatives with joint travel, totaling to 691 alternatives as shown in Table 15. Note that the choices are available based on household size. An illustrative graphic of the choice model for a three-person household is shown in Figure 10. Each choice is a vertical selection of activity patterns for each of the three members (27 choices in total). Plus, for each choice in which at least two of the members are active (shaded in the picture), there is also a choice of whether the household has at least one fully joint tour. That corresponds to 20 of the alternatives, so the total number of alternatives for a three-person household is 47.

TABLE 15: ALTERNATIVES IN COORDINATED DAILY ACTIVITY PATTERN MODEL

Household size	Alternatives without joint travel	Alternatives with joint travel	Total alternatives
1	3	0	3
2	3x3=9	3x3-(3x2-1)=4	13
3	3x3x3=27	3x3x3-(3x3-2)=20	47
4	3x3x3x3=81	3x3x3x3-(3x4-3)=72	153
5+	3x3x3x3x3=243	3x3x3x3x3-(3x5-4)=232	475
Total	363	328	691

FIGURE 10: DAILY ACTIVITY PATTERN CHOICES FOR A THREE PERSON HOUSEHOLD



The Coordinated DAP model contains explanatory variables that include person and household attributes, accessibility measures, and density/urban form variables. Since the model features intra-household interactions, a subset of the parameters in the model are specified as interaction terms. These terms are based on the contribution to the total utility of an alternative from either a two-person interaction, a three-person interaction, or an entire-household interaction. For example, the contribution of a two-worker interaction to the utility for each worker to stay home on the simulation day is typically positive, indicating that it is more likely that both workers will attempt to coordinate their days off to engage in recreational opportunities together.

The Coordinated DAP model covers up to five persons in the household. For households with more than five persons, the persons covered by the model are selected according to the following rules:

- First, two workers are selected (full-time worker and part-time worker person types). Full-time workers are ranked higher than part-time workers. If there are more than two workers of a given type, older workers are chosen over younger workers.
- Next, three children are selected (driving-age student, non-driving age student, child too young for school person types). If there are more than three children, they are ranked by

age from youngest to oldest. This assumes that the youngest child in the household has a more significant impact on activity patterns of household members than older children.

- If there are people left after applying the above criteria, they are selected randomly from the remaining members.

Anyone not selected according to the above criteria for a household with more than five persons have their activity pattern chosen by a lookup from a probability distribution which varies by person type.

Individual Mandatory Tour Frequency

Number of Models:	1
Decision-Making Unit:	Persons
Model Form:	Multinomial Logit
Alternatives:	5 (1 Work Tour, 2 Work Tours, 1 School Tour, 2 School Tours, 1 Work/1 School Tour)

Based on the DAP chosen for each person, individual mandatory tours, such as work, school and university tours are generated at person level. The model predicts the exact number and purpose of mandatory tours (e.g., work and school/university) for each person who chose the mandatory DAP type at the previous decision-making stage. Since the DAP type model at the household level determines which household members engage in mandatory tours, all persons subjected to the individual mandatory tour model implement at least one mandatory tour. The model has the following five alternatives:

- One work tour,
- One school tour,
- Two or more work tours,
- Two or more school tours,
- One work tour plus one school tour.

Alternatives with work tours are only available to workers, while alternatives with school tours are only available to students.

DAPs and subsequent behavioral models of travel generation include various explanatory variables that relate to household composition, income, car ownership, location of work and school activities, land-use development, residential and employment density, and accessibility factors.

School Pickup and Dropoff Model

Number of Models:	3 (outbound, inbound, outbound conditional)
Decision-Making Unit:	Households
Model Form:	Multinomial Logit
Alternatives:	157

Dropping off and picking up school children at school accounts for a significant portion of all shared ride travel. According to the 2016-17 household travel survey for San Diego Association of Governments, 24% percent of shared ride trips made by county residents were for part of a tour involving a pickup or dropoff of a child at school.

The school escort model determines whether children are dropped-off at or picked-up from school, simultaneously with the chaperone responsible for chauffeuring the children, which children are bundled together on half-tours, and the type of tour (pure escort versus rideshare). The model is run after work and school locations have been chosen for all household members, and after work and school tours have been generated and scheduled. The model labels household members of driving age as potential 'chauffeurs' and children with school tours as potential 'escortees'. The model then attempts to match potential chauffeurs with potential escortees in a choice model whose alternatives consist of 'bundles' of escortees with a chauffeur for each half tour.

It should be noted that transit on-board surveys, commonly used to generate calibration targets for transit tours and trips, are limited to collecting origin-destination data, resulting in incomplete information for understanding this market on transit. However, for this reason we use the words 'chauffeurs' and 'chaperones' interchangeably in this document, since chauffeurs implies auto mode and chaperone is a broader term that could include non-motorized and transit travel where a child is accompanied by an adult.

The model classifies each child's school tour into three types:

1) No escorting: the child walks, bikes, takes transit, drives, or takes a school bus to/from school (subject to data availability as described above).

2) Pure escort: the child is accompanied to/from school by a chaperone, where the purpose of the chaperone's tour is solely for the purposes of picking up or dropping off the child.

3) Rideshare: the child gets a ride to/from school, where the child is dropped-off or picked-up on the way to or from the chaperone's work or school primary destination.

The model considers up to three children with school tours and up to two potential chaperones in each household. If there are more children in the household with school tours, the model selects the youngest three who are most likely to require escorting.

A weighted calculation is used to select the most likely chaperone in households with more than two potential chaperones. Potential chaperones are limited to the following person types with an active (non-home) travel pattern predicted by the Coordinated Daily Activity Pattern model:

- Person type 1: Full-time worker
- Person type 2: Part-time worker
- Person type 3: University/College student
- Person type 4: Non-Working Adult
- Person type 5: Retired Adult

Currently the procedure calculates a score for each chaperone, and selects the three chaperones with the highest score, according to the following formula:

Weight = β_1**PersonType* + β_2**gender* + β_3**age(0,1)*

Where

PersonType is the person type number from 1 to 5 according to the person types shown above. This variable has the most weight

Gender is 1 for male and 2 for female

Age is a binary indicator equal to 1 if age is over 25 else 0.

β_1, β_2, β_3 are parameters, currently set to 100, 10, and 1 respectively. This has the effect of making person type the overriding consideration for priority, where full-time workers are the least likely to be included in the set of chaperones if there are more than two chaperones of higher person type number (e.g. retired adults, non-workers) available.

The potential choice set is also truncated based on scheduled work and school times for Rideshare tours, where only drivers whose departure time from home (or arrival time back at home) is within 30 minutes of the child requiring escorting are considered as potential combinations of chaperones\escortees. Only drivers with open time windows are allowed as potential chaperones for Pure Escort.

The model uses the following terminology:

- Bundling: Grouping escorted children into travel episodes
- Tasks: The group (bundle) of one or more children that is assigned to a chaperone
- Type: The type of tour that the chaperone uses to drop off or pick up the bundle of children

In summary, the model bundles which children are escorted by which chaperones and by what type of school escort type. Figure 11 shows an example of bundling children by chaperone for a household with three children attending school and two eligible chaperones. The first row of the alternatives shows different combinations of children being escorted. For example, in the left-most alternative, all three children are escorted, whereas in the right-most alternative, no children are escorted. The dark blue boxes under each of the first row alternatives show different combinations of bundling children by tour; in the first box underneath the left-most alternative, both children are escorted on one half-tour (one task). In the next alternative, child 1 and 2 are escorted on one tour whereas child 3 is escorted on another tour (two tasks). In the outbound direction, there are 15 mutually exclusive bundling alternatives and a total of 22 potential escorting tasks.

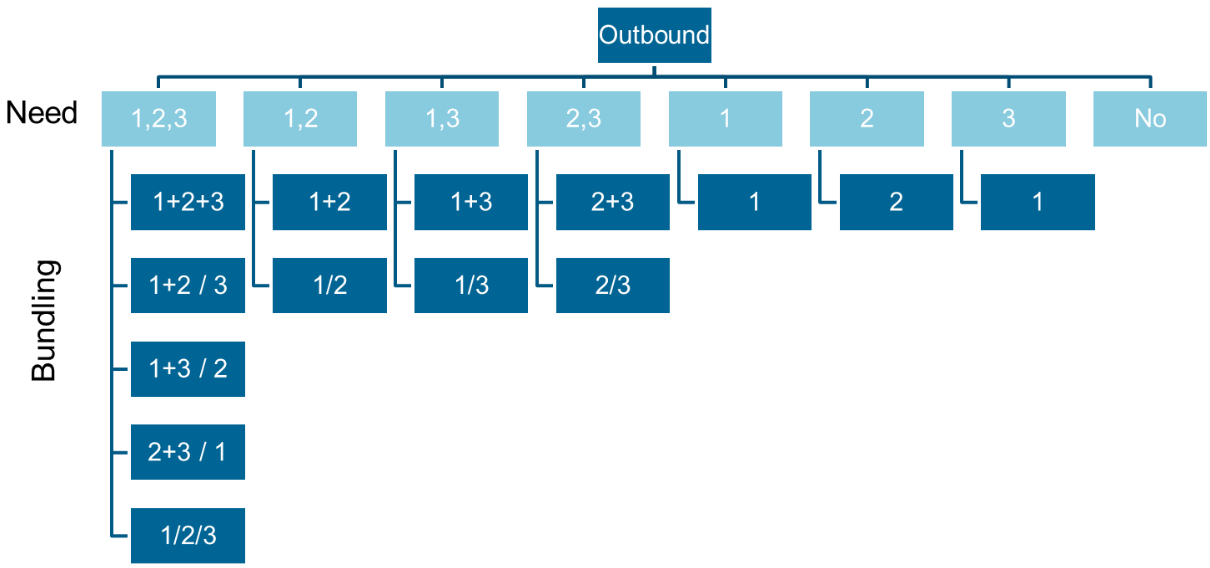


FIGURE 11: SCHOOL DROPOFF/PICKUP MODEL EXAMPLE OF BUNDLING CHILDREN BY HALF-TOUR

Each task is matched with a chaperone by tour type (Pure Escort vs Rideshare), as shown in Figure 12. Table 16 shows the combinations of alternatives for assignment of two tasks to a household with two potential chaperones. The first alternative has both tasks assigned to chaperone 1. The second alternative has both tasks assigned to chaperone two. In the third alternative, chaperone 1 is assigned task 1 and chaperone 2 is assigned task 2, and vice-versa for the fourth alternative. For each case in which there is a task assigned, there is a choice of the two tour types. In other words, there are four total alternatives for each row in Table 16, yielding 16 total chaperone and tour type alternatives in the case of two tasks.

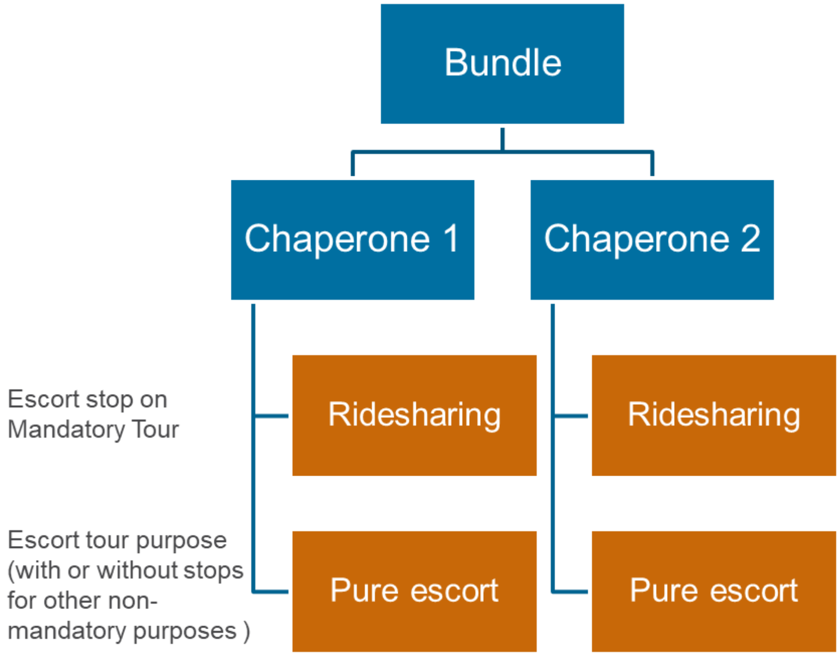


FIGURE 12: CHAPERONE ASSIGNMENT AND TOUR TYPE

TABLE 16: ALTERNATIVES FOR ASSIGNMENT OF TWO TASKS TO CHAPERONES

Alternative	Chaperone 1	Chaperone 2
1	1 & 2	None
2	None	1 & 2
3	1	2
4	2	1

Table 17 shows the number of combinations of tasks assigned to chaperones when assigning three tasks to two chaperones; there are eight total alternatives, and since there are three total tasks, each row in the table expands to six potential chaperone and tour type combinations, for a total of 48 options.

TABLE 17: ALTERNATIVES FOR ASSIGNMENT OF THREE TASKS TO CHAPER ONES

Alternative	Chaperone 1	Chaperone 2
1	1	2 & 3
2	2	1 & 3
3	3	1 & 2
4	2 & 3	1

5	1 & 3	2
6	1 & 2	3
7	1 & 2 & 3	None
8	None	1 & 2 & 3

In total, there are 173 total alternatives by direction, as follows:

1 bundling alternative for no escorting

+ 7 bundling alternatives with 1 task * 2 drivers * 2 escort types

+ 6 bundling alternatives with 2 tasks * 16 chaperone/escort type combinations

+ 1 bundling alternative with 3 tasks * 48 chaperone /escort type combinations

Total 1 + 7 * 2 * 2 + 6 * 16 + 1 * 48 = 173 alternatives

The model is run for each direction separately. Since a strong symmetry effect is observed in the data, the model is run iteratively; first for the outbound direction, then for the inbound direction, and again for the outbound direction, considering the outcomes of the inbound direction. Therefore, there are three models to estimate.

The model specification includes a number of terms that turn off alternatives based upon a set of conditional checks:

- Bundles are turned off based on the number of children going to school.
- Ridesharing escort types of tours are only available for chaperones with a mandatory activity pattern type, which indicates that a work or school tour is generated for them.
- Ridesharing is only available for bundles of alternatives where the child's preferred school departure time and the chaperones preferred mandatory tour departure time is within a certain maximum difference (currently set to 1 period or 30 minutes)
- Pure escorting is not available for any combination of bundles with chaperones if the chaperone's activity pattern is Home
- Pure escorting is not available for any combination of bundles with chaperones if the chaperone is unavailable during the child's preferred departure time to school
- Pure escorting must occur before mandatory travel in the outbound direction for chaperones

The explanatory variables in the model include the following:

- Household variables influencing likelihood of school escorting, such as income, autos available, etc.
- Person variables influencing the likelihood of school escorting, such as the gender of the chaperone, the age of the child being escorted, etc.
- Chaperone disutility for ridesharing – out-of-direction distance and time
- Escortee utility for non-rideshare (non-motorized time to school)
- Bundling utilities (the utility of driving each child separately versus taking children together)

A key aspect of the model is that tours are formed directly from the model results. In the case of multiple pickups or drop-offs on a half tour, the children are arranged by proximity to home; the nearest child is dropped off first or picked up last. The occupancy is calculated based on the number of children in the car for each trip. ActivitySim will explicitly link the drivers to the children and write all relevant information to the tours and trips to the relevant files.

Note that estimation functionality is not currently implemented in ActivitySim for this model; it will need to be added.

Generation of Joint Household Tours

In the current model structure, joint travel for non-mandatory activities is modeled explicitly in the form of fully joint tours, where all members of the travel party travel together from the very beginning to the end and participate in the same activities along the way. Partially joint travel for school trips is handled in the school pickup/dropoff model, as described above. Other types of joint travel like carpooling of workers or ridesharing between members of different households are not explicitly considered currently, though they are handled implicitly through shared-ride alternatives in mode choice.

Each fully joint tour is considered a unit of modeling with a group wise decision-making for the primary destination, mode, frequency, and location of stops, etc. Formally, modeling joint activities involves two linked stages:

- A joint tour frequency, purpose, and composition model that generates the number of fully joint tours by primary activity type and travel party composition made by the entire household. This model is constrained by the presence of at least one fully joint tour at the household level as predicted by the Coordinated Daily Activity Pattern Model. The model considers up to two fully joint tours and three composition levels (adults only, children only, or mixed adults and children). The purpose of the joint party composition model is to narrow down the set of possible person participation choices modeled by the subsequent sub-model.
- A person participation model that predicts which household members participate in the tour.

> Participation choice is modeled for each person sequentially. In this approach, a binary choice model is calibrated for each activity, party composition and person type. The model iterates through household members and applies a binary choice to each to determine if the member participates. The model is constrained to only consider members with available time-windows overlapping with the generated joint tour.

The joint tour frequency, composition, and participation models are described below. Note that estimation functionality has not been implemented yet for these models in ActivitySim.

Joint Tour Frequency and Composition

Number of Models:	1
Decision-Making Unit:	Households
Model Form:	Multinomial Logit
Alternatives:	150 (1 Tour segmented by 5 purpose combinations and three participation categories (5 * 3), plus 2 tours segmented by 5 purpose combinations and three participation categories (5 * 3^2 * 3)

Joint tour frequencies are generated by households and include the number and purposes of the joint tours. The alternatives in the model are combinations of the number of joint tours (1 or 2) by purpose (shop, maintenance, eating out, visiting, discretionary). Note that escort purpose is not available since it violates the definition of fully joint tours (except in cases where the pickup or dropoff is a non-household member, which is rare).

The explanatory variables in the joint tour frequency model include household variables, accessibilities, and other urban form type variables. These variables are associated with one or more alternatives, segmented by tour purpose. One of the most significant variables in the joint tour frequency model is the presence and size of overlapping time-windows, which represent the availability of household members to travel together after mandatory tours have been generated and scheduled. This formulation provides ‘induced demand’ effects on the generation and scheduling of joint tours; the frequency and duration of mandatory tours affects whether joint tours are generated.

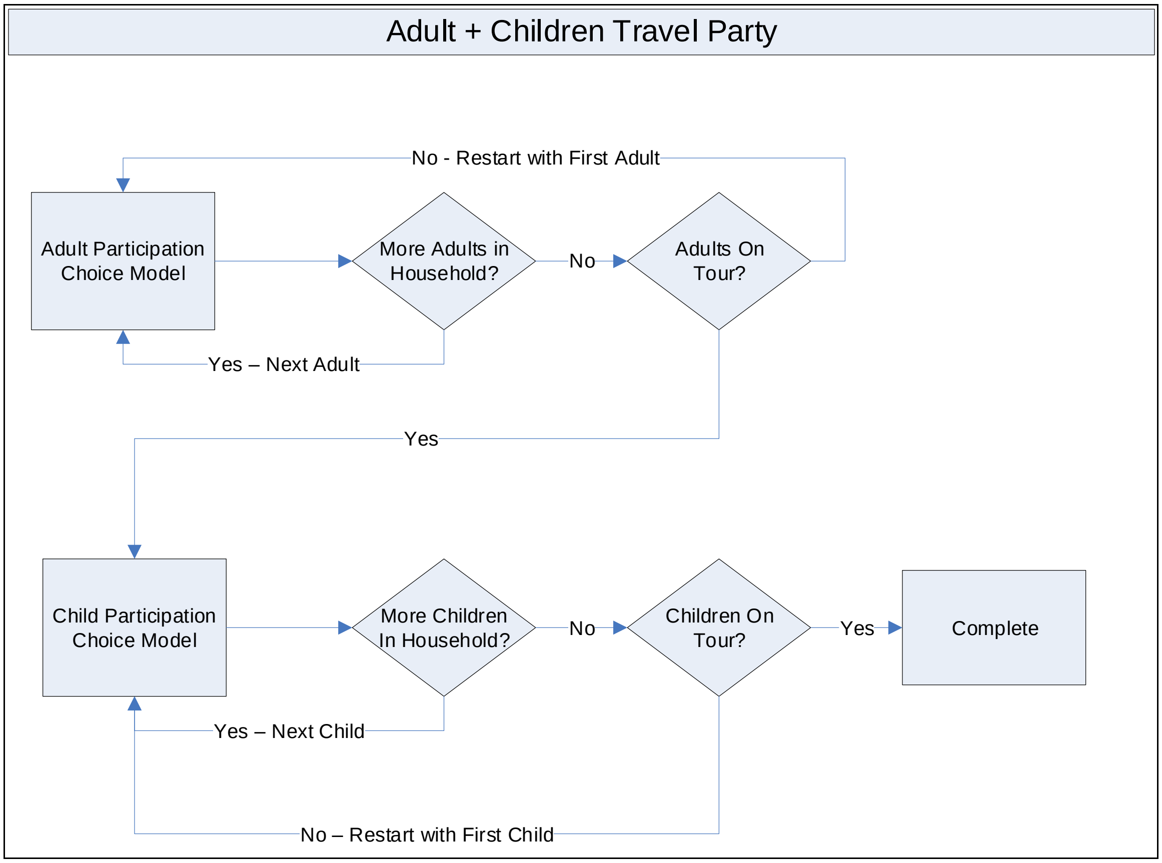
Joint Tour Participation

Number of Models:	1
Decision-Making Unit:	Persons
Model Form:	Multinomial Logit
Alternatives:	2 (Yes or No)

Joint tour participation is modeled for each person and each joint tour. If the person does not correspond to the composition of the tour determined in the joint tour composition model, they are ineligible to participate in the tour. Similarly, persons whose daily activity pattern type is home are excluded from participating. The model relies on heuristic process shown in Figure 13 to assure that the appropriate persons participate in the tour as per the composition model.

A tour starts with no participants. Each person in the household decides whether to participate using a logit model. After all persons in the household have made their participation choice, the software checks to see whether the party type constraint is met. For example, on an adults-only tour, at least two household adults must participate. On a children-only tour, at least two children must participate. On mixed tour, at least one household child and one household adult must participate. Explanatory variables include the person type of the decision-maker, the maximum pair-wise overlaps between the decision-maker and other household members of the same person type (adults or children), household and person variables, and urban form variables.

FIGURE 13: JOINT TOUR PARTICIPATION MODEL



Individual Non-Mandatory Tour Frequency

Number of Models:	7 (segmented by 7 person types)
Decision-Making Unit:	Person
Model Form:	Multinomial Logit
Alternatives:	Depends – typically the most frequently observed combinations of number of individual maintenance and discretionary tours by purpose, approximately between 100 and 150 alternatives

The non-mandatory frequency model generates all non-mandatory individual tours at the person level. This model determines the number of both maintenance and discretionary tours simultaneously, at the person level, by purpose. There are six different types of maintenance and discretionary activities (escort, shopping, other maintenance, eat out, social, other discretionary), and a large number of possible combinations of each (assuming a maximum of 4 individual maintenance/discretionary tours per day, the number of possible combinations is 6^4 = 1,296 alternatives, many of which are not observed in the data). Therefore, a simplification of the alternatives is used in which only the most frequently observed combinations of tours by purpose and number are available in the logit model.

Certain alternatives are defined as “one or more tours” of a certain purpose. If such alternatives are chosen, a subsequent frequency model determines the exact number of tours for those cases (either 1 or 2), based on the person type and the number of mandatory and fully joint tours already generated for the decision-maker.

At-Work Sub-Tour Frequency

Number of Models:	1
Decision-Making Unit:	Persons
Model Form:	Multinomial Logit
Alternatives:	6 (None, 1 eating out tour, 1 business tour, 1 maintenance tour, 2 business tours, 1 eating out tour + 1 business tour)

Work-based sub-tours are modeled last and are relevant only for those persons who implement at least one work tour. These underlying activities are mostly individual (e.g., business-related and dining-out purposes), but may include some household maintenance functions as well as person and household maintenance tasks. There are six alternatives in the model, corresponding to the most frequently observed patterns of at-work sub-tours. The alternatives define both the number of at-work sub-tours and their purpose. Explanatory variables include household and person attributes, duration of the parent work tour, the number of joint and individual non-mandatory tours already generated in the day, and accessibility and urban form variables.

Non-Mandatory Tour External Identification Model

Number of Models:	1
Decision-Making Unit:	Tour
Model Form:	Multinomial Logit
Alternatives:	2 (0: tour destination is internal to region, 1: tour destination is external)

This model predicts whether the tour destination is outside the region. The following explanatory variables should be tested in estimation:

- Household income
- Person variables (age, gender, license status)
- Distance to closest external station
- Size (base-year count of IE non-work tours) at external station from statewide model

- Region specific variables

If a tour is predicted to be external, it will choose its primary activity destination from one of the external stations. The identification of the tour as external should be tested as an explanatory variable in the Tour Scheduling Model and the Intermediate Stop Frequency model. We expect external tours to have a longer work tour duration than internal tours.

External Non-Mandatory Tour Destination Choice Model

Number of Models: 1
Decision-Making Unit: Tour
Model Form: Multinomial Logit (destination choice)
Alternatives: Equal to number of external stations in model area

This destination choice model predicts the external station for each external non-mandatory tour. The choices are 'dummy' zones corresponding to external stations. For each zone there must be a TAZ and an MAZ. The size of each zone is equal to the number of external non-mandatory tours who commute to a location outside the modeling region using that external zone from the Statewide Integrated Model (SWIM). The size terms should be compared carefully to traffic counts and other data sources, such as big data, prior to developing estimation data bundles for model estimation. The other key terms in the model are distance to each external station and the mode choice logsum to each external station.

Non-Mandatory Tour Primary Destination Choice

Number of Models: 7 (Escort, Shop, Other Maintenance, Social/Visit, Eating out, Other discretionary, At-work subtour)
Decision-Making Unit: Person
Model Form: Multinomial Logit
Alternatives: Microzones

The non-mandatory tour primary destination choice model determines the location of the tour primary destination. The model works at an MAZ level, and sampling of destination alternatives is implemented in order to reduce computation time. Explanatory variables include household and person characteristics, the tour purpose, logged size (i.e., attraction) variables, round-trip mode choice logsum, distance, and other variables. Note that the mode choice logsum used is based on a 'representative' time period for individual non-mandatory tours, since the actual time period is not chosen until later in the system.

Tour Time of Day Choice

Number of Models: One for each purpose and one for joint tours
Decision-Making Unit: Persons

Model Form: Multinomial Logit
Alternatives: 1176 (48 half-hour periods^2/2 + 48 periods/2)

The tour time of day choice model selects the departure time from home and arrival time back at home simultaneously (or for work subtours, the time departing work and returning to work). Note that it is not necessary to select the destination of mandatory tours, as this has already been determined. The model is a discrete choice construct that operates with tour departure-from-home and arrival-back-home time combinations as alternatives. The utility structure is based on “continuous shift” variables and represents an analytical hybrid that combines the advantages of a discrete-choice structure (flexible in specification and easy to estimate and apply) with the advantages of a duration model (a simple structure with few parameters, and which supports continuous time). The model has a temporal resolution of 30 minutes, for a total of 48 half-hour periods. The model utilizes direct availability rules for each subsequently scheduled tour, to be placed in the residual time window left after scheduling tours of higher priority. This conditionality ensures a full consistency for the individual entire-day activity and travel schedule as an outcome of the model.

A unique condition applies when applying the time-of-day choice model to joint tours. That is, the tour departure and arrival period combinations are restricted to only those available for each participant on the tour, after scheduling mandatory activities. Once the tour departure/arrival time combination is chosen, it is applied to all participants on the tour.

The model utilizes household, person, and zonal characteristics, most of which are generic across time alternatives. However, network level-of-service (LOS) variables vary by time of day and are specified as alternative-specific based on each alternative’s departure and arrival time. By using generic coefficients and variables associated with the departure period, arrival period, or duration, a compact structure of the choice model is created, where the number of alternatives can be arbitrarily large depending on the chosen time unit scale, but the number of coefficients to estimate is limited to a reasonable number. Duration variables can be interpreted as “continuous shift” factors that parameterize the termination rate in such a way that if the coefficient multiplied by the variable is positive, this means the termination rate is getting lower and the whole distribution is shifted to the longer durations. Negative values work in the opposite direction, collapsing the distribution toward shorter durations.

The tour scheduling model is placed after destination choice and before mode choice. Thus, the destination of the tour and all related destination and origin-destination attributes are known and can be used as variables in the model estimation.

The choice alternatives are formulated as tour departure from home/arrival at home in hour combinations (g,h), and the mode choice logsums and bias constants are related to departure/arrival periods (s,t). Tour duration is calculated as the difference between the arrival and departure hours ($h-g$) and incorporates both the activity duration and travel time to and from the main tour activity, including intermediate stops.

The tour time-of-day (TOD) choice utility has the following general form:

$$V_{gh} = V_g + V_h + D_{h-g} + \mu \ln\left(\sum_m V_{stm}\right) \quad \textit{Equation 2}$$

where:

V_g, V_h	=	departure and arrival time-specific components
D_{h-g}	=	duration-specific components
m	=	tour modes (SOV, HOV, walk to transit, drive to transit, non-motorized)
V_{stm}	=	mode utility for the tour by mode *m*, leaving home in period *s* (containing hour *h*) and returning home in period *t* (containing *g*)
μ	=	mode choice logsum coefficient

The network simulations to obtain travel time and cost skims are implemented for five broad periods: Early A.M., A.M. Peak, Midday, P.M. Peak, and Evening. Mode choice logsums are calculated for each of the 15 outbound and return combinations by defining a representative half-hour within each period for each tour purpose. These 15 logsums are then associated with each of the 1176 alternatives in the model, depending on the relationship of the half-hour period to the network simulation period.

As noted above, each explanatory variable is associated with either the tour departure period, the tour arrival period, or the tour duration. At most any given variable can be associated with two out of the three dimensions. The association works by specifying a reference period for each dimension; this is typically the most frequently observed alternative for that dimension. For example, for work tours, the most frequently observed departure period may be the half hour starting at 7:30 A.M. The effect can be a pre-reference/earlier shift or a post-reference/later shift. The pre-reference shift variable counts the number of periods before the reference period; the post-reference variable counts the number of periods after the reference period.

For example, to model the effect of part-time (compared to full-time) work status on departure time, a binary (0,1) variable indicating part-time status is defined. It is then multiplied by the pre-reference variable for a pre-reference shift, and by the post-reference variable for a post-reference shift. Typically the coefficient for part-time workers would be negative for pre-reference shift and positive for post-reference shift, indicating a preference for part-time workers to depart home for work later than full-time workers. Note that splitting the shift effects into two parts (earlier/later) provides flexibility for asymmetrical shift effects around the reference period. Most shift effects are specified as linear, but non-linear effects can also be tested, by multiplying the explanatory variable by the square or cube of the shift variable.

Note that it is helpful to plot utility distributions for variables, particularly if the effects are not straightforward. These are plotted as changes in utility (on the y-axis) compared to the base case by period before or after the reference period (on the x-axis).

Following are some of the typical variables to be tested (in addition to constants):

- Household variables: income, size, number of workers, autos owned, presence of children
- Person variables: person type, age, gender, work status, school status, license-holding status, work-from-home status, external worker status, industry, telecommute frequency
- Tour frequency variables: number of work tours, number of school tours, number of non-mandatory tours to be scheduled (to the extent that these have already been modeled)
- Residual time window size: how many periods are left unscheduled after scheduling previously scheduled tours
- Mode choice logsum: Note that often the logsum coefficient must be asserted due to endogeneity (travel in peak periods is necessary due to external constraints and therefore indicates insignificant congestion effects on timing of travel). We have found via sensitivity testing that mode choice logsum coefficients need to be 0.5 or higher in order to ensure that time-of-day choice is reasonably sensitive to congestion

Note that estimation functionality for this model in ActivitySim is currently limited to the person-level variables and the interaction effects. The joint tour choice model component coefficients are not currently estimated and this will need to be addressed in the software.

Vehicle Allocation Model

Number of Models: Maximum of five alternatives (equal to the number of autos owned by the household plus one alternative for non-household vehicles)
Decision-Making Unit: Tour
Model Form: Logit
Alternatives: 13

The vehicle allocation model determines what vehicles would be used for the tour should an auto mode be selected. It is run prior to mode choice so that the auto operating costs for the most likely vehicles (and autonomous vehicle parameters if running a scenario in which autonomous vehicles exist) can be used in tour mode choice and downstream models.

The model alternatives are the vehicles owned by the household in the vehicle table, as predicted by the vehicle type choice model, plus one additional option for a non-household vehicle. The model is run once for each occupancy level in mode choice (drive-alone, shared 2, shared 3+).

Explanatory variables include the following:

- Total round trip tour distance, interacted with fuel type
- Auto parity, interacted with body type and vehicle age
- Auto occupancy, interacted with body type
- A set of alternative-specific constants, relating to vehicle body type, fuel type, and age

Estimation functionality is not currently implemented for this model.

Tour Mode Choice Model

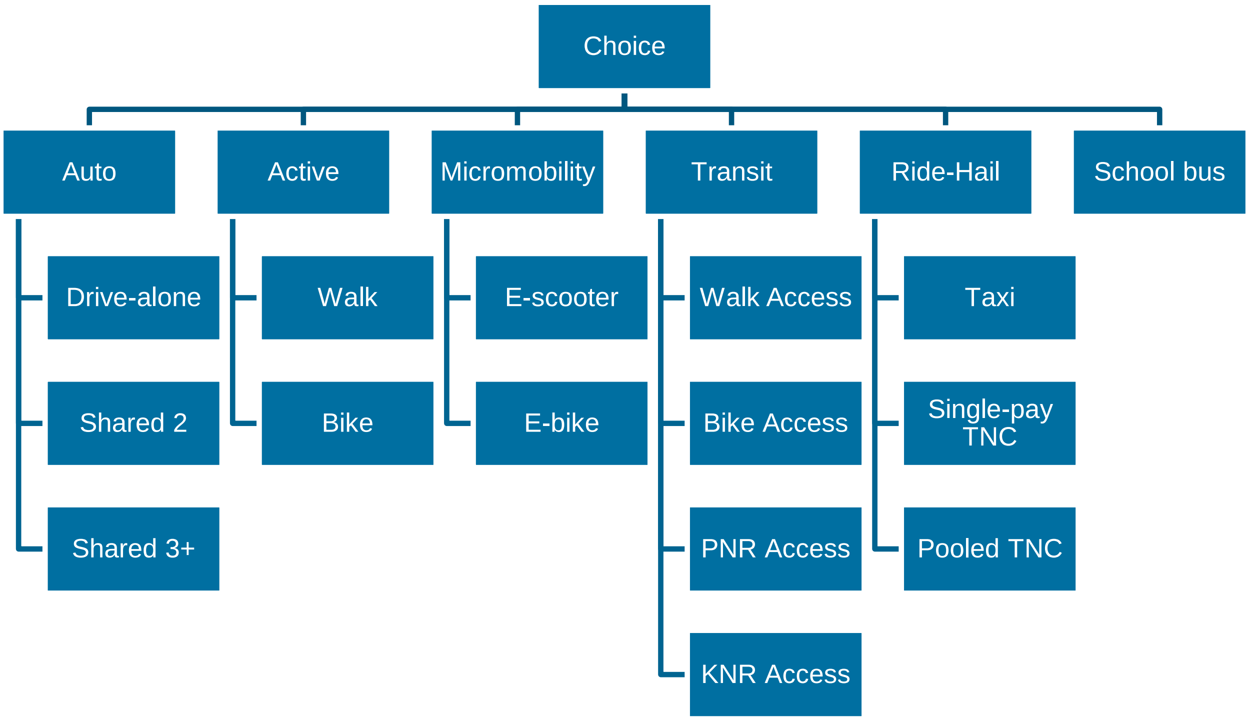
Number of Models:	Coefficients segmented by purpose (10 with at-work subtours)
Decision-Making Unit:	Tour
Model Form:	Nested Logit
Alternatives:	13

By means of this model, the “tour mode” used to get from the origin to the primary destination and back is determined. The tour-based modeling approach requires a certain reconsideration of the conventional mode choice structure. Instead of a single mode choice model pertinent to a four-step structure, there are two different levels where the mode choice decision is modeled:

- The tour mode level (upper-level choice),
- The trip mode level (lower-level choice conditional upon the upper-level choice).

The tour mode level can be thought of as a mode preference model, while the trip mode choice model can be thought of as a mode switching model. Tour mode choice is used to constrain stop location choice as well as trip mode choice. The modes for both models are the same, but the higher level of the nesting structure constrains lower-level decisions. Figure 14 shows the proposed alternatives and potential nesting structure for both tour and trip mode choice.

FIGURE 14: PROPOSED MODE CHOICE STRUCTURE



Tour modes are defined based on a set of rules pertaining to the combination of modes reported for each trip on the tour. The following set of rules are proposed, subject to modification and finalization based on observed combinations of trip modes on tours. Note that micromobility and ride-hail modes are constrained at that level, rather than the lower level trip modes.

1) If any mode is PNR-transit, the tour mode is PNR-transit
2) If any mode is bike-transit, the tour mode is bike-transit
3) If any mode is KNR-transit, the tour mode is KNR-transit
4) If any mode is walk-transit transit, the tour mode is walk-transit
5) If any mode is ride-hail (taxi, single pay TNC, shared TNC), the tour mode is ride-hail
6) If any mode is micromobility (e-scooter or e-bike), the tour mode is micromobility

7) If any mode is school bus, the tour mode is school bus.
8) If any mode is shared 3+, the tour mode is shared 3+
9) If any mode is shared 2+, the tour mode is shared 2+
10) If any mode is drive alone, the tour mode is drive alone
11) If any mode is bike, the tour mode is bike
12) If any mode is walk, the tour mode is walk

The tour mode choice model is based on the round-trip level-of-service (LOS) between the tour anchor location (home for home-based tours and work for at-work sub-tours) and the tour primary destination. The tour mode is chosen based on LOS variables for both directions according to the time periods for the tour departure from the anchor and the arrival back at the anchor. This is one of the fundamental advantages of the tour-based approach. For example, a commuter can have very attractive transit service in the a.m. peak period in the outbound direction, but if the return home time is in the midday or later at night, the commuter may prefer private auto due to lower off-peak transit service. The appropriate skim values for the tour mode choice are a function of the TAZ/MAZ of the tour origin and TAZ/MAZ of the tour primary destination. The mode choice model alternatives and skims are shown Table 18.

TABLE 18: MODE CHOICE ALTERNATIVES AND NETWORK LEVEL-OF-SERVICE VARIABLES

ALTERNATIVE	DESCRIPTION	NETWORK LEVEL-OF-SERVICE (SKIMS)*
Drive-alone	Single occupant auto	Auto time, distance by drive-alone (no HOV lanes) by three value of time bins
Shared 2	Auto with two occupants	Auto time, distance, and cost by shared-ride 2 (no HOV 3+ lanes) by three value of time bins
Shared 3+	Auto with three or more occupants	Auto time, distance, and cost by shared-ride 3+ by three value of time bins
Walk	Walk, skateboard, human powered scooter, wheelchair	Walk distance and time calculated across an all-streets network between MAZs within a certain distance threshold
Bike	Human powered bike	Bike distances and logsums calculated across an all-streets network between MAZs and TAZs within certain distance thresholds

e-scooter	Electric scooter	e-scooter distance and time calculated across an all-streets network between MAZs within a certain distance threshold
e-bike	Electric bike	e-bike distances and logsums calculated across an all-streets network between MAZs and TAZs within certain distance thresholds
Walk-transit	Transit by walk-access on both ends of the trip	Walk transit skims: in-vehicle by transit technology, total in-vehicle time, first wait, transfer wait, auxiliary walk time, transit fare, walk access time, walk egress time
Bike-transit	Transit by bike-access on both ends of the trip	Bike transit skims: in-vehicle by transit technology, total in-vehicle time, first wait, transfer wait, auxiliary walk time, transit fare, bike access time, bike egress time
PNR-transit	Transit by PNR-access on one end of the trip and walk on the other end	PNR transit skims: in-vehicle by transit technology, total in-vehicle time, first wait, transfer wait, auxiliary walk time, transit fare, PNR access (or egress) time, walk egress (or access) time
KNR-transit	Transit by KNR-access on one end of the trip and walk on the other end	KNR transit skims: in-vehicle by transit technology, total in-vehicle time, first wait, transfer wait, auxiliary walk time, transit fare, KNR access (or egress) time, walk egress (or access) time
Taxi	Traditional taxi mode	Auto time, distance, and cost by shared-ride 2 (no HOV 3+ lanes) by high value of time bin
TNC-single	Single-pay TNC mode, such as Uber or Lyft	Auto time, distance, and cost by shared-ride 2 (no HOV 3+ lanes) by high value of time bin
TNC-pool	Shared TNC mode, such as Uber Pool, Lyft Line or other micro-transit service	Auto time, distance, and cost by shared-ride 2 (no HOV 3+ lanes) by low value of time bin

School bus	Yellow bus	Auto time, distance, and cost by shared-ride 3+ by low value of time bin

* Notes on skims:

- Auto and transit skims are differentiated by five time of day periods and calculated across a planning network between TAZs.
- Auto skims are additionally differentiated by three value-of-time bins to represent travel time and cost heterogeneity. This is described more fully below.
- Auto costs include tolls on the auto network. Note that for specific scenarios, the user may wish to include discounts for certain user groups. These can be handled by applying modifiers to the tolls in mode choice utility equations, or in pre-processors as described in Section 4.6.
- Transit network level-of-service skims are differentiated by access and egress mode. Walk-transit skims are based on walk access and walk egress (walk-transit-walk). Bike-transit skims assume most travelers take their bike on-board with them, so they assume bike as both access and egress modes (bike-transit-bike). Park-and-ride and kiss-and-ride skims assume auto is used at the home end of the tour. Skims must be built by direction for each time period (PNR-transit-walk and walk-transit-PNR, KNR-transit-walk and walk-transit-KNR).
- Walk to transit times are calculated from an all-streets network based on the MAZ centroid and the nearest stop(s), as differentiated by broad transit stop types (local, premium).

Travel time and cost heterogeneity

The proposed design includes recommendations contained in research sponsored by the Strategic Highway Research Program (SHRP) C04 track on pricing and reliability[26]. The final report recommended a number of key features:

- Travel time heterogeneity: Sensitivity to travel time should be represented as a distribution reflecting personal preference and contextual conditions

[26] Parsons Brinkerhoff, Northwestern University, Mark Bradley, University of California at Irvine, RSG, University of Texas at Austin, Frank Koppelman, GeoStats, SHRP2 C04 - Improving Our Understanding of How Highway Congestion and Pricing Affect Travel Demand, Transportation Research Board of the National Academies (TRB), The Second Strategic Highway Research Program (SHRP2) Capacity Research, April 2013

- Continuous representation of income: Sensitivity to cost should ideally be represented as a continuous function of income rather than a global average
- Vehicle occupancy effects: Sensitivity to cost is also dependent upon the occupancy of the vehicle, but it is not a linear relationship.

Each person in the synthetic population will draw a mandatory and a non-mandatory travel time sensitivity parameter from log-normal distribution with a mean of 1.0. The mandatory travel time sensitivity parameter is used to factor skimmed times identified in Table 18 for work and school tours; the non-mandatory travel time sensitivity factor will be used to factor skimmed times for all other tour purposes. These factors are used to represent heterogeneity in travel time sensitivity across the population. So the travel time sensitivity coefficient is:

$$\beta time_{,purpose,person} = \beta time_{purpose} * \beta time\_sensitivity_{person}$$

The cost coefficient is calculated based on a cost coefficient that varies by tour/trip purpose, the household income of the traveler, the mode occupancy, and an income exponent:

$$\beta cost_{,purpose,person} = \beta cost_{purpose} / max(income,1000)^{income\_exponent} * occupancy\_factor$$

The value-of-time is calculated as:

$$\beta time_{,purpose,person} / \beta cost_{,purpose,person} * 60/100$$

Because travel time sensitivity varies across the population, the value-of-time (derived from time and cost sensitivity) also varies across the population. The variance in sensitivity to time and cost provides greater sensitivity to pricing, rather than relying on a single value of time for all travelers or a value-of-time that only varies according to trip or tour purpose. It also recognizes the value of time is a function of household income but is not solely determined by household income. Note that the time sensitivity distribution applies to all components of time, not just auto time. After calculating the value of time for each person tour or trip, the tour or trip is associated with the appropriate auto skims for that value of time category. These calculations are performed in the ActivitySim tour and trip mode choice preprocessors.

Transit modes and utilities

The current design specifies 4 access modes for transit – walk, bike, park-and-ride (PNR), and kiss-and-ride (KNR). A fifth mode – transportation network company or TNC – can be added for future forecasting if desired, but it is likely that current observed trips using TNC as an access mode to transit are very few which would make estimation challenging for this option. As mentioned above, skims must be built by direction for each time period; this is different from most trip-based models in which skims are built only in the AM peak period and assume auto can only be used on the access end of the transit trip. In activity-based models, skims are built in both directions for drive access – PNR-transit-walk and walk-transit-PNR; same with KNR.

We recommend utilizing the Oregon Metro approach for skimming transit. The approach reflects non-included attributes (comfort, reliability, information) of transit associated with the stop and the vehicle type, as follows:

- In-vehicle weights represented by segment-specific in-vehicle time parameters
- Stop wait times represented by node-specific wait time parameters
- Stop constants represented by node-specific variables, compiled additively along path and divided by boardings to calculate average constant (do not influence paths)
- Wait time calculation = headway/2 * stop factor

The approach tends to favor the use of premium stops and premium (fixed-guideway) transit modes in the path choice, which is useful since the recommended mode choice structure does not explicitly represent competition between bus and premium transit modes. The transit skims should include in-vehicle time by mode so that additional calibration constants can be added if necessary.

The skimmed walk access and egress times are replaced by MAZ-stop times calculated across an all-streets network and written to an MAZ file that has one row for each MAZ and the time to the nearest stop, differentiated by type of stop (bus versus premium transit). Alternatively, the time can be an average of all stops within a certain walk distance. Experimentation may be necessary with this approach to ensure that averaging stop times result in a better walk time estimate than relying on the closest stop.

Walk mode

Walk mode times will be based on MAZ distances within walking distance calculated across an all-streets network.

Bicycle mode treatment and bike comfort level

As described above, a bike comfort level model predicts the comfort level of biking for each person in the synthetic population. Each bike comfort level is associated with a bike route choice logsum and bike distance as determined by path choice across the all-streets network for MAZs and TAZs within bike distance. Distance on each type of bike facility can also be traced and used in the model estimation process.

Micromobility modes

Micromobility modes include e-scooters and e-bikes. An interesting aspect of micromobility is that these modes can be privately owned or shared. If privately owned, we assume the traveler can use the mode for any tour starting at home for free, with no access time. If shared, we include the cost of the mode in the utility for using it, as well as an access time that varies according to the origin zone. These access times must be specified in the input zonal file.

E-scooter time will be based on a distance calculated as shortest path across the all-streets network. E-bike utility can be specified simply (using shortest path calculation) or a route choice logsum (possibly as a later enhancement).

Different alternative-specific constants can be calibrated for owned versus shared micromobility modes.

Ridehail modes

Ridehail modes include taxi, single-pay TNC (such as Uber or Lyft) and pooled TNC (Uber Pool, Lyft Line) Pooled TNC is also referred on ‘on-demand transit’ or ‘micro-transit’, and can include minivans and minibuses in addition to passenger cars. We recommend keeping these modes separate because the wait times and availability can vary considerably across the region, and due to fare differences between the modes. Auto skims are used for these modes, with an additional wait time and cost function. ActivitySim provides the ability to relate the wait time according to a distribution based on area type of the origin zone. Alternatively, one can vary the wait time based on the origin zone directly, as a zonal file input. The utility of pooled TNC includes an extra factor on both wait time and in-vehicle time, due to the need to divert vehicles to pick up the requested ride. The cost of each mode is specified in the constants.yaml file and includes an initial fee (meter-drop), a cost per mile, a cost per minute, and a minimum cost if relevant.

Pooled TNC is a unique mode whose availability has changed in recent years as Uber and Lyft have pulled these services back from most metropolitan areas due to the COVID pandemic. However, there continues to be interest in this mode, particularly in scenarios in which autonomous vehicles are expected to be available. It is possible that shared autonomous vehicles could reduce private auto ownership if the price reduction (due to elimination of labor costs) would make the mode more attractive. Microtransit is also of interest to many regions, as a substitute for more expensive bus service in corridors with low ridership.

Alternative-specific constants

The calculation of calibration targets for trip mode choice are based on the tour mode coding. Currently in ActivitySim, we limit PNR access tours to only one trip in each direction; this is done to ensure that the same parking location is accessible at both ends of the tour. Therefore after tour mode is coded, any non-PNR transit trips on PNR tours must be moved to the tour mode that is consistent with their mode. For example, if there are drive-alone trips on PNR tours, they would be moved to the drive-alone tour mode prior to model calibration. Similar steps are taken for cells with very low counts of trips within a tour mode. For example, we may find a very small number of shared 2 trips on school tours. These would be moved to shared 2 tour mode.

Intermediate Stop Frequency Model

Number of Models:	10 (By purpose plus one model for at-work subtours)
Decision-Making Unit:	Person
Model Form:	Multinomial Logit
Alternatives:	Maximum 6 total, 3 per tour

The stop frequency choice model determines the number of intermediate stops on the way to/from the primary destination up to a maximum of 3 per direction, for a total of 8 trips per tour (four on each tour leg). However, for many tour purposes, the number of intermediate stops observed in the data is significantly less than 3 per direction. An additional constraint placed on intermediate stop models is that no stops are allowed on drive-transit tours. This is enforced to ensure that drivers who drive to transit pick up their cars at the end of the tour.

Stop frequency is based on a number of explanatory variables, including household and person attributes, the duration of the tour (with longer durations indicating the potential for more stop-making) the distance from the tour anchor to the primary destination (with intermediate stop-making positively correlated to tour distance), and accessibility and urban form variables.

Once the number of intermediate stops is determined, each intermediate stop is assigned a purpose based on a frequency distribution derived from the household travel survey. The distribution is segmented by tour purpose, tour direction (outbound versus return) and person type and is based on survey data summaries. Work tours are also segmented by departure or arrival time period.

Intermediate Stop Location Choice Model

Number of Models:	10 (By purpose plus one model for at-work subtours)
Decision-Making Unit:	Person
Model Form:	Multinomial Logit
Alternatives:	Microzones

The stop location choice model predicts the location of stops along the tour other than the primary destination. The stop-location model is structured as a multinomial logit model using a zone attraction size variable and route deviation measure as impedance. The alternatives are sampled from the full set of zones, subject to availability of a zonal attraction size term. The sampling mechanism is also based on accessibility between tour origin and primary destination and is subject to certain rules based on tour mode. All destinations are available for auto tour modes, so long as there is a positive size term for the zone. Intermediate stops on walk tours must be within 3 miles of both the tour origin and primary destination zones. Intermediate stops on bike tours must be within 8 miles of both the tour origin and primary destination zones.

The intermediate stop location choice model works by cycling through stops on tours. The level-of-service variables (including mode choice logsums) are calculated as the additional utility between the last location and the next known location on the tour. For example, the LOS variable for the first stop on the outbound direction of the tour is based on additional impedance between the tour origin and the tour primary destination. The LOS variable for the next outbound stop is based on the additional impedance between the previous stop and the tour primary destination. Stops on return tour legs work similarly, except that the location of the first stop is a function of the additional impedance between the tour primary destination and the tour origin. The next stop location is based on the additional impedance between the first stop on the return leg and the tour origin, and so on.

Trip Departure Time Model

Number of Models:	1
Decision-Making Unit:	Person
Model Form:	Lookup from probabilities
Alternatives:	35

The trip (stop) departure time model simulates the half-hour departure time for each outbound and inbound trip on a tour based on a lookup of probabilities by tour purpose, inbound versus outbound indicator, time remaining on the tour (number of time periods between the departure of previous trip and arrival of the tour), and number of stops remaining on the half-tour. These probabilities are created from survey data from CMAP, MWCOG (Washington, DC), SEMCOG (Detroit, MI), and SANDAG (San Diego, CA) HTS data. The (up to) 35 alternatives are in reference to the previous trip, from being in the same half-hour period as the previous trip (0-period offset) up to 34 half-hour periods later than the previous trip. This method (scheduling in reference to the previous trip instead of absolute time periods) prevents scheduling failures when a later trip can't be scheduled due to a scheduling conflict. The probability table is too large to fit into this document but can be found online at https://github.com/CMAP-REPOS/cmap_abm/blob/activitysim/activitysim/configs/trip_scheduling_probs.csv.

Trip Mode Choice Model

Number of Models:	10 (By purpose plus one model for at-work subtours)
Decision-Making Unit:	Person
Model Form:	Nested logit with constraints by tour mode
Alternatives:	13

The trip mode choice model determines the mode for each trip along the tour. Trip modes are constrained by the main tour mode. The linkage between tour and trip levels is implemented through correspondence rules (which trip modes are allowed for which tour modes). The model can incorporate asymmetric mode combinations, but there is a great deal of symmetry between outbound and inbound modes used for the same tour. In particular, symmetry is enforced for drive-transit tours, by excluding intermediate stops from drive-transit tours.

The tour and trip mode correspondence rules must be created after coding tour modes in the survey data and analyzing the resulting tabulation of trips by trip mode and tour mode. Cells for which there are an insufficient number of observed trips (due to low shares) are handled by moving those trip targets to other tour modes.

Note that in the trip mode choice model, the trip modes are exactly the same as the modes in the tour mode choice model. However, every trip mode is not necessarily available for every tour mode. Also, while tour mode choice requires round-trip levels of service, trip mode choice is based on only the level of service between the trip origin and destination, corresponding to the trip departure time period.

The trip mode choice model explanatory variables include household and person variables, level-of-service between the trip origin and destination according to the time period for the tour leg, urban form variables, and alternative-specific constants segmented by tour mode.

Parking Location Choice Model

Number of Models:	1
Decision-Making Unit:	Tour
Model Form:	Multinomial Logit (destination choice)
Alternatives:	MAZs

The parking location choice model determines the parking location for each tour. The model is intended to represent traveler behavior to locations where parking is not necessarily available in the same MAZ as the activity location, such as downtown zones. Tours with activities in such areas are allowed to select a different zone than the activity location for parking; this zone is then used as the destination for vehicle trips to the parking area (and origin for vehicle trips leaving the parking area). The alternatives are MAZs, typically constrained by a ‘parking constrained area’ indicator that is drawn around MAZs with parking costs and designated parking lots/spaces. Explanatory variables include the distance of the MAZ to the activity location, the actual cost of parking in the MAZ, and the number of spaces in the MAZ. The model is currently unconstrained, but it can be applied in a constrained manner via time-based simulation or via application of shadow prices in an iterative algorithm (See section 3.15).

3.10 MODELING OVERNIGHT VISITOR POPULATION

We recommend a separate model for representing travel for overnight visitors, rather than modeling visitors within the same model system as residents, for the following reasons:

- Overnight visitors typically have very different travel characteristics, even when explanatory variables such as person type are the same as residents. For example, school age children participate in non-mandatory tours throughout the day.
- Overnight visitors are attracted to very different land-uses than residents, even for similar types of tours. For example, discretionary tours made by residents are often made to urban parks, whereas visitors are typically attracted to tourist destinations
- It is important to distinguish visitors who are in Oregon for business purposes from visitors who are here for personal and recreational travel, as they have very different travel patterns which are not represented in the resident model.
- Data describing overnight visitor travel is typically limited, so a simpler model system (one that does not include intra-household constraints, for example) is warranted.

We recommend applying the overnight visitor model developed for San Diego Association of Governments for regions that wish to model visitors explicitly. This model was recently implemented in ActivitySim. The following section describes the overnight visitor model in more detail.

Visitor Model Description

This section describes the model system briefly. The design is shown in Figure 15. A key aspect of the model system is that it does not track 24-hour time windows of visitors; once generated, tours are scheduled via probability distributions. It also avoids issues related to intra-household coordination and ridesharing by instead modeling travel parties, and assuming these are fixed for each tour.

Probability distributions are used to generate tours and stops, rather than logit models. This makes the model much easier to develop and results in faster runtimes than if the resident model structure was applied to visitors, and reduces the data burden for development of the models.

Visitors are generated for two visitor segment types:

- **Business:** Self-identified as business traveler, or self-identified as 'Both Business and Personal' but took at least one 'business' purpose trip on travel day
- **Personal:** Self-identified as personal traveler, or self-identified as 'Both Business and Personal' but took no business purpose trips on travel day. A few self-identified Personal travelers have reported Work tours.

There are three visitor tour purposes, which were coded based on the reported trip purpose in the survey, as follows:

- **Work:** Business travel made by Business travelers
- **Recreational:** All other recreational purposes besides dining
- **Dining:** Travel to eating establishments

Each travel party can generate one or more tours of each purpose on any given day. Tour generation rates (probabilities for each combination of number of tours by purpose) are specified as a fixed input by segment.

The visitor model component sequence is described below.

1. **Visitor Tour Enumeration:** Visitor travel parties are created by visitor segment (business versus personal/recreation) based upon input accommodations employment and households. Travel parties are attributed with household income. Tours by purpose are generated for each party. Each tour is attributed with auto availability and party size. The tour origin MAZ is set to the MAZ where the tour was generated.
2. **Tour Level Models**
   - 2.1. Tour Time of Day: Each tour is assigned a time of day, based on probability distribution.
   - 2.2. Tour Destination choice: Each tour is assigned a primary destination, based on the coefficients estimated through a multinomial logit model, similar to resident tour destination choice models, but without the external identifier/destination component. Also destination choice models for visitors are typically heavily influenced by special attractions such as convention centers, tourist attractions, and large shopping centers.
   - 2.3. Tour Mode Choice: Each tour selects a preferred primary tour mode, based on an asserted nested logit model (the resident tour mode choice model). Constants are stratified by tour purpose and auto availability. Auto modes are turned off if an auto is not available.
3. **Stop Models**
   - 3.1. Stop Frequency Choice: Each tour is attributed with a number of stops in the outbound direction and in the inbound direction, based upon sampling from a distribution.

3.2. Stop Purpose: Each stop is attributed with a purpose, based upon sampling from a distribution.

3.3. Stop Location Choice: Each stop is assigned a location based upon a multinomial logit model (similar to resident stop location choice models)

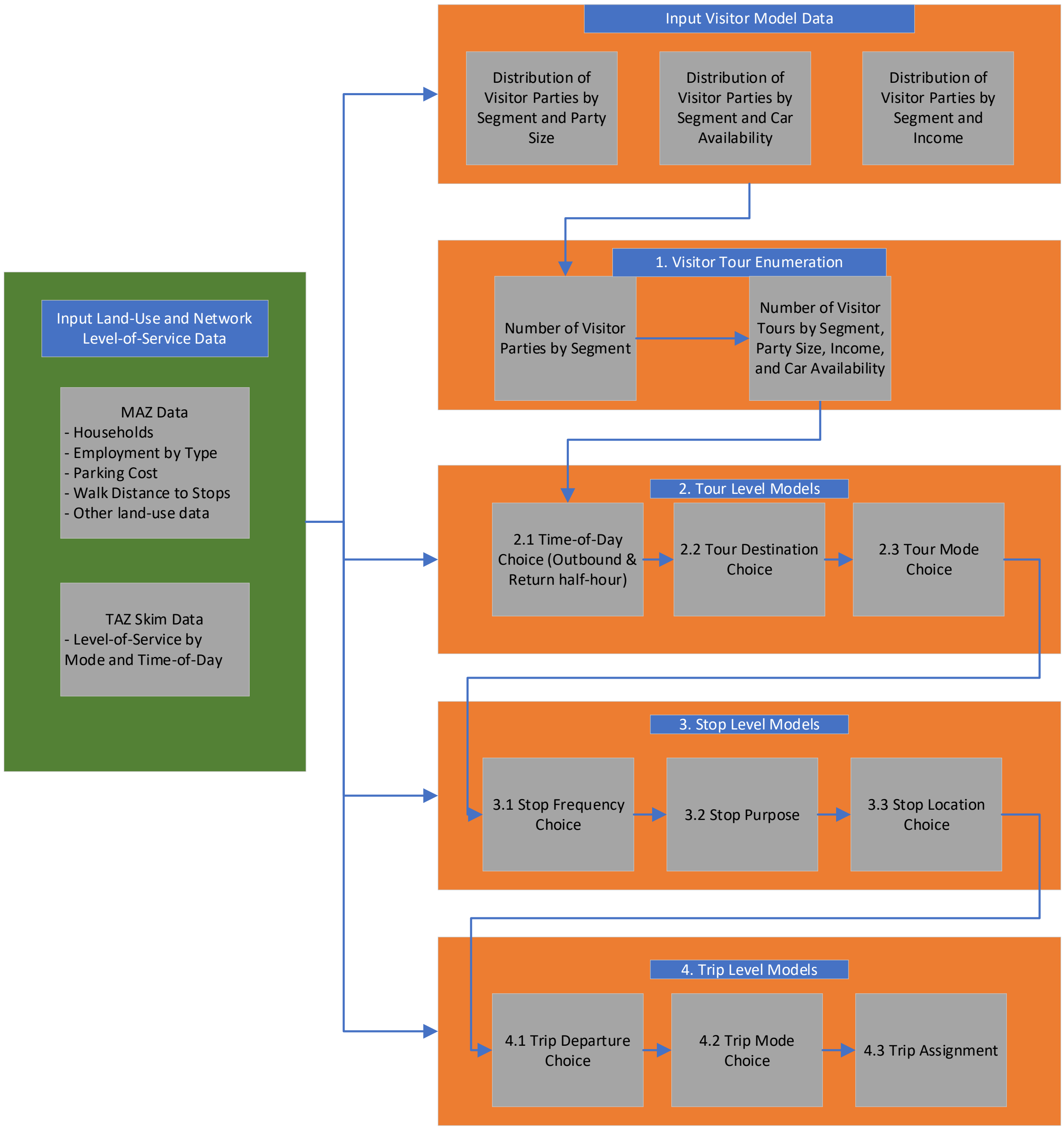
4. **Trip Level Models**

4.1. Trip Departure Choice: Each trip is assigned a departure time period based upon sampling from distributions.

4.2. Trip Mode Choice: Each trip within the tours selects a preferred trip mode, based on an asserted nested logit model.

4.3. Trip Assignment: The visitor trip list is aggregated into trip tables and assigned to the network along with resident trips and trips from auxiliary models (trucks, commercial vehicles, through trips, external-internal trips, etc.).

FIGURE 15: SANDAG VISITOR MODEL DESIGN



3.11 EXTERNAL TRAVEL

The above design explicitly generates travel made by residents to external zones for work and non-work purposes. The Statewide Integrated Model (SWIM) will be used to generate inputs to ActivitySim that influence the generation and distribution of resident travel to external zones. Trip tables representing non-resident travel to internal zones and through travel will also be derived from SWIM. The following list describes the linkages and inputs required from SWIM. All of these inputs will be created using the SWIM select link estimator tool[27].

- SWIM estimates of travel made by resident workers to each external station for work commuting are used to as inputs to the external worker identification model and the external workplace location choice model – to identify both workers whose regular workplace is external to the region and the distribution of their work tours to external stations.
- SWIM estimates of travel made by residents to each external station for non-work purposes are inputs to the external non-mandatory tour identification model and the external non-mandatory tour location choice model – to identify non-mandatory tours with a destination external to the region and the distribution to external stations.
- Trip tables will be developed for non-resident travel to internal zones. We recommend stratifying them these trip tables by work and non-work purpose. The select link tool will be used to create the initial trip tables; internal trip ends must then be allocated to the MPO MAZ system (see below) and aggregated to TAZs for assignment.
- Work trip ends from the external-to-internal trip tables created in the above step will be used to estimate the percentage of jobs in each MAZ that are ‘consumed’ by non-resident workers. These reductions will be applied to MAZ employment estimates to reduce the size terms used in work location choice in order to reflect in-commuting on the distribution of resident work location choice.
- Through trip tables developed from SWIM will be added to resident travel, overnight visitor travel, non-resident-internal travel, trucks and commercial vehicles prior to assignment.

3.12 FREIGHT\COMMERCIAL VEHICLE TRAVEL

While the ActivitySim kernel will not by itself directly treat truck and light commercial vehicle (LCV) travel, MPOs’ complete model systems will generally need to estimate truck and LCV trip making and assign those trips to the network. Oregon’s MPO models now include a variety of

[27] https://github.com/tlumip/tlumip/wiki/SL

truck/LCV options ranging from the trip-based JEMnR model's aggregate truck submodel[28] to Metro's tour-based truck microsimulation submodel.[29] Both types estimate single-unit (medium) trucks, multi-unit (heavy) trucks, and LCVs (which include "light trucks" such as commercial-use pickups plus passenger autos or vans used for commercial purposes). Both types are generally run as part of their model system's global iteration loops, feeding their vehicle trip matrices into the general assignment module and thus incorporating trucks and LCVs as part of the traffic that contributes to congestion in the system. The system's global feedback loop sends impedance skims back both to the passenger demand components and the truck submodel, ensuring that both passenger and commercial travel are sensitive to roadway performance. While both truck submodel options depend heavily on employment by TAZ by category as a key input, they differ in how they estimate truck trip creation, truck trip distribution (or destination choice), and in their detailed treatment of the three individual truck classes. The only explicit integration to date of Oregon MPOs' truck submodels with the passenger elements of their overarching model systems has been the unified assignment and skimming of trucks and passenger vehicles together, plus the use of employment data inputs for both passenger and truck model components.

The ActivitySim design assumes that ActivitySim instances in Oregon will be in model systems that will include one of the truck submodel options (presumably the submodel already in use by a given MPO) and that, at minimum, the model systems will continue to apply their truck submodels as described above. However, the ActivitySim platform offers the opportunity for additional integration of the truck submodel with the ActivitySim passenger demand components, specifically household and individual day pattern formulation. The most obvious option is for the day pattern model to estimate at-home activities (e.g. e-commerce purchases) that generate and affect the destination of commercial vehicle trips (package delivery). The amount and location of these e-commerce activities could be fed into a (modified) truck submodel to influence the creation and distribution of a portion of the total estimated LCV travel (only a portion since LCV travel includes service trips such as plumbers). The necessary truck submodel modifications would vary depending on the type at hand: the JEMnR version could alter its use of input truck/LCV production and attraction rates and the format of those inputs to make use of the ActivitySim-generated e-commerce activities. The corollary tour-based truck submodel modification could treat the spatial locations of LCV destination choice based on the ActivitySim-generated e-commerce household locations and, perhaps, modify the generation of synthetic firms that handle package delivery in ways that appropriately increase or decrease the overall amount of package delivery activity. The Oregon Travel Survey (OTS) now in the field would enable a basic ActivitySim day pattern formulation given its question "We want to know how shopping impacts travel. On <traveldate> , did <you/name> buy anything online?"[30]

[28] JEMnR · tlumip/CALM Wiki · GitHub
[29] https://rsginc.github.io/portland_freight_model/Metro_Freight_Model_User_Guide.html
[30] RSG. *Oregon Household Travel Survey: Survey Questionnaire.* 2023.

Portland State University researchers are also collaborating with the OHTS team to devise and pilot a more-comprehensive adjunct e-commerce survey that seeks to illuminate additional details. Those details are likely to include the frequency, volume, and share of household and individual purchasing of e-commerce.[31] Should the e-commerce pilot prove successful and that survey be fully deployed statewide, it will increase the potential explanatory variables and outputs of any e-commerce-related day pattern treatment.

One can conceive of additional possible integration features, for example estimating the non-typical workplace location of workers and using that data to condition the employment inputs to the truck submodels, but such steps would require additional research so the e-commerce integration may be the simplest and highest-value option.

3.13 UNIVERSITY TRAVEL

University students are modeled as a specific person type in ActivitySim. University students living off-campus in family and non-family households are included in the non-group quarters synthetic population, while students living in group quarters, which includes dormitories as well as fraternity and sorority houses are included in the group quarters population. Therefore general travel patterns to and from colleges and universities is represented explicitly in the proposed model design. However, the current ActivitySim software has some limitations with respect to university travel – especially larger universities that have a large student population that lives in close proximity to campus and where the campus spans multiple zones. These limitations restrict the ability of the model to accurately represent walking, biking, and transit trips between parking and campus or within campus. The following section lists key limitations of the current ActivitySim design (and software) and the potential enhancements related to university student travel.

1. The accuracy of residential location of university students is mostly dependent upon the age controls used in to create the non-group quarters synthetic population and estimates of group quarters households for university students. The proposed population synthesis procedure includes 18-24 year old age controls at the block group level for non-group quarters households, and university student households for group quarters households. The synthetic population for areas around University of Oregon (UO) and Oregon State University (OSU) should be analyzed to determine whether these constraints are sufficient for accurately representing the student population. If not, we recommend development of a a university student residential location choice model which would be run prior to running population synthesis. Outputs from the model can then be used as an additional control to ensure that the right students are located in close proximity to major universities.

[31] Portland State University. *E-Commerce Pilot Survey*. (internal working draft, 11/6/23)

2. ActivitySim currently only selects one school zone for each student and sends all school tours to that zone. This restricts the ability to generate intra-campus trips for students that attend a university that spans multiple zones, such as UO, OSU, and Portland State University. This has implications for the ability of the model to generate the right number of walk, bike, and transit/campus shuttle trips. The ActivitySim model developed for Southeast Michigan Council of Governments (SEMCOG) was enhanced to address this problem. The model re-samples a new school zone for each school activity based on classrooms and other space data by zone. This allows the model to generate travel across campus and model university transit ridership.

3. ActivitySim treats parking location as a special zone and travel between the lot and the activity location isn't explicitly modeled; walk is assumed as the mode. Campus parking is quite unique; the choice of parking is often limited based on type of pass owned, parking lot space constraints are more significant, and parking can be quite far from campus requiring a campus shuttle. The SEMCOG model has an enhancement to address this issue, It includes a parking allocation model for all auto trips onto campus where an intermediate stop is inserted into the tour before and after the first and last on-campus activities respectively. The location of each stop is set equal to the parking zone. All intra-campus trips are modeled as if the tour mode is ‘walk-transit'; their choice of mode is limited to walk and walk-transit trip modes. That allows for a more accurate representation of mode choice for intra-campus travel.

The above enhancements are all potentially valuable additions for more accurate modeling of university travel. However, they are not part of the core ActivitySim code base and would therefore require additional software development. The Oregon Travel Survey is specifically targeting data collection for university students; this data will be invaluable to develop core ActivitySim estimation and calibration datasets and enhance the ActivitySim representation of student travel.

3.14 LAND-USE MODEL CONNECTIONS

There is currently no standard land-use modeling software or approach used in Oregon; each region has their own methods for generating land-use data for input to the travel demand model. However, ActivitySim offers some potential methods for integrating model outputs to land-use models such that accessibilities impact the distribution of households and jobs. The following lists the potential two-way linkages between travel and land-use models:

1. Future-year synthetic population controls can be derived from or provided by land-use models. These controls can be limited to just total households, or can include the socio-economic distributions used in PopulationSim. It is possible that geographic crosswalks would need to be developed and maintained to ensure that controls are being provided at appropriate geographic levels for use in the travel model.

2. Future-year employment data can be derived from or provided by land-use models.

3. ActivitySim can provide both aggregate and disaggregate origin-based accessibilities by purpose and market segment (see above). These accessibilities can be used by land-use models to influence the distribution of population and employment.

4. The travel time and cost skims by mode, period, and value-of-time bin can be used in land-use models to influence the distribution of population and employment.

There may be other linkages between the model systems as well; these depend upon the exact type of land-use model developed and the variables used by the model. We recommend further defining these interactions for each region depending upon what tools are used by the region.

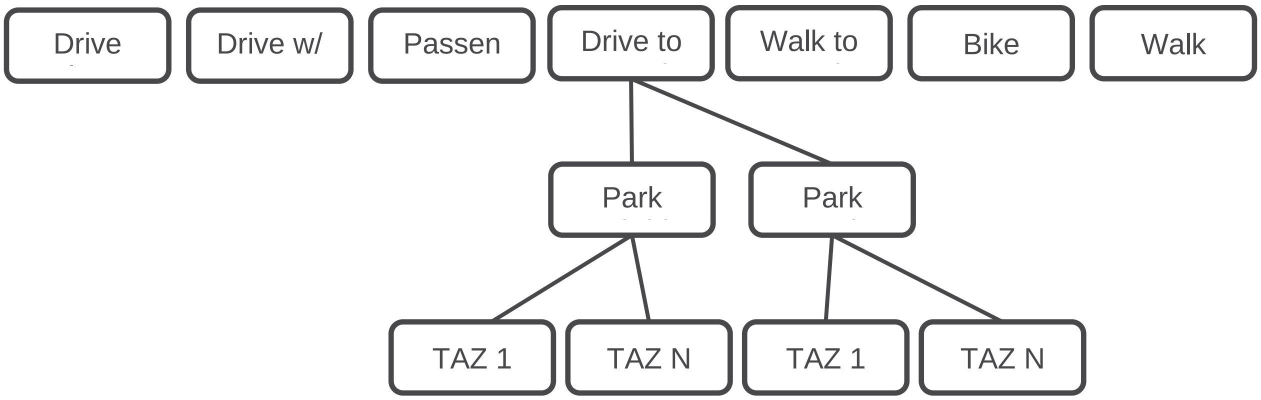
3.15 PARKING CONSTRAINTS

The current Oregon Metro model has an explicit representation of parking lot choice with capacity constraints. However, the current version of ActivitySim does not explicitly consider parking constraints in either transit park-and-ride choice or downtown parking. Parking supply (number of spaces) influences parking cost and location but does not strictly constrain the location of parking, the number of auto trips to destinations or the number of park-and-ride trips at specific lots. Below we discuss potential options for enhancing ActivitySim to address parking lot choice and capacity. They are organized into two main categories; those involving the calculation of shadow prices, and those using simulation-based constraints.

Shadow Pricing Approaches

One potential way to address lot choice is via an extension of the mode choice model, similar to the current Oregon Metro implementation. The lot choice model allocates trips to lots between ODs by trip purpose based on the composite utility of the auto + lot + transit path. It do so by a priori trimming the lot alternative set to an N defined number of likely lots, and then letting the mode choice model distribute the demand among the candidate lots. The full auto + transit path via each likely lot is considered in the mode choice model. Figure 16 shows a graphic of the extended mode choice model. Before running the lot choice model within mode choice, a set of N likely lots for formal (PNR) and informal (PNH) is created for each OD pair. This pre-processing creates a much smaller lot alternative choice set in order keep mode choice run times and memory requirements down. Then in mode choice, the utility is re-calculated including the auto utility to the lot, the transit utility from the lot to the destination, and the cost of parking at the lot including a shadow price which reflects capacity constraints. The model must be iterated to convergence, where the shadow price is adjusted at each iteration until parking demand is within a pre-determined range of parking supply at each lot.

FIGURE 16: OREGON METRO PARK AND RIDE LOT CHOICE MODEL



Implementation of this methodology in ActivitySim would require an extension of the mode choice model and development of supporting scripting and procedures in commercial transport software.

An entirely network-based approach is also possible. In such an approach, all park-and-ride trips are funneled onto either an auto link with a fixed capacity equal to the lot capacity, or bus route with a fixed capacity. In both cases, the demand is assigned to the network, a comparison is made to the capacity of the auto link or bus route, a cost is calculated and skimmed, and the assignment and/or demand model is re-run until some measure of convergence is reached. In the case of an auto link, an auto volume-delay function calculates delay; in the case of the bus route, the transit capacity restraint mechanism in the commercial transport software is used to calculate the capacity constraint cost. The exact number of assignment iterations and assignment-demand iterations must be evaluated and defined.

Simulation-based Approaches

Simulation can also be used to represent parking constraints. In a simulation-based approach, the demand model is run, over-capacity alternatives are eliminated from the choice set, and the demand model is re-run for only decision-makers who selected over capacity alternatives. This approach is currently used in ActivitySim to constrain work location choice models to match input jobs by zone. The advantage of simulation-based constraint mechanisms is that they tend to match hard constraints much closer than shadow pricing methods in less computational time than shadow pricing, because shadow pricing methods require re-running the entire population and because shadow prices are not hard constraints. Shadow prices are also particularly challenging in the context of parking location choice, because of the temporal nature of parking demand, with travelers arriving and departing throughout the day.

A simulation-based park-and-ride approach was recently developed in the activity-based model for Metropolitan Transportation Commission (MTC). After the demand model is run, a post-processing application reads in the output trip list and resorts the demand by departure time, disaggregated into 5-minute time intervals. Arrivals at parking lots for premium transit stations (BART, light-rail, commuter rail, ferries) are explicitly tracked and compared to parking supply in each time increment. Any lot that fills up is flagged for removal from the next demand iteration and any tour that arrives in that lot after it fills up is flagged for re-simulation. In the next demand iteration, only tours that selected lots that were full in the previous iteration are re-simulated, after removing those lots from their choices. The re-simulation in subsequent iterations is limited to tour mode choice and downstream models – the tour destination and all upstream models are held fixed. This allows the tours to change their tour mode based on the change in utility resulting from removal of the originally chosen parking lot, and also adjust their stop frequency, location, and trip mode accordingly.

Application of the approach in ActivitySim would require either implementing something akin to the lot choice model for Oregon Metro described above, or via iterating the assignment model with the parking post-processor and re-simulation approach. Because ActivitySim uses a modular design where individual components can be turned on or off, re-simulating only certain decisions in ActivitySim should be possible without any core software changes.

Typically, shadow prices and simulation-based constraints are not considered in model estimation; therefore the specific approach to be used for representation of parking constraints is not strictly necessary to proceed with joint model estimation. However, the approach can certainly influence model calibration since it impacts park-and-ride demand. Therefore we recommend that the specific approach be further defined prior to model calibration for those regions that intend to adopt it.

4.0 ANTICIPATED INPUT AND OUTPUT SPECIFICATION

4.1 NETWORK ATTRIBUTES

Below we discuss network attributes useful for development of transport skims input to the Oregon Activity-Based Model system specified in Section 3.

Highway Network Attributes

Highway network attributes should be developed and maintained to support the five time periods identified for skimming and assignment discussed in Section 3. The network should also be coded to allow for calculating paths by auto occupancy, so HOV lanes must be identified. Auto paths may also include toll facilities. Below we discuss relevant link level network attributes in more detail. Other attributes: A-node, b-node, distance, maximum speed, etc. are assumed to be available.

- Number of lanes: Number of lanes could vary over the day. For example, on-street parking may be prohibited during peak periods, where the parking lane is used for auto travel, and available for other time periods. Lanes may have occupancy restrictions for part of the day and be unrestricted for other periods. Coding number of lanes by period in each direction (e.g. AB_LANES_EA, AB_LANES_AM, etc.) provides maximum flexibility to calculate capacity for each period, should agencies wish to have this functionality. Note that Oregon Metro considers center turn lanes as half a lane, but this representation is not possible in VISUM since number of lanes is required to be an integer value. A workaround would need to be implemented in VISUM as an adjustment to number of through lanes.
- Auto occupancy restrictions: These attributes are in place to identify HOV lanes, which are typically coded as separate network links. Since HOV restrictions could vary by time of day, it is useful to code these similar to number of lanes. HOV restrictions should be coded for HOV 2 versus HOV 3+ facilities to provide maximum flexibility for forecasting.
- Truck restrictions: These attributes are used to restrict heavy trucks from using facilities which are weight restricted. There can be more than one weight restriction category depending on the trip tables produced by the heavy truck model. Steep grades can also be identified via a link field to add additional generalized cost for heavy trucks.

- Toll costs: Toll costs are typically coded in cents for each tolled link. Some commercial transport software allows coding tolls by entry-exit gate. In any event, tolls should be coded in such a way to allow the analyst to vary them by time of day and occupancy (AB_TOLL_SOV_AM, AB_TOLL_SOV_EA, AB_TOLL_SR2_EA, AB_TOLL_HTRUCK, etc.).
- Reliability: If reliability is included in the network, the static aspect of reliability (the portion that is not dependent upon volume-capacity ratio) would be stored as one or more link attributes. For SANDAG, that includes distance to and from major freeway interchanges, which is a measure of weaving that leads to accidents and unreliable conditions.
- Control type, cycle length, green/cycle ratio, and intersection capacity-related fields. These node fields are useful if inclusion of node-based delay are to be included in calculation of travel paths and times, and also in the case that control type is used in the bicycle path calculations. An example of such coding in VISUM can found on the Southern Oregon activity-based model github wiki[32]. The attributes are used to calculate capacity at each intersection and are used in the volume-delay function, originally developed by University of Arizona and Pima Association of Governments[33].

Transit Network Attributes

Each commercial transport software package has specific fields which must be coded for transit skimming and assignment. Below we address key considerations for skimming transit level-of-service for the Oregon activity-based model design.

- Route coding: Transit routes and headways should be coded for each of the five time periods. Routes that do not operate in a given time period should be removed from the route file for that period or otherwise made unavailable. Often transit networks for trip-based models are coded in the AM peak direction for peak skims and the midday direction for off-peak skims. Because transit will be skimmed and assigned for five periods, routes must be coded by direction in each period. Transit-only links are used to represent grade-separated facilities and road segments where passenger vehicles are prohibited (Tillicum Crossing, etc.). Rail times can be coded as station-to-station times or by a fixed time per link.
- Transit technology: It is useful to define a set of transit technologies to be used to differentiate time spent in different vehicle times, which can be skimmed separately and used in mode choice and post-processing steps. Oregon Metro’s transit path-building

[32] https://github.com/RSGInc/SOABM/wiki/auto-network-coding
[33] https://github.com/RSGInc/SOABM/wiki/soabm-volume-delay-function-definition

procedures uses separate path weights by transit technology. A core set of available technologies in Oregon would include local bus, express bus, bus rapid transit, light rail, and commuter rail. The list could be expanded in the future if other transit modes are considered (i.e. urban rail, ferry, etc.).

- Vehicle capacity: Transit vehicle capacities are required inputs for manual or automated demand-supply equilibration (transit capacity restraint). Transit vehicle capacities are also useful for post-model demand-supply summaries and performance metrics. Vehicle capacities are expressed as seated capacity and standing capacity. The potential number of standing passengers (‘standees’) is an estimate which should take into account North American comfort levels with respect to personal space requirements.
- Park-and-ride lot capacity: Number of spaces per transit park-and-ride lot are required inputs for manual or automated park-and-ride demand-supply equilibration. They are also useful for post-model demand-supply summaries and performance metrics. Number of spaces includes both total physical spaces as well as any informal on-street parking if available at a formal lot.
- Transit reliability: If transit reliability is modeled, certain network attributes may be useful to describe reliability. Area type was used in the Los Angeles Metro models as a proxy for measuring reliability affected by passenger loadings, for example. Other models may measure reliability as a function of actual boardings and alightings, which would not require additional link or node attributes beyond those provided by transit assignment models.

Non-Motorized Network Attributes

The non-motorized network is used for skimming walk and bike paths and assigning walk and/or bike trips output from ActivitySim. Since the network will be used to calculate bicycle route utilities, it must be attributed with bicycle facility types and other variables used in bike utility calculations. Currently we do not anticipate taking into account sidewalks in walk time calculations, but these can be added at a later date if desired. One would need to make sure that walking paths are represented reasonably well in the all-streets network as well.

Because non-motorized path calculations are computed between MAZs, the non-motorized network must be attributed with MAZ centroids and these must be attached to regular nodes so that a path can be found between them. Because walk times are also calculated between MAZs and transit stops, the non-motorized network must also be attributed with transit stops. Both MAZ centroids and transit stop nodes can be added to the all-streets network using an automated procedure. There are two approaches to maintaining the all-streets network so that these calculations can be made. The preferred method is to utilize a master network with all links and nodes; in this case, there is only one network to maintain. A subset of links are tagged with an indicator to identify if the link should be used in auto and transit skimming and assignment, since these skims are TAZ-TAZ and for efficiency not all links should be included in path calculations. Lower facility type links (residential streets) can be excluded from the network if they do not carry significant amounts of inter-zonal demand and if they are not needed to assign demand between zones (they are not on any competitive path in the equilibrium auto assignment solution for any time period). For example, Oregon Metro's assignment network includes collectors and above. The subset of links and nodes is typically referred to as the 'planning network'. However, the drawback with the master network approach is that commercial transport planning software may impose node and link limits or variable pricing, whereby large-scale networks become more costly to use. Such pricing varies by vendor.

An alternative option is to maintain two separate networks – a planning network for auto and transit coding, and an all-streets network for non-motorized calculations. In this case, network processing must be used to attribute the all-streets network with features from the planning network such as intersection control type, transit stop flags, and so on. The attribution should be done in an automated script so that changes to the transit system will automatically be reflected in the all-streets network. Note that one-way prohibitions may not apply to bicycle travel (SANDAG's bike route choice model allows traveling the wrong way on a one-way street but imposes a significant negative penalty) and should not apply to the walk mode.

An initial set of node and link attributes for the non-motorized network are given in Table 19 and Table 20 respectively. These will be updated as details are worked out regarding non-motorized calculations. Note that in the case that certain attributes cannot be identified, default coding procedures may need to be developed to impute attributes (such as intersection control).

TABLE 19: NON-MOTORIZED NETWORK NODE ATTRIBUTES

Variable Name	Description
ID	Node ID
MAZ	MAZ number of centroid
CONTROL_TYPE	1 = signalized intersection. Includes bike-specific signals but NOT pedestrian crossing aids only (e.g. RRFBs). Cross traffic must be prohibited by a red signal to be considered signalized for bikes. 2 = stop-controlled intersection 3 = roundabout 4 = two-way yield 5 = none
LOCAL	1 = local bus stop

PREMIUM	1 = premium transit stop

TABLE 20: NON-MOTORIZED NETWORK LINK ATTRIBUTES

Variable Name	Description
ANODE	From node
BNODE	To node
ONEWAY	1 = one-way street
BIKELANE	1 = conventional, striped bike lane.
BIKEPROT	1 = protected bike lane or street-level cycle track, requires some measure of physical separation from adjacent motor vehicle travel lanes for most of its length
BIKEBLVD	1 = bicycle boulevard/neighborhood greenway. Localized, discontinuous paths should also receive this designation
BIKEPATH	1 = off-network, paved, regionally significant multi-use path
BRIDGE_CENTER_SPAN	1 = Major river crossing. Note that only one link is designated as the center span for each one-way crossing. These are used to apply bridge crossing penalties.
TRMSBRIDGE [optional]	1 = above grade link reflecting unreliable elevation/slope calcs.
TRAFFICVOLUME	ADT volume on link
SLOPE	Aggregate elevation change on link

4.2 LEVEL-OF-SERVICE SKIMS

Level-of-service matrices for auto and transit are created for each of five time periods between TAZs. Non-motorized skims are created from the all-streets network between MAZs and are not differentiated by time of day. Skim matrix files must be in OMX or CSV format for ActivitySim. We recommend packing the auto and transit skims into separate OMX files (auto_skims.omx and transit_skims.omx) and using CSV format for the MAZ skims.

Passenger Auto skims

Each auto skim table in the auto_skims.omx file is stratified by occupancy, value-of-time bin, and period. The naming convention is OCC_VOT_TABLE__PER where OCC is one of {DA,S2,S3}, VOT is one of {LOW, MED, HIGH}, TABLE is one of the auto skim tables shown in Table 21, and PERIOD is one of {EA, AM, MD, PM, EV}. Note that there are two underscores ("__") before PERIOD; this is an ActivitySim requirement used to correctly index into skims in preprocessor and utility expressions.

TABLE 21: AUTO SKIM TABLES

Table Name	Description
TIME	Time (minutes) on congested path
DIST	Distance (miles) on congested path
TOLL	Toll cost ($2023 cents) on congested path
REL	(Optional) Reliability on congested path

Truck Skims

To fully support truck submodels such as the agent-based version in use by Portland Metro, a truck_skims.omx file should contain distance, travel time, and tolls skims for each of the assigned truck classes. The typical truck submodel classes are LT—light trucks also known as light commercial vehicles; MD—medium trucks defined to be single-unit goods-bearing trucks; and HV—heavy trucks defined as multi-unit goods-bearing trucks. The LT or LCV category contains service vehicles, e.g. plumbers and gardeners, plus deliveries made by passenger cars or other light-duty vehicles, e.g. meal delivery firms. However, the LT class generally uses the passenger auto skims since such vehicles are not differentially affected by restrictions or toll policies. Skimming for the MD and HV separately is advantageous for truck submodels that can make use of the difference, e.g. by supporting scenario analysis with input network differences for those classes in route restrictions, weight limits, and tolls. However, many truck submodels only have one skim for MD and HV trucks combined. While not all truck submodels treat all individual model time periods (Portland Metro's approach is to have a peak skim and an off-peak skim), to support those that do, another dimension of the truck skim OMX file could include breaking out the skims by time of day as in Passenger Auto skimming. A useful table naming convention would be CLASS_TABLE_PER where CLASS is one of {MD, HV}, TABLE is one of the skim types in the table below, and PER has the same options as in the Passenger skims or a reduced version of PK for peak and OP for off-peak. A reliability (REL) data element could be added to this structure if needed (or contemplated) by any of the Oregon MPO truck submodels.

TABLE 22: TRUCK SKIM TABLES

Table Name	Description
TIME	Time (minutes) on congested path
DIST	Distance (miles) on congested path

TOLL	Toll cost ($2023 cents) on congested path

Transit skims

Each auto skim table in the auto_skims.omx file is stratified by access mode, egress mode, and period. The naming convention is ACC_TABLE_EGR__PER where ACC and EGR is one of {WLK, PNR, KNR, BIKE}, TABLE is one of the transit skim tables shown in Table 23, and PERIOD is one of {EA, AM, MD, PM, EV}. Note that there are two underscores ("__") before PERIOD; this is an ActivitySim requirement used to correctly index into skims in preprocessor and utility expressions.

TABLE 23: TRANSIT SKIM TABLES

Table Name	Description
IVT	In-vehicle time (minutes)
BUSIVT	In-vehicle time on bus (minutes)
BRTIVT	In-vehicle time on bus-rapid transit (minutes)
LRTIVT	In-vehicle time on light rail (minutes)
COMIVT	In-vehicle time on commuter rail (minutes)
FWAIT	First wait time (minutes)
XWAIT	Transfer wait time (minutes)
XFERS	Transfers
ACC	Walk or drive access time (minutes)
EGR	Walk or drive egress time (minutes)
AUX	Auxiliary walk time (minutes)
FAR	Fare ($2023 cents)

Non-motorized skims

There are two non-motorized skim files, one for walk mode (maz_maz_walk.csv) and one for bike mode (maz_maz_bike.csv). Because these files contain non-motorized level of service, only MAZ-pairs within the maximum walk and bike distance are listed in the file. The maximum distances should be gleaned from survey data; typical maximum walk distance is 3-4 miles and maximum bike distance is 10-12 miles. Fields for each file are shown in Table 24 and Table 25 respectively.

TABLE 24: MAZ_MAZ_WALK.CSV FIELDS

Field Name	Description
OMAZ	Origin MAZ
DMAZ	Destination MAZ
TIME	Time
DISTANCE	Distance

TABLE 25: MAZ_MAZ_BIKE.CSV FIELDS

Field Name	Description
OMAZ	Origin MAZ
DMAZ	Destination MAZ
TIME_*BCL*	Time by bicycle comfort level (BCL)
DISTANCE_*BCL*	Distance on least-cost path by bicycle comfort level
LOGSUM_*BCL*	Route choice logsum or utility on least-cost path by bicycle comfort level

4.3 LAND-USE DATA INPUTS

Land-use data is specified in an MAZ data file, with one record per MAZ. Fields are shown in Table 26. The fields which can be calculated based on maintained MAZ or other data such as network information are identified as “(Calculated)”. Procedures to calculate these fields are described below.

Note that external MAZs should be created for external stations. Any facility crossing the area boundary (including freeways coded as two one way links) should be connected to one external TAZ and one external MAZ to represent all traffic into and out of the region at that crossing point. These MAZs are coded as EXTERNAL=1.

TABLE 26: MAZ LAND-USE DATA FILE

Field Name	Description
MAZ	Micro-analysis zone number
TAZ	(Optional) Transportation analysis zone number
BLKGRP	(Optional) Census block group number (documentation field – optional)
TRACT	(Optional) Census tract number (documentation field – optional)
PUMA	(Optional) Census Public Use Microdata Number (documentation field – optional)
COUNTY	Census county number (documentation field – optional)
DISTRICTXX	(Optional) District field where xx is the number of districts. Used for application of district-level constants if necessary and/or for summaries of model results.
TOTNGQHHS	(Calculated) Total non-group quarters households, calculated from synthetic population file.
TOTGQHHS	(Calculated) Total group quarters households/population, calculated from synthetic population file.
TOTHHS	(Calculated) Total households, calculated from synthetic population file.

TOTPOP	(Calculated) Total population, calculated from synthetic population file.
EMP_NRM	Natural resources and mining employment (NAICS 21)
EMP_CON	Construction employment (NAICS 23)
EMP_NHTMFG	Non-high tech manufacturing employment (NAICS 31-33 except NAICS 334)
EMP_HTMFG	High tech manufacturing employment (NAICS 334)
EMP_WT	Wholesale trade employment (NAICS 42)
EMP_TWU	Transportation, warehousing, and utilities employment (NAICS 22, 48, 49)
EMP_RCS	Retail Trade employment (NAICS 44-45,)
EMP_ IFRPBS	Information, finance, real estate, professional, and business service employment (NAICS 51-56)
EMP_HCS	Health care and social assistance employment (NAICS 62)
EMP_OSV	Other service employment (NAICS 81)
EMP_GOV	Government employment (no education, NAICS 92)
EMP_EDU	Education employment (NAICS 61)
EMP_AER	Arts, entertainment, and recreation employment (NAICS 71)
EMP_ACC	Accommodations employment (721)
EMP_FSD	Food services and drinking places employment (NAICS 722)
EMP_TOT	Total employment (Can be calculated from above fields)
ENROLL_K8	Grades K-8 enrollment (public and private)
ENROLL_912	Grades 9-12 enrollment (public and private)

ENROLL_COLL	College enrollment (Can add fields to denote classroom space, university parking supply, and major university indicator if implementing the university travel model)
TERMTIME	Terminal time (average minutes to walk to/from parked car)
PRKCST_HR	Hourly parking cost ($2023 cents)
PRKCST_DAY	Daily parking cost ($2023 cents)
PRKCST_MNTH	Monthly parking cost ($2023 cents)
PRKSPACES	Total parking spaces
ACRES	Size of MAZ in acres (Can be calculated from MAZ shapefile)
ACTIVE_ACRES	Active park acres (include urban parks, golf courses, basketball/tennis/pickleball courts)
OSPC_ACRES	Open space acres (include areas of open space with trail systems)
ESCOOACCTIME	E-scooter access time (average number of minutes to walk to a shared e-scooter, 9999 if not available)
EBIKEACCTIME	E-bike access time (average number of minutes to walk to a shared e-bike, 9999 if not available)
EXTERNAL	1 if an external MAZ, else 0
EXT_WORK_SIZE	Size term for work tours at external MAZ (total resident work tours from SWIM at the external station)
EXT_NWRK_SIZE	Size term for non-work tours at external MAZ (total resident non-work tours from SWIM at the external station)
TOTINT	(Calculated) Intersection count (total 3-way and 4-way intersections in MAZ): Note this can be calculated from the all-streets network. For future years, this can be updated based on application of intersection count/acres from other similar MAZs for greenfields.

EMPDEN	(Calculated) Sum(EMP_TOT)/Sum(ACRES) for all MAZs within 0.5 miles of this MAZ
RETDEN	(Calculated) Sum(EMP_RCS)/Sum(ACRES) for all MAZs within 0.5 miles of this MAZ
HHDEN	(Calculated) Sum(TOTHHS)/Sum(ACRES) for all MAZs within 0.5 miles of this MAZ
POPDEN	(Calculated) Sum(TOTPOP)/Sum(ACRES) for all MAZs within 0.5 miles of this MAZ
POPEMPDEN	(Calculated) Sum(TOTPOP + EMP_TOT)/Sum(ACRES) for all MAZs within 0.5 miles of this MAZ
WLKTIME_BUS	(Calculated from all-streets network) Walking time to local bus (9999 if not available)
WLKTIME_BRT	(Calculated from all-streets network) Walking time to bus-rapid transit (9999 if not available)
WLKTIME_LRT	(Calculated from all-streets network) Walking time to light-rail transit (9999 if not available)
WLKTIME_CR	(Calculated from all-streets network) Walking time to commuter rail (9999 if not available)
EXPPRK_MNTH	(Calculated) Expected average monthly parking cost from origin to all destinations ($2023 cents)
EXPPRK_DAY	(Calculated) Expected average daily parking cost from origin to all destinations ($2023 cents)
EXPPRK_HR	(Calculated) Expected average hourly parking cost from origin to all destinations ($2023 cents)
PARKAREA	(Calculated) Parking constrained area code: 1: Constrained parking area (destinations in area have an expected cost and parking location choice model is applied. 2: Buffer area (free parking in area is included in expected cost for destinations in PARKAREA 1, but destinations in area do not have an expected cost and parking location choice model not

	applied), 3: Not parking constrained or buffer area (free unconstrained parking).
Built Form (not sure on naming)	We need some indication for how walkable / comfortable the area is. In the past this has been intersection density in the base year, and a proxy for the future. I would like to see us use maybe something from Smart Location Database definitions, but open to different ideas here.

Calculated fields in MAZ data file

Density fields: These can be calculated by dividing the appropriate land-use field by area. They are calculated on a 'floating' basis by buffering around the MAZ and summing the land-use field and acreage for all MAZs within a ½ mile radius of the current MAZ (note that the current MAZ values are always included in the calculation despite how far the nearest MAZ may be from the current MAZ). The distance can be gleaned from the all-streets network using a shortest path by distance calculation, or the straight-line distance between MAZ centroid coordinates are required, which can also be calculated from the all-streets node layer. Other fields may be coded if useful in estimation.

Transit walk-time fields: These can be calculated from the all-streets network which has MAZ centroids and transit stop nodes in the network. The transit stop nodes should be updated at the start of a model run since the transit network coding could change between runs. The walk times are calculated by finding the distance to either the closest transit stop by type from the MAZ, or to the average of multiple stops. The average of stop-level boardings from a previous assignment could be used as a weight on walk time, though that would bias the results against stops introduced as part of a new alternative and would introduce simulation variance into the results. We suggest a 2.8 mile per hour walking speed to calculate time. These times are used as an override on the skimmed TAZ times.

Expected parking costs and park area: Parking costs (PRKCST_HR, PRKCST_DAY, PRKCST_MNTH) and parking spaces (PRKSPACES) are the actual cost of parking in the MAZ and the actual number of spaces in the MAZ. Expected parking cost fields (EXPPRK_HR, EXPPRK_DAY, EXPPRK_MNTH) are the average parking costs that a traveler might expect to pay when they park either in the MAZ or somewhere close to the MAZ and walk to the MAZ from their parking location. The parking cost equation is shown below, and graphically in Figure 17.

$$\text{expected cost} = \frac{\sum Cost \ * Spaces \ * \ e^{dist_{walk} \, * \, \beta_{walkdist}}}{\sum Spaces \ * e^{dist_{walk} \, * \, \beta_{walkdist}}}$$

FIGURE 17: PARKING COST GRAPHIC

Expected parking cost calculations should take into account the potential for parking for free to the extent that travelers seek out free on-street parking and walk to their destination which has an expected cost. This requires defining a ‘parking constrained area’ where MAZs in this area have a parking cost calculated for them, a ‘buffer area’ around the parking constrained area where MAZs have free on-street parking, and a ‘completely free’ area which includes all other MAZs. These parking areas are defined in the PARKAREA field as 1, 2, and 3. They can be manually coded in the MAZ file or calculated via automated geometric algorithm available in Python, as was recently done for SANDAG. The buffer area free parking calculation requires an estimate of on-street parking capacity per mile which is applied to links in the all-streets network.

4.4 SYNTHETIC POPULATION INPUTS

Inputs to PopulationSim include seed household and person data files, geographic equivalency files, total households by MAZ, a set of control files for each geography for which controls are specified, a file that maps household and person attributes in the seed data to each control, and a yaml file with properties that control the PopulationSim run. The seed household file and person file for initial model implementation are generated directly from ACS 5-year (2018-2022) Public Use Microdata files. The seed data can be updated to the most recent 5-year average for model calibration. Data for the PUMAS containing the region to be modeled are retained from the Oregon data files and the entire household and person file in CSV format is read into PopulationSim. This provides the user the option to merge any household and person attributes from the original PUMS data to the ActivitySim input or output household and person files if desired. PUMS household and person data file formats are available on the internet[34].

The MAZ data file will be used for the input total households by MAZ and the geographic equivalencies for PopulationSim. A script will pre-process each set of controls such that the total households matches the household estimates by MAZ and that total population is consistent with distributions of households by household size at appropriate geographic levels.

There is one control file for each geographic level for which controls are specified. The controls are described in Section 3.7 and are not repeated here.

4.5 SYNTHETIC POPULATION OUTPUTS

Table 27 and Table 28 show ActivitySim household and person input files respectively. These files are created from PopulationSim and post-processing and contain the attributes of synthesized households and persons derived from the seed data, plus any additional fields coded by the user used in ActivitySim. For example, the user may wish to represent over-rides for certain choice models. These can be specified by adding fields noted as model outcomes at the household or person levels (specified in Table 34 and Table 35 respectively) and setting those fields to a user-specified value, then changing the ActivitySim configuration to use those fields from the input files rather than run the choice models that create those fields. We have also added two fields that allow the user to apply person-level factors to toll costs and transit fares. Their default value should be 1.0.

[34] https://www2.census.gov/programs-surveys/acs/tech_docs/pums/2017_2021ACS_PUMS_User_Guide.pdf

TABLE 27: SYNTHETIC HOUSEHOLDS (SYNTHETIC_HHS.CSV) FILE FIELDS

NAME	TYPE	DESCRIPTION
HOUSEHOLD_ID	Integer	Unique household ID number
(Optional) SERIALNO	Integer	PUMS household number
MAZ	Integer	Household MAZ ID
(Optional) TAZ	Integer	Household TAZ ID
TYPE	Integer	Type of housing unit 1 = Household 2 = Institutional group quarters 3 = Noninstitutional group quarters
NP	Integer	Number of people in household
HHT	Integer	The type of household - Census PUMS HHT variable: N/A: Not applicable (group quarters or vacant) 1: Married couple household 2: Other family household: Male householder, no spouse present 3: Other family household: Female householder, no spouse present 4: Nonfamily household: Male householder: Living alone 5: Nonfamily household: Male householder: Not living alone 6: Nonfamily household: Female householder: Living alone 7: Nonfamily household: Female householder: Not living alone
HHINCADJ	Integer	The annual household income in dollars in the household, adjusted to the model base year ($2023)
WORKERS	Integer	The number of workers in the household

TABLE 28: SYNTHETIC PERSONS (SYNTHETIC_PERSONS.CSV) FILE FIELDS

Name	Type	Description
HOUSEHOLD_ID	Integer	Household ID number (must agree with household file)
PERSON_NUM	Integer	Person number in household (1…n where n is total number of persons in household)
PERSON_ID	Integer	Person ID number used as index (1…n where n is total number of persons in synthetic population)
PUMA_GEOID	Integer	Household PUMA Geographic ID
TAZ	Integer	Household TAZ ID
MAZ	Integer	Household MAZ ID
AGEP	Integer	Age of person in years
SEX	Integer	Sex of person in household (1 = male, 2 = female)
WKHP	Integer	Usual hours worked per week in past 12 months
WKW	Integer	Weeks worked during past 12 months -9: under 16 years old or did not work in past 12 months 1: 50-52 weeks worked 2: 48-49 weeks worked 3: 40-47 weeks worked 4: 27-39 weeks worked 5: 14-26 weeks worked 6: 13 or fewer weeks worked
ESR	Integer	Employment Status Recode -9: Under 16 years old 1: Civilian employed, at work 2: Civilian employed, employed but not at work 3: Unemployed 4: Armed forces, at work

		5: Armed forces, employed but not at work 6: Not in labor force
SCHG	Integer	School grade -9: Not in school 1: Preschool 2: Kindergarten 3-14: grades 1-12 15: Undergrad college 16: College beyond bachelor's degree
MIL	Integer	Military Service -9: under 17 years old 1: On active duty 2: Active in past, but not currently 3: Only active for training in Reserves/National Guard 4: Never served in military (currently used in SANDAG, but if not used in Oregon – this should be removed from the user guide, as an ABM person level input)
NAICSP	Integer	NAICS ID of work occupation (Census Code)
INDP	Integer	Work industry code (Census code)
OCCP	Integer	Work occupation code (Census code)
OCCCAT	Integer	1: Management, business, science, and arts; 2: Service; 3: Sales and office; 4: Natural resources, construction, and maintenance; 5: Production, transportation, and material moving; 6: Military specific occupations

DDRS	Integer	Self-care difficulty (-9: N/A (Less than 5 years old), 1: Yes, 2: No
DEAR	Integer	Hearing difficulty (-9: N/A (Less than 5 years old), 1: Yes, 2: No
DEYE	Integer	Vision difficulty (-9: N/A (Less than 5 years old), 1: Yes, 2: No
DOUT	Integer	Independent living difficulty (-9: N/A (Less than 5 years old), 1: Yes, 2: No
DPHY	Integer	Ambulatory difficulty (-9: N/A (Less than 5 years old), 1: Yes, 2: No
DREM	Integer	Cognitive difficulty (-9: N/A (Less than 5 years old), 1: Yes, 2: No
TOLLFACTOR	Float	A user-defined factor to be applied to skimmed toll values. Default should be 1.0.
FAREFACTOR	Float	A user-defined factor to be applied to transit fare (in addition to transit subsidy or transit pass factors). Default should be 1.0.

4.6 ACTIVITYSIM SETTINGS

ActivitySim Global Parameters

There are a number of global parameters stored in the ActivitySim constants.yaml file that are used in various places in the model, but mostly in tour and trip mode choice. While some of these parameters are typically held constant in future years (such as travel time sensitivity factors), others may vary by scenario. They are described in Table 29. Note that this table is not necessarily a complete description of every single property that could be in constants.yaml. There may be other properties required to implement the model system, such as codes to map 30-minute periods to five skim/assignment periods, mappings between variable codes, or parameters to be used in model specifications that should be parameterized in the file. The table will need to be finalized after the models are fully estimated and implemented. However the philosophy should be that any parameter that may need to be exposed to the user should be put in the constants.yaml file rather than specified in a utility specification file for maximum transparency.

TABLE 29: ACTIVITYSIM GLOBAL PARAMETERS

Property	Description	Default Value
Distributed travel time sensitivity and cost parameters		
distributed_time_factor_work_mean	Mean travel time for distributed time sensitivity for work tours	1
distributed_time_factor_work_stddev	Standard deviation of travel time for distributed time sensitivity for work tours	0.7
distributed_time_factor_nonwork_mean	Mean travel time for distributed time sensitivity for non-work tours	1
distributed_time_factor_nonwork_stddev	Standard deviation of travel time for distributed time sensitivity for non-work tours	0.6
distributed_time_factor_min	Minimum travel time sensitivity factor	0.1
distributed_time_factor_max	Maximum travel time sensitivity factor	10
costShareSr2	Factor for cost sharing for shared-ride 2	1.75
costShareSr3	Factor for cost sharing for shared-ride 3+	2.5
vot_threshold_low	Upper breakpoint for lowest value-of-time bin, used to match VOT with skims and for trip table segmentation ($2023 cents)	881
vot_threshold_med	Upper breakpoint for medium value-of-time bin, used to match VOT with skims and for trip table segmentation ($2023 cents)	1800
c_auto_operating_cost_per_mile	Default auto operating cost in cents per mile ($2023) to be used for non-household vehicles	18.29
Ridehail settings		
Taxi_baseFare	Base/meter drop fare for taxi mode ($2023 cents)	220
Taxi_costPerMile	Taxi mode cost per mile ($2023 cents)	23
Taxi_costPerMinute	Taxi mode cost per minute ($2023 cents)	10
Taxi_waitTime_mean	Taxi mean wait times are specified by area type (1-5). They can also be specified in the MAZ input file if more control is desired.	5.5, 9.5, 13.3, 17.3, 26.5
Taxi_waitTime_sd	Taxi standard deviation of wait time is also specified by area type (1-5)	6.4, 6.4, 6.4, 6.4, 6.4

TNC_single_baseFare	Base/meter drop fare for single-pay TNC mode ($2023 cents)	220
TNC_single_costPerMile	Single pay TNC mode cost per mile ($2023 cents)	13
TNC_single_costPerMinute	Single pay TNC mode cost per minute ($2023 cents)	24
TNC_single_costMinimum	Single pay TNC minimum fee ($2023 cents)	720
TNC_single_waitTime_mean	Single pay TNC mean wait times are specified by area type (1-5). They can also be specified in the MAZ input file if more control is desired.	3.0, 6.3, 8.4, 8.5, 10.3
TNC_single_waitTime_sd	Single pay TNC standard deviation of wait time is also specified by area type (1-5)	4.1, 4.1, 4.1, 4.1, 4.1
TNC_pooled_baseFare	Base/meter drop fare for pooled TNC mode ($2023 cents)	220
TNC_pooled_costPerMile	Pooled TNC mode cost per mile ($2023 cents)	53
TNC_pooled_costPerMinute	Pooled TNC mode cost per minute ($2023 cents)	10
TNC_pooled_costMinimum	Pooled TNC minimum fee ($2023 cents)	300
TNC_pooled_IVTFactor	Factor to apply to auto in-vehicle time to reflect extra time for diverting for passenger pickup and dropoff for pooled TNC	1.5
TNC_pooled_waitTime_mean	Pooled TNC mean wait times are specified by area type (1-5). They can also be specified in the MAZ input file if more control is desired.	5.0, 8.0, 11.0, 15.0, 15.0
TNC_pooled_waitTime_sd	Pooled TNC standard deviation of wait time is also specified by area type (1-5)	4.1, 4.1, 4.1, 4.1, 4.1
min_waitTime	Minimum waiting time for any ride-hail mode	0
max_waitTime	Maximum waiting time for any ride-hail mode	50
Active and micromobility mode parameters		
walkSpeed	Default speed for walking (MPH). Note that this field can be calculated as a person-based attribute based on age and/or mobility status in the person preprocessor instead of using a global walking speed.	2.8
bikeThresh	Maximum distance for manual biking. Note that this field can be calculated as a person-based attribute based on age and/or mobility status in the person preprocessor instead of using a global maximum bike distance threshold.	10
bikeSpeed	Default speed for human powered bike (MPH). Not necessary if using bike logsum.	7.8
ebikeSpeed	Default speed for ebike (MPH). Not necessary if using bike logsum.	10
escooterSpeed	Default speed for escooter (MPH). Not necessary if using bike logsum.	6.7

driveSpeed	Default speed for drive to/from transit (MPH). Not necessary if skimming time from auto network.	25
escooterVarCost	Shared e-scooter cost per mile ($2023 cents)	20
escooterFixedCost	Shared e-scooter base/meter drop cost ($2023 cents)	100
escooterRentTime	Shared e-scooter rental time in minutes	1
escooterAccessThreshold	Shared e-scooter maximum access time threshold after which it is unavailable	30
ebikeVarCost	Shared e-bike cost per mile ($2023 cents)	20
ebikeFixedCost	Shared e-bike base/meter drop cost ($2023 cents)	100
ebikeRentTime	Shared e-bike rental time in minutes	1
ebikeAccessThreshold	Shared ebike maximum access time threshold after which it is unavailable	30
escooterownership	Share of households that own an escooter	0.006
ebikeownership	Share of households that own an ebike	0.008
ebikeMaxDist	Maximum ebike distance in miles	10.5
escooterMaxDist	Maximum escooter distance in miles	2
Autonomous vehicle factors		
useAV	Parameter to control whether AV ownership model is applied	0
autoIVTFactorAV	Parameter to factor in-vehicle time for tour and trip mode choice using AVs (comfort, productivity)	0.75
autoParkingCostFactorAV	Parameter to factor parking costs for tour and trip mode choice using AVs (assuming some portion of AVs will park for free or return home)	0.5
autoTerminalTimeFactorAV	Parameter to factor terminal time for tour and trip mode choice using AVs (assuming some portion of AVs will drop off traveler at activity location)	0.65
minAgeDriveAloneAV	Minimum age for driving alone for tour and trip mode choice if using an AV	13
Other		
monthly_transit_pass_cost	Cost of monthly transit pass. Used in transit pass ownership model ($2023 cents).	10000

Household, Person, and Land Use Preprocessors

At the start of ActivitySim model runs, a number of columns are added to the household, person and landuse tables. These columns are derived from the available data in the input data tables and are then used in subsequent ActivitySim models. The preprocessor files are csv files that specify the name of the new column and the python expression that is used to calculate that column. The following tables list the common variables that are often included in ActivitySim implementations. Since these preprocessors are completely open to the users, additional columns that may be useful for new models or re-estimated models could be added.

TABLE 30: COMMON HOUSEHOLD PREPROCESSOR COLUMNS

Property	Description
income_in_thousands	Household income in thousands
income_segment	Income segmented in to bins [0, 30k, 60k, 100k, >100k]
number of non_workers	Number of household non-working adults]
number of drivers	Number of household drivers
num_adults	Number of household adults
num_children	Number of household children
num_young_children	Number of household children aged 5 or younger
num_gradeschool	Number of grade school children in the household
num_highschool	Number of high school children in the household
num_university_students	Number of college students in the household
num_fullTime_workers	Number of full time workers in household
num_partTime_workers	Number of part time workers in household
num_retired_adults	Number of retired adults in household
home_is_urban	True if household is in an urban area
home_is_rural	True if household is in a rural area
hh owns an ebike	True if household owns an E-Bike

TABLE 31: COMMON PERSON PREPROCESSOR COLUMNS

Property	Description
adult	True if person is aged 18 or older
male	True if person is male
female	True if person is female
pstudent	Student status categorized into non-student, grade/high school, or university
pemploy	Employment status characterized into
ptype	Person Type
has_non_worker	presence of non_worker other than self in household
has_retiree	presence of retiree other than self in household
has_preschool_kid	presence of preschooler other than self in household
has_driving_kid	presence of driving_kid other than self in household

has_school_kid	presence of school_kid other than self in household
has_full_time	presence of full-time worker other than self in household (independent of person type)
has_part_time	presence of part-time worker other than self in household (independent of person type)
has_university	presence of university student other than self in household
student_is_employed	True if a student person type is also employed
nonstudent_to_school	True if a part-time or full-time worker is also a student
is_student	True if person is student
is_preschool	True if person is a preschool person type
is_gradeschool	True if person is in grade school
is_highschool	True if person is in high school
is_university	True if person is taking university classes
school_segment	Student status categorized into preschool, grade school, high school, or university
is_worker	True if person is a worker
is_fulltime_worker	True if person is a full-time worker
is_parttime_worker	True if person is a part-time worker
home_zone_id	home zone id of the person
time_factor_work	travel time sensitivity factor for work travel sampled from a lognormal distribution
time_factor_nonwork	travel time sensitivity factor for non-work travel sampled from a lognormal distribution
occupation	occupation category based on SOC codes

TABLE 32: COMMON LANDUSE PREPROCESSOR COLUMNS

Property	Description
density_index	(household_density * population density) / (household density + person density)
is_cbd	True if zone is in the central business district
tot_collegeenroll	total college enrollment for mandatory constraint target
preschool_target	preschool target for mandatory constraint - sum of all size terms
is_parking_zone	Is a zone with parking
emp_total	Total employment excluding non-salaried workers

ActivitySim model-related files

ActivitySim requires a set of files related to each model component that is run. These include the following, which are described more fully on the ActivitySim github page[35]:

- An optional preprocessor file that create data items in a 'chooser' or decision-maker table which are necessary to run a model, for example converting continuous variables to bins used in a certain model specification, or appending data from a zone file onto the choosers table. These are named variously 'annotate_[tablename]_[modelname]_preprocessor.csv' to indicate which table they are annotating, and a model name to indicate for which model the preprocessor is run, or '[modelname]_[tablename]_preprocessor.csv'.
- A model specification file, typically named '[modelname].csv', containing one row per calculation and a set of columns for each alternative in the model.
- A model parameters file, typically named '[modelname].yaml', containing properties and parameters required to run the model.
- A model coefficients file, typically named '[modelname]_coefficients.csv', containing the coefficients applied in utility equations.
- An optional coefficient template file, typically named '[modelname]_coefficients_template.csv', containing a mapping from generic coefficients (rows) to purpose or person-type specific coefficients (columns) so that one utility specification file can be used for models with multiple segments. An example would be the tour mode choice model, which is segmented by tour purpose. The utility file lists the generic parameters while the template file maps those generic parameters to purpose-specific parameter names. The coefficients file lists each purpose-specific parameter value.
- An optional alternatives file, used in the case that the model has many alternatives which would be impractical to represent as columns in a utility specification file. These are commonly used for location choice models, destination choice models, time-of-day choice models, the school pickup/dropoff model, non-mandatory tour frequency, joint tour frequency and composition, and the vehicle type choice model. These files are typically held constant for every model run, except for cases where the zone system changes (which requires changing the location choice and destination choice alternatives files) or in the case of the vehicle type choice model where scenarios may be run which require a change to the vehicle choices or attributes of those choices (number of makes and models, costs of vehicles, operating ranges of electric vehicles, etc.). The columns in the vehicle type alternatives file can be seen in Table 33 below.

[35] https://activitysim.github.io/activitysim

Vehicle type data is only used in “option 4” which simultaneously predicts body type, fuel type, and age. The “option 2” version which supplies probabilities for fuel types does not use this vehicle type data.

- An optional post-processor file that appends data to the output chooser table. This is used to calculate variables that are derived from the model choice and may be used in downstream models or reporting. A common example of this is to calculate the distance from home to work and school after the mandatory location choice model and append those columns to the person data. Another example would be to append the appropriate disaggregate accessibility values based on the auto ownership of the household after the auto ownership model is run.

TABLE 33: VEHICLE ALTERNATIVES FILE

Field	Description
body_type	Vehicle body type with categories: car, truck, suv, van, motorcycle
fuel_type	Vehicle fuel type with categories: gas, diesel, hybrid, plug-in hybrid, electric
vehicle_year	Year the vehicle was made
NumMakes	Number of manufacturers making this alternative
NumModels	Total number of models made across all manufacturers
MPG	Miles per gallon (or equivalent for electric vehicles)
Range	Range in miles for electric vehicles
NewPrice	Prices in dollars ($2023) of average new car for this vehicle type
auto_operating_cost	Cost in cents per mile ($2023). Excludes insurance and registration.

4.7 ACTIVITYSIM OUTPUTS

ActivitySim outputs include a household file, a person file, a vehicle file, a tour file, and a trip file. Each file is attributed with unique IDs so that files can be linked to each other, some fields that are provided either in the input data from PopulationSim or derived from input attributes, and the choices made by each decision-maker in the ActivitySim model system. Additionally, ActivitySim creates both aggregate and disaggregate accessibilities.

The output household file (final_households.csv) is shown in Table 34. The table includes some of the attributes from the input household file and some derived attributes that are useful for summaries of data. Other attributes from input household data file created by PopulationSim can be appended to the table by merging with the PopulationSim output household data table by household_id. Outputs from household-level choice models are clearly indicated in the table. Outputs from person, tour, and trip level choice models can be appended to the household table by manipulating (summing, averaging, etc.) the results from those files and merging the resulting data table(s) with the household data by household_id.

The output person file (final_persons.csv) is shown in Table 35. Similar to the household file, the person file also contains some attributes of the input persons file and some derived attributes such as person type that are useful for data summaries. Also like the household file, outputs from person-level models including mobility models and tour frequency models are appended to the table. Other variables can be calculated and appended from merging tour and trip files by household_id and person_id if necessary.

The output vehicle file (final_vehicles.csv) is shown in Table 36. This file has one record for each vehicle owned by each household. The file is created entirely from the vehicle type choice model. The output tour file (final_tours.csv) is shown in Table 37. The file has one record per tour; joint tours are also represented as one record per joint tour. In other words, there are at least two participants on each joint tour, but the joint tour is represented only once on the tour file. The tour_id field is unique in the tour file and can be used to join the tour file to any calculations made in the trip file. The output trip file (final_trips.csv) is shown in Table 38. The file has one record per trip, with joint tours also represented as one record which includes multiple participants. The trip_id field is unique. A number of tour fields are provided in the trip file in order to allow an analyst to summarize trip data more easily.

ActivitySim can be configured to write out travel data used in model calculations such as the times and costs experienced by travelers as a result of the choices they made. The SANDAG ActivitySim model was configured to write out this data and will be a good starting point for the Oregon models.. The data can be summarized to understand total VMT by person or household, fuel consumption, energy consumption, greenhouse gas emissions, total travel time expenditures, total travel cost expenditures, and so on. The data can also be annualized if desired. Annualization factors, energy consumption parameters, and greenhouse gas generation parameters would need to be developed and applied to the results. These data items are stored as additional fields in the output trip file shown in Table 38.

TABLE 34: ACTIVITYSIM OUTPUT HOUSEHOLD FILE FIELDS

Field	Description
household_id	Household ID
home_MAZ	Household MAZ
GQ	Indicator for group quarters household (0,1)
income	Household income in dollars ($2023)
hhsize	Number of persons in household
HHT	Household dwelling unit type
num_workers	Number of workers in household
sample_rate	Sample rate for household
num_drivers	Number of licensed drivers in household
num_adults	Number of adults (age >= 18) in household
num_children	Number of children (age < 18) in household
auto_ownership	(Model output) Auto ownership
escooter_owner	(Model output) escooter ownership
ebike_owner	(Model output) ebike ownership

Field	Description
av_ownership	(Model output) AV ownership indicator (0,1) from AV ownership model
numAVowned	(Model output) Number of AVs owned from vehicle type choice model
workplace_location_accessibility	(Disaggregate accessibility output) Work location choice logsum
shopping_accessibility	(Disaggregate accessibility output) Shop destination choice logsum
othdiscr_accessibility	(Disaggregate accessibility output) Other discretionary choice logsum
num_travel_active	(Model output) Number of household members from CDAP with an active travel day (CDAP pattern M or N)
num_travel_active_adults	(Model output) Number of adults in household (age 18+) from CDAP with an active travel day (CDAP pattern M or N)
num_travel_active_children	(Model output) Number of children in household (age <18) from CDAP with an active travel day (CDAP pattern M or N)
school_escorting_outbound	(Model output) School escorting indicator (0,1) in outbound direction from school pickup/dropoff model
school_escorting_inbound	(Model output) School escorting indicator (0,1) in inbound direction from school pickup/dropoff model
school_escorting_outbound_cond	(Model output) School escorting indicator (0,1) in outbound direction from school pickup/dropoff conditional model
num_hh_joint_tours	(Model output) Number of fully joint tours from joint tour frequency/composition model

Field	Description
joint_tour_frequency_composition	(Model output) Joint tour frequency composition from joint tour frequency/composition model

TABLE 35: ACTIVITYSIM OUTPUT PERSON FILE FIELDS

Field	Description
person_id	Person ID
household_id	Household ID
home_MAZ	Household MAZ
age	Age in years
PNUM	Person number in household (1…n where n is total persons in household)
sex	1=male, 2=female
pemploy	Employment status: 1: full-time, 2: part-time, 3: not employed, 4: underage child
pstudent	Student status: 1: K12 student, 2: college student, 3: not a student
ptype	Person type:

Field	Description
	1: full-time worker, 2: part-time worker, 3: college student, 4: not a worker, 5: retiree, 6: driving-age student, 7: non-driving age student, 8: pre-schooler
educ	Educational attainment: 1 = No schooling completed 9 = High school graduate 13 = Bachelor's degree
naics_code	Census NAICS codes; 0 for unemployed
soc2	Census SOC2 codes; 0 for unemployed
occupation	0 unemployed; 1: Management, business, science, and arts; 2: Service; 3: Sales and office; 4: Natural resources, construction, and maintenance; 5: Production, transportation, and material moving; 6: Military specific occupations
is_adult	TRUE if age>18, else FALSE
is_student	TRUE if student, else FALSE
is_worker	TRUE if worker, else FALSE
time_factor_work	Simulated travel time sensitivity factor for work tours
time_factor_nonwork	Simulated travel time sensitivity factor for other tours
has_license	(Model output) TRUE if has a driver’s license, else FALSE
bike_comfort_level	(Model output) Bike comfort level 1: “No way no how”, 2: “Interested but concerned”, 3: “Enthused and confident”, 4: “Strong and fearless”
work_from_home	(Model output) TRUE if worker works from home, else FALSE

Field	Description
is_out_of_home_worker	(Model output) TRUE if worker has a usual out of home workplace else FALSE
is_external_worker	(Model output) TRUE if external worker, else FALSE
external_workplace_MAZ	(Model output) external station MAZ if external worker, else 0
external_workplace_location_logsum	(Model output) work location choice logsum from external workplace location choice model, else 0
external_workplace_modechoice_logsum	(Model output) work mode choice logsum from external workplace mode choice model, else 0
school_MAZ	(Model output) school MAZ from school location choice model, else 0
school_location_logsum	(Model output) school location choice logsum from school location choice model, else 0
school_modechoice_logsum	(Model output) school mode choice logsum from school mode choice model, else 0
distance_to_school	(Model output) Distance from home to school in miles, else 0, using midday SOV medium value of time skim
workplace_MAZ	(Model output) work MAZ from work location choice model, else 0
workplace_location_logsum	(Model output) work location choice logsum from work location choice model, else 0
workplace_modechoice_logsum	(Model output) work mode choice logsum from work mode choice model, else 0
distance_to_work	(Model output) Distance from home to work in miles, else 0, using the midday SOV medium value of time skim (Midday distance is typically used in destination choice as a calibration tool so as not to double count congestion which affects mode choice logsums)

Field	Description
expected_daily_parkcost	(Model output) expected parking cost at work in cents ($2023)
transit_pass_subsidy	(Model output) 1 if transit pass is subsidized, else 0
transit_pass_ownership	(Model output) 1 if owns a transit pass, else 0
free_parking_at_work	(Model output) TRUE if has free parking at work, else FALSE
telecommute_frequency	(Model output) Days worker telecommutes per week, 0-4
cdap_activity	(Model output) Daily activity pattern M: Mandatory, N: Non-mandatory, H: Home
mandatory_tour_frequency	(Model output) Mandatory tour frequency choice (0: no mandatory tours, 1: 1 work tour, 2: 2 work tours, 3: 1 school tour, 4: 2 school tours, 5: 1 work tour and 1 school tour
num_work_tours	(Model output) Number of work tours (0, 1, or 2)
num_school_tours	(Model output) Number of school tours (0, 1, or 2)
escorted_in_outbound_dir	(Model output) Result from school pickup/dropoff model. TRUE if student is escorted in the outbound direction, else FASE
escorted_in_inbound_dir	(Model output) Result from school pickup/dropoff model. TRUE if student is escorted in the inbound direction, else FASE
num_pure_escort_chap_tours	(Model output) Result from school pickup/dropoff model. Number of pure escort tours as chaperone
num_pure_escort_chap_outbound_dir	(Model output) Result from school pickup/dropoff model. TRUE if performs a pure escort activity in the outbound direction, else FALSE
num_pure_escort_chap_inbound_dir	(Model output) Result from school pickup/dropoff model. TRUE if performs a pure escort activity in the inbound direction, else FALSE

Field	Description
rideshare_escort_chap_outbound_dir	(Model output) Result from school pickup/dropoff model. TRUE if performs a dropoff activity on a work tour in the outbound direction, else FALSE
rideshare_escort_chap_inbound_dir	(Model output) Result from school pickup/dropoff model. TRUE if performs a pickup activity on a work tour in the inbound direction, else FALSE
num_joint_tours	(Model output) Joint Tour Frequency\Composition Choice Model result: Total number of fully joint tours
num_joint_shop_tours	(Model output) Joint Tour Frequency\Composition Choice Model result: Total number of fully joint shop tours
num_joint_maint_tours	(Model output) Joint Tour Frequency\Composition Choice Model result: Total number of fully joint maintenance tours
num_joint_eatout_tours	(Model output) Joint Tour Frequency\Composition Choice Model result: Total number of fully joint eating out tours
num_joint_social_tours	(Model output) Joint Tour Frequency\Composition Choice Model result: Total number of fully joint social tours
num_joint_discr_tours	(Model output) Joint Tour Frequency\Composition Choice Model result: Total number of fully joint discretionary tours
non_mandatory_tour_frequency	(Model output) Individual Non-Mandatory Tour Frequency Choice Model result: Total number of individual non-mandatory tours
num_indiv_escort_tours	(Model output) Individual Non-Mandatory Tour Frequency Choice Model result: Number of individual escort tours
num_indiv_shop_tours	(Model output) Individual Non-Mandatory Tour Frequency Choice Model result: Total number of individual shop tours

Field	Description
num_indiv_maint_tours	(Model output) Individual Non-Mandatory Tour Frequency Choice Model result: Total number of individual maintenance tours
num_indiv_eatout_tours	(Model output) Individual Non-Mandatory Tour Frequency Choice Model result: Total number of individual eating out tours
num_indiv_social_tours	(Model output) Individual Non-Mandatory Tour Frequency Choice Model result: Total number of individual social tours
num_indiv_discr_tours	(Model output) Individual Non-Mandatory Tour Frequency Choice Model result: Total number of individual discretionary tours
num_indiv_tours	(Model output) Individual Non-Mandatory Tour Frequency Choice Model result: Total number of individual non-mandatory tours
num_atwork_subtours	(Model output) At-work subtour Frequency model results: Total at-work subtours
total_home_tours	(Model output) Total home-based tours
total_tours	(Model output) Total tours including mandatory tours, tours generated by the school pickup/dropoff model, joint tours, individual tours, and at-work subtours

TABLE 36: ACTIVITYSIM OUTPUT VEHICLE FILE FIELDS

Field	Description
vehicle_id	Vehicle ID
household_id	Household ID
vehicle_num	Vehicle number from Vehicle Type Choice Model alternatives file

vehicle_type	A string field consisting of [Body type]_[age]_[fuel type] and an optional extension "_AV" if the vehicle is an autonomous vehicle
auto_operating_cost	Auto operating cost in cents ($2023)
range	Electric vehicle range in miles
MPG	Average miles per gallon or equivalent
KWH	Kilowatt hours of energy consumed

TABLE 37: ACTIVITYSIM OUTPUT TOUR FILE FIELDS

Field	Description
tour_id	Tour ID
person_id	Person ID
household_id	Household ID
parent_tour_id	Parent tour ID if this is a work-based subtour, else 0
tour_category	The category string of the primary activity on the tour. "mandatory", "joint", "non_mandatory", "atwork"
tour_type	Purpose string of the primary activity on the tour: For home-based tours, the purposes are: "work", "school", "escort", "shopping", "othmaint", "eatout", "social", and "othdiscr". For work-based subtours, the purposes are "business", "eat", and "maint".
primary_purpose	Recoding of tour_type where all atwork subtours are identified as "atwork" regardless of destination purpose

Field	Description
tour_type_count	The total number of tours within the tour_type
tour_type_num	The sequential number of the tour within the tour_category. In other words if a person has 3 tours; 1 work tour and 2 non-mandatory tours, the tour_type_num would be 1 for the work tour, 1 for the first non-mandatory tour and 2 for the second non-mandatory tour.
number_of_participants	Number of participants on the tour for fully joint tours, else 1
participant_IDs	A space-delimited string file listing IDs of each joint tour participant; blank for individual tours
origin_MAZ	MAZ number of tour origin
destination_MAZ	MAZ number of primary destination
start	Time period of departure from tour origin (1…48)
end	Time period of arrival back at tour origin (1…48)
duration	Duration of the tour including all activity episodes and travel in periods (1…48)
school_esc_outbound	For school tours where the child is being escorted according to the school pickup/dropoff model, this string field indicates the type of escorting in the outbound direction: “pure_escort” or “rideshare”
school_esc_inbound	For school tours where the child is being escorted according to the school pickup/dropoff model, this string field indicates the type of escorting in the inbound direction: “pure_escort” or “rideshare”
num_escortees	Number of children being escorted on this tour (max of outbound and inbound direction)
chaperone_ID_outbound	Tour ID of chaperone if child is escorted to school in outbound direction
chaperone_ID_inbound	Tour ID of chaperone if child is escorted to school in inbound direction
composition	Composition of tour if joint “adults”, “children”

Field	Description
is_external_tour	TRUE if primary destination activity is external to region, else FALSE
destination_logsum	Logsum from tour destination choice model
tour_mode	Tour mode string “DRIVEALONE”, “SHARED2”, “SHARED3”, “WALK”, “BIKE”, “ESCOOTER”, “EBIKE”, “WLKTRAN”, “PNRTRAN”, “KNRTRAN”, BIKTRAN”, “TAXI”, “TNC_SINGLE”, “TNC_POOLED”, “SCHBUS”
mode_choice_logsum	Logsum from tour mode choice model
selected_vehicle	Selected vehicle from vehicle type choice model; a string field consisting of [Body type]_[age]_[fuel type] and an optional extension “_AV” if the vehicle is an autonomous vehicle
atwork_subtour_frequency	At-work subtour frequency choice model result; a string field with the following values: “no_subtours”, “business1”, “business2”, “eat”, “eat_business”, “maint”, or blank for non-work tours.
atwork_subtours	Number of at-work subtours generated
stop_frequency	Stop frequency choice model result; a string value of the form [0…n]out_[0…n]in where the first number is the number of outbound stops and the second number is the number of inbound stops
stops_outbound	Number of outbound stops
stops_inbound	Number of inbound stops
VOT	Value of time for tour in dollars/hour ($2023)
time_sensitivity	Time sensitivity factor for trip

TABLE 38: ACTIVITYSIM OUTPUT TRIP FILE FIELDS

Field	Description
trip_id	Trip ID
person_id	Person ID
household_id	Household ID
tour_id	Tour ID
vehicle_id	Vehicle ID if auto trip
outbound	TRUE if trip is in the outbound direction, else FALSE
trip_num	Sequential number of trip by direction (1…n where n is maximum trips on half-tour, e.g. max stops + 1)
primary_purpose	Primary purpose of tour (see tour table)
origin_MAZ	MAZ of trip origin
destination_MAZ	MAZ of trip destination
escort_participants	Space delimited string field listing person IDs of other children escorted on this trip, else null
chaperone_tour_ID	Tour ID of chaperone’s tour if trip is a child being escorted to or from school
school_escort_direction	String field indicating whether child is being dropped off at school (“outbound”) or picked up from school (“inbound”). “null” if not a child being picked up or dropped off.
destination_logsum	Logsum from trip destination choice model. -9 if destination is tour origin or primary destination.
depart	Departure time period (1…48)
trip_mode	Trip mode string “DRIVEALONE”, “SHARED2”, “SHARED3”, “WALK”, “BIKE”, “ESCOOTER”, “EBIKE”, “WLKTRAN”, “PNRTRAN”, “KNRTRAN”, BIKTRAN”, “TAXI”, “TNC_SINGLE”, “TNC_POOLED”, “SCHBUS”

Field	Description
mode_choice_logsum	Logsum from trip mode choice model
VOT	Value of time for trip in dollars per hour ($2023)
time_sensitivity	Time sensitivity factor for trip
parking_MAZ	Parking MAZ from parking location choice model, else -1
trip_period	A string indicating the skim period for the trip (“EA”,”AM”,”MD”,”PM’,”EV”)
tour_participants	Number of joint tour participants if joint tour, else 1
tour_participant_IDs	Space delimited string field of person IDs for each joint tour participant, else null
sample_rate	Sample rate for trip
trip_veh_body	Body type of vehicle used for trip, else “null”
trip_veh_age	Age of vehicle used for trip, else “null”
trip_veh_fueltype	Fuel type of vehicle used for trip, else “null”
origin_purpose	Origin purpose of trip (see primary purpose codes)
destination_purpose	Destination purpose of trip (see primary purpose codes)
auto_time	Time spent in autos on the trip (including any auto mode time, PNR and KNR access/egress time, time in taxi/TNC-single, TNC-shared, or school bus. TNC-shared time should account for TNC-shared time factor)
auto_distance	Distance traveled in autos on the trip (including any auto mode distance, PNR and KNR access/egress distance, distance in taxi/TNC-single, TNC-shared, or school bus. TNC-shared distance should account for TNC-shared time factor)
auto_tolls	Skimmed toll costs for auto modes on the trip (including any auto mode tolls, PNR and KNR access/egress tolls, tolls for taxi/TNC-single, TNC-shared, or school bus). Considers toll discount specified in input person file.

Field	Description
walk_time	Time spent walking (includes walk mode, transit walk access/egress/auxiliary, terminal time for auto trips, and walk time from parking zone to activity location zone)
walk_dist	Distance spent walking (includes walk mode, transit walk access/egress/auxiliary, distance calculated from terminal time, and distance from parking zone to activity location).
bike_logsum	Bicycle logsum used in trip mode choice model
bike_time	Bicycle time including bike mode time and bike-transit access/egress time
bike_distance	Bicycle distance including bike mode distance and distance calculated from bike-transit access/egress time
escooter_time	e-scooter time
escooter_distance	e-scooter distance
ebike_time	e-bike time
ebike_distance	e-bike distance
wait_time	Time spent waiting for transit (first wait + transfer wait), school bus, and ride-hail modes, as well as time spent accessing shared micromobility
parking_cost	Parking costs at trip origin and destination, calculated as one-half of the costs at each end, with subsidies considered.
auto_operating_cost	Auto operating cost for trip, including auto modes and PNR/KNR-transit auto operating costs
transit_fare	Transit fare (including discounts) paid for trip
ridehail_fare	Ride-hail fare paid for trip
micromobility_cost	Cost of using shared micromobility
MPG	Miles per gallon of vehicle used for trip
fuel_consumed	Fuel consumed for trip

5.0 DESIGN IMPLEMENTATION SPECIFICATIONS

5.1 SKIMMING AND ASSIGNMENT

Time Periods

ActivitySim must have a mapping between the half-hourly periods used for mode choice, time-of-day choice, and other models and the level-of-service skims used in those models. These mappings can vary between regions to reflect different levels of congestion and transit services that typically vary between regions. Note that these periods must be mutually exclusive and exhaustive, covering the entire day, which starts at 3:00 A.M. and ends at 2:59 A.M. The time periods used to map half-hourly periods to skim periods by region are shown in Table 39.

Note that it is possible to aggregate trips into periods other than those identified for skimming and assignment. For example, a peak hour auto trip table can be created and assigned to the network using one-hour capacities. It is up to the user whether to skim this assignment and use it to represent congestion throughout the entire A.M. peak period as defined in Table 39. In most ActivitySim implementations to date, the assignment periods are consistent with the skim periods in this table; in other words, the trips assigned to the network are aggregated from the periods identified in Table 39 and assigned using capacities that are consistent with those periods. Total daily auto volumes and transit boardings can then be aggregated across those periods. Skims created from those assignments are therefore representative of average congestion levels experienced by travelers within that time period. However, the Oregon Modeling Statewide Collaborative has suggested that there may be some interest in assigning trips to the network for periods other than those identified in Table 39. Therefore we provide two additional tables; Table 40 shows the time periods used for auto assignment and Table 41 shows periods for transit assignment. For example, for the SKATS region, a one-hour auto assignment between 7 and 8 AM would be used to build auto skims for all trips occurring between 6 AM and 9 AM. Furthermore, an agency may wish to assign peak-hour trip tables for certain applications; this can be easily accommodated within the ActivitySim structure.

TABLE 39: MAPPING BETWEEN HALF-HOURLY PERIODS AND SKIMS

Period	Oregon Metro		SKATS		Lane COG		S. Oregon		Bend MPO	
	START	END	START	END	START	END	START	END	START	END
EA	3:00 AM	5:59 AM	3:00 AM	5:59 AM	3:00 AM		3:00 AM	6:29 AM	3:00 AM	
AM	6:00 AM	8:59 AM	6:00 AM	8:59 AM			6:30 AM	8:29 AM		
MD	9:00 AM	2:59 PM	9:00 AM	3:59 PM			8:30 AM	4:29 PM		
PM	3:00 PM	5:59 PM	4:00 PM	6:59 PM			4:30 PM	6:29 PM		
EV	6:00 PM	2:59 AM	7:00 PM	2:59 AM		2:59 AM	6:30 PM	2:59 AM		2:59 AM

TABLE 40: TIME PERIODS USED FOR AUTO ASSIGNMENT

Period	Oregon Metro		SKATS		Lane COG		S. Oregon		Bend MPO	
	START	END	START	END	START	END	START	END	START	END
EA	3:00 AM	5:59 AM	3:00 AM	6:59 AM	3:00 AM		3:00 AM	6:59 AM	3:00 AM	
AM	6:00 AM	8:59 AM	7:00 AM	7:59 AM			7:00 AM	7:59 AM		
MD	9:00 AM	2:59 PM					8:00 AM	16:29 PM		
PM	3:00 PM	5:59 PM	5:00 PM	5:59 PM			4:30 PM	5:29 PM		
EV	6:00 PM	2:59 AM	6:00 PM	2:59 AM		2:59 AM	5:30 PM	2:59 AM		2:59 AM

TABLE 41: TIME PERIODS USED FOR TRANSIT ASSIGNMENT

Period	Oregon Metro		SKATS		Lane COG		S. Oregon		Bend MPO	
	START	END	START	END	START	END	START	END	START	END
EA	3:00 AM	5:59 AM	3:00 AM	6:59 AM	3:00 AM		3:00 AM	6:59 AM	3:00 AM	
AM	6:00 AM	8:59 AM	7:00 AM	7:59 AM			7:00 AM	7:59 AM		
MD	9:00 AM	2:59 PM					8:00 AM	4:29 PM		
PM	3:00 PM	5:59 PM	5:00 PM	5:59 PM			4:30 PM	5:29 PM		
EV	6:00 PM	2:59 AM	6:00 PM	2:59 AM		2:59 AM	5:30 PM	2:59 AM		2:59 AM

Highway Assignment

Auto assignments will be performed using static equilibrium methods; typically bi-conjugate Frank Wolfe is used. Path cost will be calculated using a generalized cost formula shown below:

$$gen\_cost_{class} = congested\_time + (toll_{class}) * value\_of\_cost_{class}$$

where:

gen_cost_{class} is the generalized cost for the user class in equivalent minutes

congested_time is the time returned by the volume-delay function

length is the link length in miles

$toll_{class}$ is the toll in 2023 cents by class (class differentiation allows the user to specify different toll costs by user class; for example for modeling HOT lanes and toll facilities which apply different toll rates by axle)

$value\_of\_cost_{class}$ is the factor used to convert cents to minutes, by class. It is the time value of money, which can be calculated by taking the inverse of the value of time (first convert value of time from dollars per hour to minutes per cent). It can be set to the mid-point of each value-of-time bin or preferably to the 90th percentile of each bin which weights cost less in the generalized cost equation. These values can be derived from the distribution of trips by value of time output from ActivitySim for passenger trips; truck values of time can be derived from research[36] or estimated from commercial vehicle survey data.

Class refers to the assignment class. For passenger vehicles, we are assuming nine classes that vary by occupancy {DA,S2,S3} and value of time {LOW, MED, HIGH}. We also assume that there would be at least one if not more heavy truck vehicle classes, stratified by weight. Also note that groups of links can be restricted for certain user classes. For example, HOV 2 lanes are made unavailable for DA class and HOV 3+ lanes are made unavailable for both the DA class and the S2 class. Links with weight restrictions are made unavailable for heavy truck classes.

Note that the example above converts cost to minutes and the generalized cost is stated in minutes. Alternatively the generalized cost can be expressed in monetary units such as dollars or cents. It is a matter of preference which to use; because path costs are relative to each other in assignment, the outcome should be the same regardless of which units are used.

In the above generalized cost formula, the congested time is calculated using a volume-delay function. We recommend implementation of the volume-delay function developed by Pima

[36] For example see https://github.com/tlumip/tlumip/wiki/CT for an overview of truck values of time

Association of Governments which takes into account both mid-link and intersection level delay. The function uses a standard Bureau of Public Roads formulation for mid-link delay. For links with end-points that are controlled by a stop sign, traffic circle, or traffic signal, there is an additive term which assumes random arrivals at the intersection and considers the capacity of the intersection as well as the cycle length and green time share for the approach. The volume-delay function is shown in Figure 18.

FIGURE 18: VOLUME DELAY FUNCTION WITH INTERSECTION DELAY

$$T_f = T_0 * \left[1 + \alpha_l * \left(\frac{v}{C_l}\right)^{\beta_l}\right] + P * \frac{c}{2} * \left(1 - \frac{g}{c}\right)^2 * \left[1 + \alpha_i * \left(\frac{v}{C_i}\right)^{\beta_i}\right]$$

where:

- T_f is link congested travel time (minutes)
- T_0 is link free-flow travel time (minutes)
- v is link volume
- C_l is mid-link capacity
- C_i is intersection capacity for controlled intersections, which is a function of intersection geometry and g\c ratio for the approach
- g/c is green time ratio for the approach for signalized intersections, or some factor representing the g/c ratio for other
- c is the cycle length in seconds
- P is a factor representing progression for coordinated signals, or lack of progression for uncoordinated signals. It is set to 1.0 for links with no progression. It can be set to less than 1.0 for one-way streets with signal progression, subject to user judgement.
- α_l, β_l, α_i, β_i are parameters calibrated to result in monotonically increasing congestion with respect to increases in volume for link (l) and intersection (i) components (see Table 42 for terms used in the Southern Oregon activity-based model).

TABLE 42: VOLUME-DELAY FUNCTION TERMS USED IN THE SOUTHERN OREGON ACTIVITY-BASED MODEL

Link Functional Class	αl term	βl Term	αi term	βi Term
Interstate (access controlled)	0.3	6.0	2.0	2.0
All Other	0.15	4.0	2.0	2.0

We should point out that while the above volume-delay function (VDF) is useful in that it takes into account intersection level of service in the calculation of delay, static equilibrium models do not consider queuing and therefore are typically not able to represent congested travel times accurately under very congested conditions where downstream delays affect upstream links.

Auto Reliability

If it is desired to include reliability in the path cost, the VDF can be modified to include a reliability component. This should be included in the VDF because research indicates that links become more unreliable as link congestion increases. The additive function includes a reliability component that varies by volume to capacity ratio. There is also a static component of reliability that is related to aspects of the network that do not vary according to the volume-capacity ratio, such as facility type. These can be calculated in advance and do not need to be included in the VDF. The path cost under unreliable conditions is then the sum of congested time, equivalent minutes of monetary cost, and unreliability. The San Diego Association of Governments activity-based model includes auto reliability; readers are directed to technical documentation[37] for more information.

Transit Assignment

Headway-based versus Schedule-based Assignment

There are two broad categories of transit assignment algorithms available in the two software packages used by agencies in Oregon (EMME and VISUM). They include headway-based assignment and schedule-based assignment. In a headway-based network, the wait time is a probabilistic function of the passenger arrival pattern at the stop (typically assumed random over the period modeled) and the line headway. The solution used in EMME and VISUM for finding paths in the headway-based assignment is referred to as optimal strategies, originally proposed by Spiess and Florian[38]. It assumes that the passenger does not have information about the arrival time of lines at a node that serve the passenger's destination. Therefore the passenger is assumed to board the first line in the optimal choice set of lines that arrives at the node. There are several options used for the assumption of distribution of headways in an optimal strategies assignment. The original algorithm assumes an exponential distribution. In VISUM this is referred to as the "No information and exponentially distributed headways" option. This

---

[37] RSG, Pricing and Travel Time Reliability Enhancements in the SANDAG Activity-Based Travel Model: Final Report, Prepared for San Diego Association of Governments, 2016. https://tfresource.org/topics/SANDAG_C04Report.pdf

[38] Heinz Spiess, Michael Florian, Optimal strategies: A new assignment model for transit networks, Transportation Research Part B: Methodological, Volume 23, Issue 2, 1989.

algorithm is suggested for modeling very irregular transit schedules. Other options include "No information and constant headways", which is a version of optimal strategies suggested for modeling more regular schedules; and two other options which assume the passenger has more information upon which to decide which alternatives to choose from and what information is included in the decision. Headway-based assignments are more commonly for modeling urban areas and in cases where transit arrival times are somewhat random rather than operate on an exact timetable such as the case for long-distance rail service.

In a schedule-based network, timetables are specified for each transit route and wait time is a deterministic function of line schedules and the passenger arrivals at transit stops. Thus, the minimum generalized cost path for a given origin/destination pair and starting time is deterministic, and every option can be enumerated precisely. The assignment results can be summarized by individual runs of each route. This assignment algorithm is commonly used for long-distance services characterized by precise headways and timed transfers between routes.

Because the headway-based assignment is more appropriate for urban modeling conditions prevalent for most transit services in Oregon's urban areas, and because it is easier to forecast, we recommend headway-based assignments for use in the Oregon activity-based model system. In every assignment software package there exists a set of parameters that are used to influence the selection of hyper-paths and the probability of choosing a hyper-path when there are multiple options[39]. A well-designed on-board survey is invaluable for determining the values for those parameters. The expanded survey can be assigned to the transit network for the relevant time period and access/egress mode combination and the boarding rate (number of boardings divided by number of linked trips) can be calculated for the assignment as well as boardings by route. Path weights can be calibrated to find values that maximize the goodness-of-fit between estimated and observed transfers and boarding rates for each assignment.

Transit Reliability

Unfortunately there are few practical examples of models that consider transit reliability in assignment or mode choice. One of the few examples is the work performed by Parsons Brinckerhoff for Los Angeles Metro[40]. The work focused on measuring schedule adherence for a subset of transit routes and incorporating those measures in transit pathfinding and mode choice. There are two aspects of reliability modeled; adherence to arriving at the boarding stop (via extra wait time due to schedule unreliability), and adherence of schedule in-vehicle time

---

[39] In EMME, these include boarding time (min), wait time factor, weight factors for in-vehicle time, waiting time, auxiliary transit and boarding time, and a spread factor. In VISUM, these include the boarding penalty for transit and for transit auxiliary modes, the mean delay, weight factors for: initial and transfer wait times; access time, egress time and walk time; transfer time; in-vehicle time; and fare.

[40] Parsons Brinckerhoff, Westside Subway Extension Project: Los Angeles Metro Crowding, Capacity, and Unreliability Summary Report Task No. 10, Prepared for LA Metro, 2014.

versus actual in-vehicle time, to account for additional delay encountered in the transit vehicle. The former component of unreliability is measured as a function of attributes of the transit network and demand prior to vehicle arrival at the boarding node; the latter component is a function of transit network attributes and demand between the boarding location and the alighting location. Automatic Vehicle Location (AVL) and Automatic Passenger Counting (APC) data from 15 of the highest ridership transit routes in the corridor was analyzed and a model was estimated to predict extra average wait time (the difference between scheduled wait time and actual wait time) as a function of headway (binned), area type, vehicle type, and time-of-day. The model was applied to transit routes in path-building and assignment, and the additional time was skimmed for mode choice.

Non-Motorized Assignment

Walk and bike paths will be calculated based on an all-streets network.

Walk Paths and Assignment

We recommend calculating walk paths using a shortest cost path algorithm where the cost is based on link distance. Although the presence of sidewalks may be used in the path (e.g. a lower path weight would be used for links with sidewalks), maintenance of a sidewalk inventory seems like overkill for a regional travel demand model. Walk trips can be assigned to the network using the shortest path if walk volumes are a useful model output. Level-of-service variables for walk paths include time and distance (MAZ to MAZ fields are shown in Table 24 while MAZ to Transit Stop fields are shown in Table 26).

MAZ to MAZ Calculations

Walk paths must be created for both MAZ-pairs (to be used in the walk mode utility calculations) and between MAZ centroids and transit stops (to be used in transit mode utility calculations for walk access and/or walk egress utility components).

MAZ to Transit Stop Calculations

For walking between MAZ centroid and transit stop node, there are several options to consider. One option is to calculate the shortest path between the centroid and the closest transit stop or set of stops with a user-defined distance buffer[41], which can be differentiated by the type of service at the stop (local or premium). This calculation would likely cause downward bias of transit walk access and egress times since the closest stop to the MAZ centroid is not necessarily the stop that is consistent with the route(s) used between the origin and destination. Another option would be to calculate the average walk time to the closest *n* stops, where *n* is a

[41] Typically set to between 0.75 and 1.25 miles, depending on type of service at the stop. The thresholds would ideally be set based on analysis of on-board survey data.

user defined parameter. These could also be segmented by stop type. Whether this increases the accuracy of the walk time would depend upon the consistency of the stop set with the actual set of stops used for all transit demand generated by the MAZ and which stops were included in the calculation. Note that this approach has the potential to cause decreases in transit accessibility and transit demand if stops in a given network alternative are added to the choice set and cause the average time to increase for one or more MAZs.

A third option would be to implement a method developed for Metropolitan Transportation Commission. This method relies on the stop-level boardings and alightings from a previous transit assignment to set weights for each transit stop which are then used to calculated a weighted average walk time for each MAZ. This method would theoretically make the walk times more consistent with transit demand but introduces Monte Carlo variance into the average walk time calculation. Given that transit demand tends to be a low probability choice, the variance could be significant.

We suggest using the closest transit stop to the MAZ centroid as a starting point for calculation of walk to transit times. We suggest investigating alternative methods as a potential refinement, such as averaging the closest set of transit stops within walking distance. Such testing should include evaluations based on on-board survey data as well as baseline versus build model applications to ensure that alternative methods result in reasonable results.

Bike Paths and Assignment

For bicycle paths, we recommend using a stochastic bike route choice model. A stochastic route choice model differs from equilibrium assignment models used in auto assignment in that demand is spread probabilistically across a set of enumerated routes. The route enumeration step is typically performed by calculating the least cost path according to a set of path weights that vary according to a distribution. There are two alternative approaches that can be considered for implementation. Each method can be extended to consider separate user classes by bike comfort level.

Oregon Metro Approach

One approach is to implement the method used by Oregon Metro. This method is relatively simple. It uses a bicycle utility equation with a set of parameters estimated from observed route choice data to assign a cost to each link. For each TAZ pair, a single shortest path (by distance) and a single path minimizing a simplified route cost function (Table 43) is calculated. Both the shortest path distance and the disutility terms are used in mode choice utility functions for the bicycle mode.

TABLE 43: OREGON METRO SIMPLIFIED BIKE PATH COST UTILITY FUNCTION PARAMETERS

Coefficient	Attribute	Value

1.0	Length	miles
0.5	Length * Bike facility type	Miles * (0 if bike path, else 1 if lane/boulevard/protected, else 2 if no bike facility)
0.3	Length * AADT_bin_1	Miles if AADT 10k – 20k and no bike facility, else 0
1.5	Length * AADT_bin_2	Miles if AADT 20k – 30k and no bike facility, else 0
3.0	Length * AADT_bin_3	Miles if AADT > 30k and no bike facility, else 0
0.01	Turn Type	1 if left or right turn, else 0

Bicycle trips generated from the demand model are then assigned to the network according to probabilities generated by a logit model that considers up to three deterministic paths. The deterministic paths are chosen by sampling path costs calculated using the above simplified utility function randomly from a normal distribution with a mean cost equal to the cost calculated by the parameter values and a user-specified standard deviation. A more complete utility function whose parameters are given in Table 44 is then applied to each sampled path and a logit model is used to calculate the probability of each path. The assignment is segmented by purpose (commute [C] versus non-commute [NC]). The bike route choice model is implemented in VISUM with Python.

The advantage of this approach is that the bike path choice function parameters can include non-additive terms such as the proportion of the path on different types of bicycle facilities; such terms cannot be included in the path cost calculations due to non-additivity. The disadvantage of the approach is that there is no guarantee that the path choice set is consistent with the least cost path used in mode choice, and the path sets can vary across baseline to build alternative scenarios resulting in bike assignment differences that may not be explained by changes in model inputs.

TABLE 44: OREGON METRO BIKE PATH CHOICE UTILITY FUNCTION PARAMETERS

Coefficient	Attribute	Value
-1.81	Bridge w/no Bike Facility	1 if bridge link with no bike facility, else 0
-1.30	Bridge w/no Separated Bike Facility	1 if bridge link with no separated bike facility, else 0
1.03	Proportion on Bike Blvd	Percent of route length on bike boulevard
1.03	Proportion on Protected Bike Lane	Percent of route length on protected bike lane
1.57	Proportion on Bike Path	Percent of route length on bike path
-2.85	Proportion with 2-4 percent upslope	Percent of route length on 2-4% positive grade
-7.11	Proportion with 4-6 % upslope	Percent of route length on 4-6% positive grade
-13.0	Proportion with 6plus % upslope	Proportion of route length on over 6% positive grade
-2.82 (C), -1.05 (NC)	Proportion on AADT_bin_1	Percent of route length on AADT 10k – 20k and no bike facility links
-7.88 (C), -4.51 (NC)	Proportion on AADT_bin_2	Percent of route length on AADT 20k – 30k and no bike facility links
-18.89 (C), -10.3(NC)	Proportion on AADT_bin_3	Percent of route length on AADT > 30k and no bike facility links
-8.98 (C), -5.22 (NC)	Ln(Distance + 1)	Natural log of distance
-.782	Left unsig 10k-20k per mile	Number of left turns at unsignalized intersections across facilities with 10k-20k AADT

Coefficient	Attribute	Value
-1.87	Left Unsig 20k plus per mile	Number of left turns at unsignalized intersections across facilities with 20k AADT or more per mile
-0.186	Left or Through at Signal per mile	Number of left turns or through turns at traffic signals per mile
-.0483	Stop Signs per mile	Number of stop signs per mile
-0.371	Left or Right Turns per mile	Number of left or right turns per mile
-0.338	Right Unsig Cross AADT 10kplus per mile	Number of right turns at unsignalized intersections across links with 10k or more AADT per mile
-0.363	Left or Through Unsig AADT 5k-10k per mile	Number of left turns or through crossings at unsignalized intersections across facilities with 5k-10k AADT per mile
-.516	Left or Through Unsig AADT 10-20,000 per mile	Number of left turns or through crossings at unsignalized intersections across facilities with 10k-20k AADT per mile
-2.51	Left or Through Unsig AADT 20,000plus per mile	Number of left turns or through crossings at unsignalized intersections across facilities with 20k or more AADT per mile

San Diego Association of Governments Approach

The following section describes the path sampling method developed for San Diego Association of Governments[42]. For each path sampling iteration, a random coefficient vector is first sampled from a non-negative multivariate uniform distribution with zero covariance and mean equal to

[42] RSG, with Hood Transportation Consulting and Parsons Brinckerhoff, Activity-Based Model Active Transport Enhancements, for San Diego Association of Governments, 2015. Available at https://github.com/SANDAG/ABM/wiki/files/at.pdf

the link generalized cost coefficients corresponding to the path choice utility function (Table 45). In the path search, the cost of each subsequent link is calculated by summing the product of the random coefficients with their respective link attributes, and then multiplying the result by a non-negative discrete random edge cost multiplier. In some stochastic methods, the number of paths sampled is set arbitrarily. For SANDAG, path sampling is repeated until both a minimum count of paths and a preset target for the total of all path sizes in each alternative list is reached. If the total path size does not reach its target after a given maximum number of sampling iterations, sampling terminates to prevent excessively long computation time. Another aspect of the SANDAG approach is the use of a bootstrapping procedure to reduce the choice set and therefore the computational time for probability calculations. The procedure uses sampling probabilities proportional to the inverse of their estimated original sampling probabilities until a specified total overlap-adjusted size of the choice set is reached. By following this procedure, calculation of the logsum rewards the availability of multiple attractive paths fairly without unfairly penalizing the presence of one or two very attractive, frequently-sampled routes. Sample sizes increase as distance increases to account for the fact that there are more options with respect to distance. To prevent path sizes from enhancing the logsum for long trips compared to short trips, the path sizes are normalized so the size of all alternative lists is equal to one. The SANDAG bike route choice logsum model is implemented in Java.

Advantages of this approach include increased number of potential paths considered in the choice set, consistency of bike path cost and route choice probability, and greater consistency of path sets between baseline and build alternatives. Disadvantages include complexity, restrictions on terms used in path choice to strictly additive terms, and software programming language used. The path parameters are also a combination of parameters from three different sources and to the extent that some variables are correlated with each other, may be biased in application. We are not aware of any commercial transport modeling software package that has implemented this algorithm.

TABLE 45: SANDAG BIKE ROUTE CHOICE COEFFICIENTS

Variable	Coef.	Source
Distance, total (mi.)	–0.858	Monterey[1]
Distance on class I bike paths	0.610	Portland[2]
Distance on class II bike lanes	0.314	Monterey
Distance on class III bike routes	0.085	Monterey
Distance on arterials without bike lanes	–1.050	Monterey
Distance on "cycle track" class II bike lanes	0.120	None
Distance on "boulevard" class III bike routes	0.430	Portland
Distance wrong way	–3.445	San Francisco[3]
Cumulative gain in elevation, ignoring declines (ft.)	–0.010	San Francisco
Turns, total	–0.083	Portland
Traffic signals, excluding right turns and through T junctions	–0.040	Portland
Un-signalized left turns from principal arterial	–0.360	Portland
Un-signalized left turns from minor arterial	–0.150	Portland
Un-signalized crossings of / left turns onto principal arterial	–0.480	Portland
Un-signalized crossings of / left turns onto minor arterial	–0.100	Portland
Access of interstate, freeway, or expressway	–999.999	Constrained
Log of path size	1.000	Constrained

Incorporation of Bike Comfort Level in Bicycle Path Choice

Either of these approaches can be extended to consider bike comfort level as user classes with different route choice cost coefficients for each class. We would expect that higher bike comfort level riders would be less sensitive to slope, traffic levels, and lack of bicycle paths. For less comfortable riders, a route with much less traffic and presence of bicycle facilities would be much more desirable. Preferably a bike route choice model would be estimated for each bicycle comfort level defined in Table 14 using observed bicycle routes for each user class. Lacking that, a set of bike route choice parameters can be asserted for each user class with starting values shown in the above tables. However it is not clear how to proceed since the estimated coefficients were estimated for all bike comfort levels (except presumably the 'no way no how' segment). An even simpler approach would be to segment the bike utility coefficient estimated in mode choice on bike comfort level; one would expect the coefficient size to be inversely proportional to bike comfort level. However, such an approach would ignore differences in path set generation or assignment for different bike comfort levels, which is less than ideal.

Dynamic Traffic Assignment

Dynamic Traffic Assignment (DTA) is a broad class of assignment models that relies on time-dependent path and route calculations as opposed to static equilibrium methods[43]. DTA models the movement of individual vehicles or groups of vehicles over time, whereas static equilibrium models are flow-based; they assign total demand for a given time period and delay is measured using a volume-delay function which takes into account total demand and total capacity for the period modeled. In a dynamic model, the fundamental diagram describes how congestion at the exit node (reduced link outflow) is propagated upstream though the link, until it spills back onto upstream links. DTA models are therefore much less forgiving than static equilibrium models in that capacity constraints must be observed. Queues build up and propagate through the network. Excessive demand can lead to gridlock.

DTA models are more computationally intensive than static equilibrium models and also more data intensive. In addition to more standard network level of detail, the following network data is required for a DTA model:

- Intersection controls and their characteristics (e.g., signal phasing and timing)
- Turning lane configurations and turning movement restrictions
- Added pocket lanes at intersections and accurate representation of roundabouts
- Ramp meters
- Time-dependent tolls for managed lanes

Additionally, demand must be represented in more disaggregate time periods and provided for trip departure time. Typically commercial transport software provides the user some flexibility over the definition of time periods; the user can specify demand for half-hour, fifteen minute, five minute, 1 minute, or even less than one minute departure times. If the DTA model is set up to

[43] For more information on DTA, see Chiu, Yi-Chang, Bottom, Jon, Mahut, Michael, Paz, Alexander, Balakrishna, Ramachandran, Waller, Steven, & Hicks, Jim (2011) Dynamic traffic assignment: A primer (Transportation Research Circular E-C153). Transportation Research Circular. Transportation Research Board, United States of America.

use a more disaggregate temporal distribution than the activity-based model, then a method must be implemented to disaggregate the activity-based model demand to be consistent with the DTA model. Similarly, if the DTA model represents more disaggregate demand segments such as individual values of time, these must be represented in the demand profile. And the converse also must be considered; if the DTA model is used as a replacement for static equilibrium skims, level-of-service data must be aggregated from the DTA model's potentially more disaggregate temporal scale and/or user classes into matrices that can fit within the memory constraints of the demand model.

A key issue in the integration of activity-based model demand with DTA is the extent to which the activity pattern generated by the demand model is consistently simulated on the network. At any point in the activity pattern, queues can build up and the traveler may experience delays that cause subsequent activity departure and arrival episodes to become infeasible. It is not necessarily the case that assigning activity-based model demand must track the activity patterns across the day; demand can be treated as a series of completely independent trips. But if the purpose of the simulation is to model demand with day-pattern fidelity, then this issue must be addressed.

The Strategic Highway Research Program (SHRP) funded a research project[44] to address these and other issues. The research resulted in the development of a schedule adjustment procedure referred to as the individual Schedule Adjustment Module (iSAM), and a method for storing and retrieving very disaggregate level-of-service data from the DTA model, referred to as the Accumulated Database of Individual Trajectories (ADIT). iSAM is iterated within the DTA assignment procedure in an inner loop, while an outer loop feeds the level-of-service information back to the activity-based model which then produces a new set of daily plan trajectories. The approach is similar to that implemented in the MATSim software[45], which iterates a DTA model with a schedule adjustment procedure. The SHRP2 program also funded a demonstration project for a dynamic transit assignment algorithm, but the implementation was not completed. It should be noted that while schedule adjustment procedures may solve the problem of ensuring that activity patterns are consistent with simulated travel times, there is a dearth of research on how to value the loss of utility resulting from such adjustments. Therefore it is not clear what advantages such procedures may yield for the purposes of providing more or better information for decision-making.

[44] TravelWorks Integrated Models Final Report. DOT-VNTSC-FHWA-18-11, FHWA-HEP-18-075, Prepared for Federal Highway Administration Office of Planning, Washington, DC, June 2018.
45 Horni, A., Nagel, K. and Axhausen, K.W., 2016. The Multi-Agent Transport Simulation MATSim. London: Ubiquity Press. DOI: https://doi.org/10.5334/baw

Assignment Dimensions and Required Resources

ActivitySim is continually improving its performance through the course of the development of the software package. In particular, the performance of the software is the primary focus of the Phase 9A development cycle of ActivitySim which is currently taking place at the time of this writing. As such, the RAM and disk space estimates here are subject to change and are a reflection of the current state of the ActivitySim code as of early April 2024.

RAM requirements in ActivitySim also requires some context. First, ActivitySim has “chunking” functionality. Chunking refers to breaking the population up into smaller pieces that can fit in RAM as opposed to trying to do the full population all at once. This helps with memory performance but adds to the model runtime and takes some additional effort to setup[46]. For the following table, RAM estimates are the amount needed to run the model without chunking.

Estimation of the RAM and disk space requirements is based on the current SANDAG ABM3 example. This model has 3.17M population, 4947 TAZs, and 1755 total skim matrices. It currently requires a total of about 700 GB of RAM, of which 175 GB is just reserved for holding the skims in memory. Assuming the same number of skims and the same RAM to population ratio, memory estimates for each Oregon region are calculated and recorded in the table below.

TABLE 46: DIMENSIONS OF EACH REGION AND LIKELY REQUIRED DISK SPACE AND RAM

REGION	BASE-YEAR POPULATION	TRANSPORTATION ANALYSIS ZONES	DISK SPACE	RAM
Portland (Metro)	2.5M	2500		460GB
Salem (SKATS)	272k	500		47GB
Eugene (LCOG)	383k	500		66GB
Southern Oregon (RVCOG/Medford)	300k	1500		65GB
Corvallis Albany Lebanon Model (CALM)	170k	1000		35GB
Bend (Bend MPO)	160k	1000		34GB

[46] see HYPERLINK "https://activitysim.github.io/activitysim/v1.2.0/core.html?#chunk"https://activitysim.github.io/activitysim/v1.2.0/core.html?#chunk for more details

ActivitySim has Sharrow to also help with RAM requirements beyond chunking. Sharrow is an interface between ActivitySim and the python package Numba that compiles the utility expressions to make the models run faster[47]. Running ActivitySim with Sharrow turned on also dramatically reduces the memory requirements for the model due to the way it handles the data. In a non-sharrow run of ActivitySim, an ”interation” data frame is built that merges each submodel’s ”choosers” with the ”alternatives”. For submodels like mandatory tour scheduling which has a large number of alternatives and many workers making a modeling choice, the combined data frame of (workers) x (scheduling alternatives) is quite large. This is where the memory requirements of ActivitySim spike, but sharrow does not have to create this merged table. Additionally, sharrow can do skim compression which also helps dramatically reduce the amount of RAM needed for the skims. Thus, running with sharrow can reduce RAM requirements by a significant amount, typically 60-80%. However, running with sharrow has some restrictions around allowed expressions and requires additional setup to compile the expressions and encode the skims. The numbers in the above table do not consider sharrow.

5.2 DONOR MODEL IMPLEMENTATION

We recommend starting from the SANDAG ActivitySim model for Oregon’s Joint Estimation effort. The SANDAG model is the closest available model to the model design described above. It includes many of the proposed mobility models, internal-external models, the vehicle type choice model, and the school pickup/dropoff model. PopulationSim will need to be implemented for each region because it is useful to deploy the model before starting to estimate components. Land-use data and skims will need to be developed for each region according to the above specifications.

The following changes will need to be made to the donor model according to the above specification:

- There are several new model components that must be implemented in the model system. These include a driver’s license holding status model; a bike comfort level model, and an auto driver identification model. Placeholder versions of these models will need to be implemented in the model system and the choices from the models will need to be introduced into downstream models, particularly mode choice.
- Person types, purposes, and modes in the model system will need to be made consistent with the design specification.

[47] see https://activitysim.github.io/sharrow/

- The donor model pre-processors and utility terms will need to be revised to be consistent with the fields from population synthesis, land-use data fields and skim tables.
- Outputs will need to be created where the SANDAG model varies from the above specification.

We suggest deployment of a visualizer to summarize model outputs and ensure that the model is functioning correctly prior to estimation. Once data is processed and calibration targets are available (see below), the model results can be compared to targets for each region in the visualizer. After the model is implemented, estimation data bundles can be prepared and models can be estimated.

5.3 DATA PROCESSING

Data is a key requirement in building an estimated, calibrated, and validated activity-based travel model. The data requirements and processing steps depend on the source of the data and what stage of the model estimation, calibration, and/or validation the data is being used for. There are five primary sources of data often used in the development of activity-based models: household travel survey data, census data, transit on-board survey data, passive data, traffic count data, transit ridership data, bicycle counts, and travel time data.

This section will go through each of the data sources, describe how they are used in the development of an ActivitySim travel demand model, and the processing that is required for the data to be of use.

Household Survey Data

Household Travel Survey (HTS) data is the primary data component used in model development. This source of data contains the socio-demographic information of the household and the persons living in the household as well as detailed travel information for all members. This rich source of data is used in all three phases of model development: HTS data is used to estimate model coefficients, calibration targets for the majority of submodels in ActivitySim are derived from the weighted distribution of the HTS data, and statistics like VMT computed from the HTS data are a vital component of model validation.

RSG’s rMove application is the source of the HTS data for this development project. rMove is a survey that can be performed on a smartphone app or through a web-based interface. Participants in the survey answer a host of questions about their household, members of their household, and their typical travel patterns. Answers are typically broken down into three categories: household, vehicle, person, and trip. Typical household questions include the income of the household, the number of vehicles owned, the location of the household,

household type, etc. Person questions include age, gender, work status, student status, work/school location, work from home and telecommuting, etc.

If the person is participating in the survey through the smartphone version of the rMove app, the app will record the person's location throughout the day. The app will then ask questions around the trips that person took including checking the accuracy of the start and end times and locations, what the purpose of the trip was, the trip mode, and whether there were other people who took that trip with them. People participating on the web-based version of rMove input this information manually instead of having the app help track location and times.

Once raw rMove app data is captured, the first step is to go through an initial cleaning phase. This includes the removal of partially completed surveys, table formatting, assigning valid ID numbers to each household/person/trip, manually inspecting and correcting trip locations and times, and geocoding the locations to the zone system being used. The result of this processing is three primary tables: households, persons, and trips.

After processing raw rMove app data into household, persons, and trip tables, a weighting process is performed and can be summarized in the following four primary steps:

1. **Initial Expansion**: Calculating an "initial weight" based on the probability of selection in the survey sample design. This step essentially "reverses" the sample plan, providing higher initial weights to areas where less sampling occurred.
2. **Reweighting to account for non-response bias**: Performing an entropy maximization-based list balancing routine to match several key household and person dimensions to ensure the weighted data accurately represent the entire survey region (and reduce sampling biases). This routine is performed using the open-source application, PopulationSim[48]. To do this step, missing values for income, gender, and race/ethnicity were imputed for those who did not provide that information.
3. **Creating day-level weights to account for multi-day survey data:** Adjusting the day-level and trip-level data to account for the fact that smartphone respondents provided multi-day travel diaries, while online respondents provided a single-day travel diary (this is the "multi-day adjustment"). These relatively simple adjustments ensure that travel analyses accurately reflect the entire survey region for a "typical" weekday (Mon-Thu) and do not over-represent smartphone respondents with multiple travel days.
4. **Adjusting for non-response bias in day-pattern and trip rates**: Adjusting the trip-level weights by data collection method (smartphone vs. online vs. call center) to account for under-reporting biases that RSG has detected in this survey and prior travel

[48] https://activitysim.github.io/populationsim/

surveys. These adjustments help make the day and trip-level data more consistent and increase the accuracy of trip rates across survey participation methods.

The next phase of HTS survey data processing is to compute variables that relate to the activity-based model structure of ActivitySim. This step is performed via a survey processing application (SPA) whose primary purpose is to group trips into tours and works by looping through the trip records for each person. Tours start and end at home (or work for atwork subtours). For each trip record that is involved with that tour, the SPA tool transforms trips into linked trips if necessary. This is often only necessary for transit trips where access and egress trips are combined with the transit trip to create a linked trip. For example, if a person drives their car to a transit center, hops on the train, and then walks to work, those three legs will be combined into one park-and-ride linked trip. The SPA tool will also check if the trip is reported to be with other household members. If so, there is logic to see if a matching trip exists for the household member that was reported to be traveling with them. In this way, joint tours and escorting can be determined by looking at which household members are present for each leg of the tour.

The SPA tool is also responsible for the general calculation of variables used in the ActivitySim framework. This includes the determination of worker status, student status, and person types for each person, the work from home and telecommute status of each person, whether the person takes mandatory, non-mandatory, or no travel during that day, calculating the tour mode based on a hierarchy of trip modes, creating fully joint tours and determining their participants, deciding on a tour purpose based on the trip purposes in the tour, and creating atwork subtours.

Output from the SPA tool consists of a format that mimics the output of ActivitySim and includes household, person, tour, trip, and joint tour tables. This data can then be summarized in the same manner as the ActivitySim output to produce direct comparisons of model output with observed distributions. Additionally, this data is used to perform model estimation where the parameters are fit to the observed records in the HTS data. For additional details on exactly how this data is used for model estimation and calibration, please see the following sections.

Transit On-board Survey Data

One challenge with a household travel survey is that the sampling rate is relatively low when it comes to people using public transportation. To get increased statistics around public transportation usage, on-board transit surveys (OBTS) are used to supplement the transit data from the HTS and provide a more accurate and complete picture of the transit ridership market.

On-board surveys are performed by selecting a set of transit routes and surveying the people using that service on that survey day. The survey level-of-detail is often much less than what is available in the HTS data and usually just consists of a couple of household/person demographic questions and questions around the surveyed transit trip. At a minimum, the questions ask about the origin and destination of the trip, where the boarding and alighting of the transit service occurred, the purpose of the trip, transit access and egress, transfers used on the trip, and minimal sociodemographic questions like household income, auto ownership, age, and gender. The survey data is weighted to match the transit boardings observed in the region.

There are a couple of uses for onboard transit survey data. The primary usage is to supplement the HTS data when creating mode choice targets. This involves coding the on-board survey data to match the tour and trip modes used in the model. For example, if the transit skimming and assignment modes are separated by access mode and transit technology, then the onboard survey data needs to be summarized accordingly. Tour mode choice in ActivitySim is often calibrated with segmentation by tour purpose and auto ownership. For this reason, the auto ownership question in the onboard survey is important.

On-board survey data is also often used in model validation. Route-level comparisons are generally compared between the survey data and the model assignment output. Transfer / boarding rates computed from the onboard survey are used to hone in on the transfer penalties in mode choice and assignment. Overall boarding comparisons between the model assignment and the onboard survey by segments like access mode, time of day, transit technology, etc. are a primary source for transit validation of the model.

On-board survey data can also be used to directly test assignment. Demand matrices can be created from the on-board survey data and fed directly into the model assignment. If the assignment procedures are working accordingly, then the output boardings should match very closely to the ridership observed in the onboard survey. This process can help in the process of tuning transit hyper-path parameters such as transfer penalties, in-vehicle time perception factors across modes, etc.

The inclusion of on-board survey data in the model development process is not required. HTS survey data generally provides all the information required, just with much lower statistics. However, on-board survey data greatly increases the accuracy of transit ridership in the model.

On-board survey data is not used in the estimation process due to the less-detailed nature of the survey around the entire daily travel pattern of the person and the lack of complete coverage around person/household information. Table 47 lists transit on-board surveys currently available or expected to be collected in Oregon in the next few years.

TABLE 47: AVAILABLE ON-BOARD SURVEYS IN OREGON

Region	Year Collected (or expected to be collected)	Coverage (system-wide, partial) for each operator	Number of observations

Census Data

Household travel survey data has its limitations due to the sample size of the survey. The onboard survey is used to supplement the transit side of the data. Similarly, census data is often used to supplement HTS data for certain model components. The Census Bureau's American Community Survey (ACS) Public Use Microdata Sample (PUMS) data is the primary source for supplemental census information.

Census data is required for the synthetic population targets used in the population synthesis procedures (see the previous section on Population Synthesis). It is additionally used in the model calibration for a subset of models. These include work from home and telecommuting rates, auto ownership, and origin-destination county worker flows to compare against the workplace location model outputs. ACS PUMS is generally provided at Public Use Microdata Area (PUMAs) level of geography – much larger than TAZs. For this reason, the ACS PUMS and model data are summarized up to county or regional level for comparison.

One area where census data is particularly helpful is in the distribution of zero auto households throughout the region. Zero auto households represent generally around 5-10% of the population and thus generally have low statistics in the household travel survey. Census data helps cover that gap. Additionally, county-to-county work flows from census are often used to calibrate the workplace location model. These county-to-county flows from HTS data are often lacking in some OD pairs, which is where census can help fill in with additional statistics.

Census data is summarized according to the specific application and can be used as calibration targets for the model or for another data source to cross-check model outputs for validation.

Traffic Count Data

Traffic count data is used in the highway validation process. The predicted model flows on links output from assignment are compared to the observed counts on the link. To make this comparison, traffic count data must be joined to the correct network links. Traffic count data is typically provided in a format that includes the xy-coordinate of the count location, some details about the link such as its facility type, name, and direction, and the observed counts on that link.

Processing of traffic count data consists of aggregating traffic counts by time-of-day periods that match what is used in model assignment, and calculating a daily average of counts across observation days for that link. Next, the traffic count data must be joined to the model network. This is typically done by performing a script that performs a spatial join between the network links and the traffic count locations, taking into account the facility type of the link and ensuring the directionality of the link matches the direction of the count station.

Validation summaries are created after traffic counts are appropriately joined to model links. A primary validation process is to create a scatter plot of model flow vs traffic counts for all count stations. If the model perfectly agreed with the traffic counts, all points would fall along the 45-degree line. Outliers far away from the 45-degree line should be manually inspected to ensure the count locations are joined to the correct model link.

Passive Data

Passive data is a source of data that is collected without the explicit participation of individuals and generally has very high statistics, but little-to-no information on the individual travelers. The primary source of passive data is from Location-Based Services (LBS) or connected vehicle data. LBS data is a collection of xy-points in time that are collected from devices like smartphones. Passive data, and output derived from passive data, can be purchased from many different venders including ATRI (American Transportation Research Institute), Replica, Streetlight, Wejo, LOCUS, and INRIX.

The uses of passive data are limited for activity-based model estimation because the data has very limited information about the trip maker and the trip itself. Information about the household and person are essentially non-existent. Trip information is limited to primarily origin, destination, and time. However, the data can still be useful in providing another calibration source for some model outputs. These include things like origin-destination flows, trip rates, time of day distributions, and trip length frequency distributions. Passive data can also be very helpful for additional uses outside of the activity-based resident model. It can be used to validate the commercial vehicle models and the internal-external models which do not have the same level of household/ person information that is used in the activity-based resident model. At a minimum, before using passive data, the analyst should carefully compare the data to all available and relevant observed data. It may be necessary to re-expand passive data to traffic counts and/or other observed data, or reject the data altogether, depending on the source of the data and the outcomes of these comparisons.

5.4 MODEL ESTIMATION

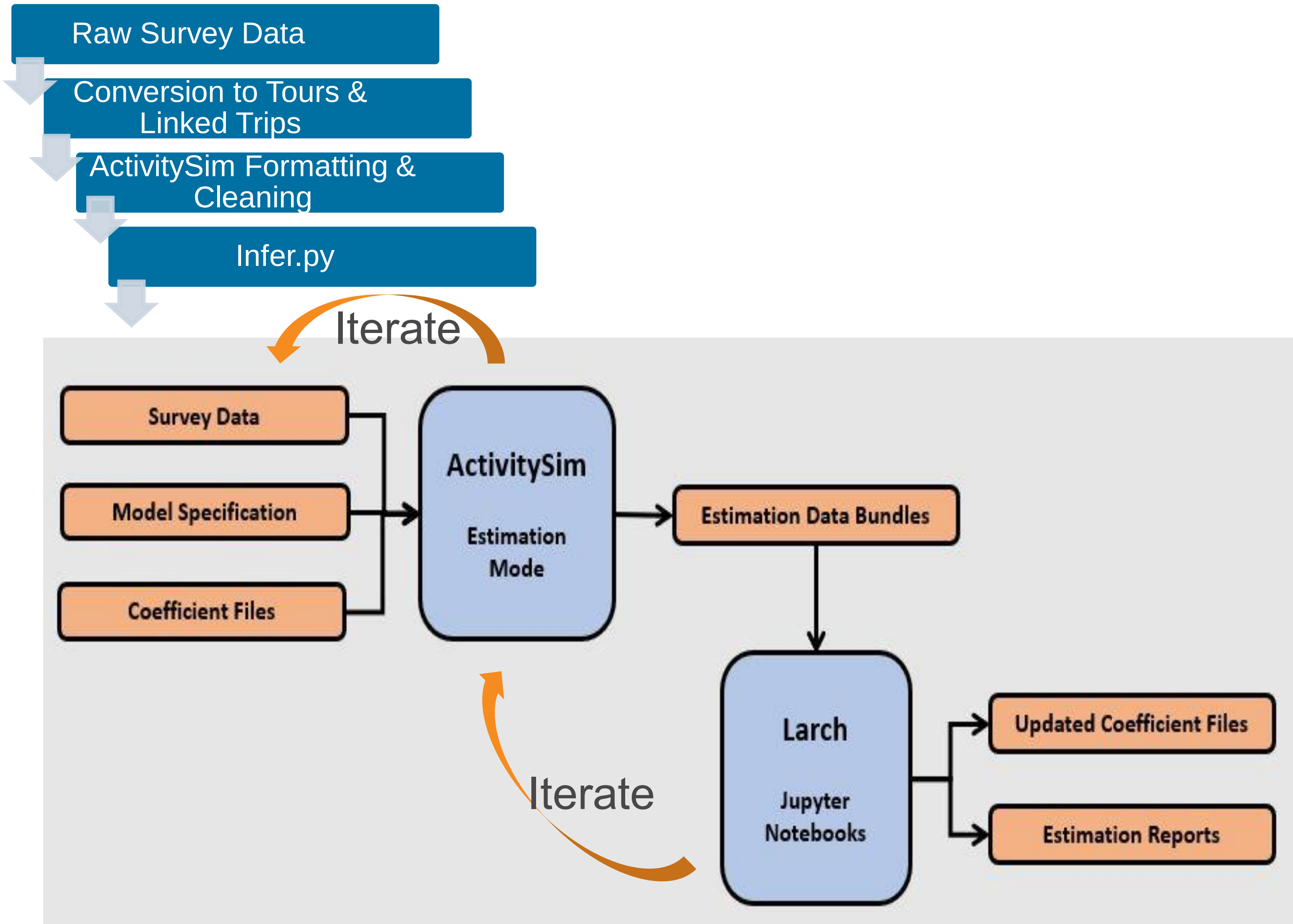
Introduction and Overview

Model estimation in the context of activity-based development is the process by which household travel survey data is fit with a given model structure to determine the coefficients of the model. These model coefficients are then used when calculating utilities to make predictions in the model application. ActivitySim allows for multinomial or nested logit model structures and each submodel in the ActivitySim framework is estimated independently.

The general process of estimation is as follows: the SPA tool processes raw survey data into tours and linked trips as described in the previous section, then the data is further cleaned to be ActivitySim compliant and processed to have all the inputs necessary to run in ActivitySim's estimation mode. Running in ActivitySim's estimation mode produces Estimation Data Bundles (EDBs) that are read by the software package Larch[49] which fits the model to the data to produce estimated model coefficients. This procedure is shown schematically in Figure 19.

[49] See https://larch.newman.me/v5.7.0/intro.html for additional documentation.

FIGURE 19: MODEL ESTIMATION WORKFLOW



There are a couple of steps in the estimation workflow that are iterative. ActivitySim is very particular about the data that it accepts in estimation. If data is missing or represents travel not captured by the ActivitySim framework, then ActivitySim will crash when trying to create the EDBs. Thus, an extensive data preparation stage is required to make sure all the data is ActivitySim compliant. Finding every single edge case for survey data that may throw off the estimation process is often an iterative procedure where ActivitySim will crash, the user needs to figure out what in the data is non-compliant, and the input survey data is modified to avoid the crash.

The other iterative part of the estimation workflow is between the model estimation and the ActivitySim config files. EDBs are created according to the specification given to ActivitySim when ActivitySim is run in estimation mode. There are often cases that exist when going through the estimation process where the user will come up with a new variable or expression to test. This often necessitates re-generating the EDBs with updated config files.

The following sections in this chapter will describe the common data preparation steps and variables needed to be able to run ActivitySim in estimation mode, actually running ActivitySim in estimation mode and the creation of Estimation Data Bundles, and the estimation procedure in Larch. Additional sections focus on how to deal with multiple survey days and multiple survey regions in the estimation process.

Data Preparation

Data preparation in this stage is centered around getting data to be completely ActivitySim compliant. In order to be compliant, every variable that is provided into ActivitySim must be a variable that could be generated by ActivitySim itself. This section will describe many of the common cases where data must be further cleaned from the output of the Survey Processing Application (SPA) described in the previous section.

If the survey data was completely accurate, and ActivitySim fully captured the travel that all individuals make, then we could just run ActivitySim in estimation mode with the formats from the SPA tool. Unfortunately, survey data is never perfect and ActivitySim does not model all possible travel scenarios. There are specific tour patterns that ActivitySim expects, and the survey data can be incomplete or inconsistent leading to ActivitySim crashes. This section contains the cleaning of the household travel survey data to get it into valid ActivitySim data and the assumptions made along the way. Considering the myriad of issues that may accompany survey data, it should be expected that other survey data sets will have issues that are not present in this document.

Households

Processing of household level data in ActivitySim is specified in the *annotate_households.csv* config file, so checking this file for region specific calculations is recommended. The household table typically requires minimal processing compared to the others, however a few modifications are still required:

- **Income:** Household income is often reported in a categorical variable in the survey and through the SPA tool. ActivitySim produces and expects actual dollar amounts for income. Income values are randomly generated by sampling from a uniform distribution of the household’s income category. Households that were missing income values are assigned a randomly selected income drawn from the distribution of income values in the survey.

- **Household Size:** The household size variable needs to match the actual number of persons in the person file for that household. If persons are removed in subsequent processing steps, this field needs to be updated.

Persons

Review of the *annotate_persons.csv* config file should be performed prior to processing person level data to understand how student and employment categories are determined and how person type is calculated. Employed persons can be identified using the ESR code, and part-time workers can be identified based on hours worked per week (WKHP) and number of weeks worked (WKW). We typically consider employed persons based on employment status code (*ESR*) equal to 1. Student status is typically determined by the person's age and grade level attending (*SCHG*) code.

Certain person and tour level models only run for people that fall within a certain employment, student, or person type category. Thus, when fixing person type issues, special care needs to be taken to ensure the correct variables are being changed in the survey processor so that the persons are annotated correctly in ActivitySim.

- **Age:** In the Household Travel Survey (HTS), there are people who decline to answer the age question. Since student status, employment, and person type each have an "age < some number" filter, these missing ages need to be fixed. If a person has a valid school TAZ and a student category less than high school, the assumption used was to set their age 12 to represent non-driving students. Otherwise, they were treated as adults and their age is set to 30.
- **School Zone:** Not all students have a valid school zone. Either the student does not provide that information in the survey or the school zone they reported has no actual enrollment in the landuse file. If a school zone was not reported, and a school trip was made, that school trip location is used as their school zone. If a school zone had no enrollment in the land use, ActivitySim would crash in estimation mode because the size term used to calculate the utility is zero. These people have their school zones replaced with the closest zone (by TAZ/MAZ centroid distance) with the appropriate level of enrollment. Typically, surveys can only get ~70% of students to have a valid school zone even after the corrections.
- **Workplace Zone:** Like school, many people who report as workers do not report a workplace zone. If these people made a work tour, the first work tour destination is selected as their workplace. If a region contains zones without employment (or households), additional logic would be needed to ensure a valid workplace zone like what was done for school. A zone needs employment data (or households) because the model needs a non-zero size term in location choice. Essentially the survey output needs to be consistent with the way the model structure is set up. If the model says that

no workers can work at the zone because the size term is zero, the survey data cannot send people there either. Zone validity can be done by segments (e.g. industry, occupation income, etc.) if deemed necessary.

There are many instances in the survey data where workers will report going to work but will not attend their primary workplace location. These tours / trips are instead changed to have the "other maintenance" purpose which contains work-related travel not associated with the primary workplace location. Work-related travel is handled in the non-mandatory tour location models and not the person-level workplace zone.

If there was no way to infer the school or workplace zones, their values were set to -1 which is ignored in ActivitySim destination choice estimation. (Workers who work from home also have their workplace location set to -1 in both the data processing and ActivitySim.) Ignoring these values is only acceptable if that person does not make a work or school tour or else the downstream models will crash with an invalid destination.

- **Person Type:** If a person is not labeled as a worker or a student, but has a work or school tour, mandatory tour frequency and scheduling models will crash because work and school location choice does not get run for those people and no tour destination is set. Everyone who makes a work or school tour needs to be labeled as a worker or student. Student and worker determination can be setup uniquely across different models. The calculation of student and worker status is typically done in the annotate_persons configuration files that get called at the start of an ActivitySim run.
- Adjusting the survey person types has been the traditional method of handling ensuring survey data is consistent with ActivitySim's allowed tour patterns. Models using census data as their synthetic population seed typically have *ESR* is set to 1 for all persons making a work tour. These models then flag them as employed in the annotate persons step. Typically, full-time workers are prohibited from taking a school tour. Thus, if a person takes a work tour and a school tour, the number of working hours and the work hours per week are decreased to fall below the part-time threshold so that both school and work tours can be performed. There are also often people who report as non-student workers but perform a school tour and not a work tour. These people are changed from workers to students. Similarly, people who report as non-worker students but only make a work tour and no school tour are changed from students to workers.
- ActivitySim has configurable settings which allow the user to choose which people are passed to the workplace and school location models. Thus, if changing person type is not the desired solution for ensuring only workers go to work and students go to school, then a custom solution can be developed. A primary change would be to reverse dependencies here – instead of calculating worker and student status from person type, determine worker and student status first and then determine person type.

Tours

In addition to determining correct school and workplace zones, tour level processing constitutes the bulk of the survey processing code. ActivitySim requires that all tours happen within a 24-hour day and that the number of tours match acceptable tour patterns.

- **Tour Purpose:** ActivitySim configuration files specify the types of tours allowed. A map between SPA output tour purposes and ActivitySim purposes must be created. Most of the categories map one-to-one with ActivitySim purposes. ActivitySim does not have a university tour type, so university tour types were changed to school. Another common tour purpose mapping is to code work-related travel as other maintenance. Loop tours are dropped from the dataset since ActivitySim does not model them.
- **Tour Category:** ActivitySim currently expects tours to be categorized into mandatory, non-mandatory, joint, and at-work subtours. Only tours that contain multiple members from the household for the entire tour are considered joint tours in ActivitySim.
- **Tour Type:** Atwork subtours contain an additional tour type variable into the ActivitySim categories *business*, *eat*, and *maint*. Work tours are mapped to *business*, eat out tours are mapped to *eat*, and all other atwork subtour purposes are mapped to *maint*.
- **Tour Mode:** Tour modes need to match the specified tour modes in the ActivitySim configuration files.
- **Tour Locations:** If tour destinations are not within the model region or are not reported, those tours do not have a valid start or end zone. These tours are removed from the resident model. If external workplace and tour destination submodels are included in the resident model, tours that leave the region and return by the end of the day are kept. Any trips that occur completely external to the region (i.e. origin and destination are outside the modeling region) are removed in the trip processing as they would not be modeled. These external workplace and tour destination submodels only account of Internal-External-Internal patterns, i.e. the person needs to start and end their day at home still. External tours that do not fit this I – E – I pattern are removed.
- **Tour Times:** The *tour_departure_and_duration_alternatives.csv* config file contains the allowed tour start and end times. Time bins from a full 24-hour day are broken down into 48 half-hour time bins starting at 3:00 A.M. Tours are typically removed if it did not fit this specification, such as tours that take place overnight or do not complete during that 24-hour period. Tour time processing could be improved by simply shifting people's time bins such that they start and end within the day. For example, people who might start at 2:30 AM could have their time shifted to 3:00 AM instead. Additional analysis would be needed to develop the exact rule structure to use as well as the implications for moving many different start / end times to the first / last time bin.

- **Tour Frequencies:** There are configuration files for each of the tour categories that specify the allowed sets of tour frequencies. ActivitySim's default setting is that a person can only take two mandatory tours with the following possible combinations: one work, two work, one school, two school, or one work and one school. A much larger set of possible alternatives exists for the non-mandatory purposes (see the config file non_mandatory_tour_frequency_alternatives.csv for a list of all allowed possibilities).

  Code was developed to count the number of tours for each person and each tour category and summarize them up in such a way that matches the ActivitySim tour frequency alternatives files. Tours are then removed if a certain tour exists outside the allowed tour frequencies. For example, if a person were to take 3 eat out tours, but the specification only allows for up to two, then the third eat out tour is removed. Tours are numbered starting at the beginning of the day and the first tours are the ones selected. A slight bias may have been introduced from this sampling method, but previous comparisons between the estimation results and survey data that did not remove these tours showed no significant difference.

  More tour frequency combinations can be allowed by adding alternatives to the non_mandatory_tour_frequency_alternatives file. This would allow the model to capture more combinations of tours but would have implications for the model's runtime and memory. Typically, the most common tour type that get discarded from this process are "work-related" tours where a person might be traveling to a bunch of different locations throughout the day. (This is especially true for work-related travel in the stop frequency model.) One alternative would be to separate out "work-related" travel from the "other maintenance" purpose and add that separately to the non-mandatory tour frequency alternatives configuration file. However, this again has implications for model runtime, memory, and downstream complexity.

If a tour was removed for any of the above reasons, all subtours for that parent tour were also removed. In a typical HTS, about 15-20% of the tour records are removed due to one or more of the above conditions not being met.

Joint Tours

ActivitySim lists joint tours only once in the tours file and instead lists the participants on each joint tour in the joint tour participants file. Additionally, all joint tours in ActivitySim are fully joint tours, meaning that all persons on the tour follow the same trip patterns and modes. Besides these formatting requirements and performing the transferrable checks on all tours including locations and start/end times, there are a few additional joint tour checks that need to be made.

- **Joint Tour Type**: ActivitySim tour types follow the same purposes as individual non-mandatory tours except there is no escorting purpose. Joint escort tours are changed to other maintenance instead.

- **Joint Tour Frequency:** Similar to non-joint tours, joint tours have restrictions on frequency.. The procedure for removing joint tours that do not fall in the frequency alternatives is the same as for non-joint tours.
- **Joint Tour Participants:** Each joint tour needs to have an adult or child on the tour to determine joint tour composition. This is achieved by ensuring only fully joint tours from the SPA output are treated as joint tours. Each fully joint tours has participants listed in the joint tour participants file.

Trips

Trips follow many of the same processing rules as tours with some exceptions around stop frequency.

- **Trip Destination:** If trip destinations are not within the model region or are not reported, those trips do not have a valid start or end zone. Trips with both their origin and destination outside of the model region are removed from the estimation data set as they are not modeled. Trips spanning the model boundary will be flagged as external trips and are included as part of the internal-external-internal tours that are part of the resident model, as discussed in the tour destination cleaning above. The external end for these trips are assigned to the corresponding external station.
- **Trip Purpose:** ActivitySim configuration files specify the types of trips allowed. A map between SPA output trip purposes and ActivitySim purposes must be created. Most of the categories map one-to-one with ActvitySim purposes, just like tours.
- **Trip Departure Time:** The departure time of the trip is given by bin number where the bins start at 3:00 AM and are in increments of half-hour. The time bin structure for trips is the same as for tours.
- **Stop Frequency:** Technically a tour-level variable, stop frequency is the number of stops on the tour in the outbound and inbound directions. The default number of stops is three for inbound and outbound directions separately, or six stops total. The default number of stops can be adjusted as part of the model implementation. Increasing the number of stops would involve adding alternatives to the stop frequency model, but this can become quite onerous because alternatives of that model are combinatorial, i.e. 0out_0in, 0out_1in,… 1out_0in, 1out_1in,… etc. After the number of allowed stops is determined, additional trips beyond the allowed number of stops are removed. Removed trips are those that take place between the maximum number of stops and the half-tour destination.
- **Trip Mode:** Trip modes need to match the modes specified in the ActivitySim configuration files.

Trips that take place on tours that were removed in the previous tour cleaning steps need to also be removed since trips must belong to a corresponding tour. Generally, the trip cleaning process removes about 30% of the available trip records in the raw survey data.

RSG has proposed new functionality as part of the next ActivitySim development cycle to be able to edit the checkpointed files created in estimation mode and run individual models with additional data. This would in theory allow for the recovery of some of this removed survey data. For example, we would be able to include a tour in destination choice even if it has bad start / end times. This checkpointing functionality would also drastically speed up the estimation mode debugging procedure by not having to start the model from scratch when running with updated data.

Infer.py

Infer.py is a python script that gets run on the cleaned survey data and adds ActivitySim specific variables to the data files before the data gets run in ActivitySim's estimation mode. This step is necessary to ensure the tour and trip records created internally in ActivitySim match the corresponding tours and trips as well as ensuring the proper observed survey values are used for all downstream models.

These tour and trip IDs are how ActivitySim matches the observed survey tours with the corresponding tour/trip generated by ActivitySim. Thus, they are an essential part of the data preparation process. Part of the infer.py module includes checking these tour frequencies and assigning ActivitySim tour ids. If there are tour or trip combinations that are not allowed, they will not be assigned an ID and the estimation process will fail.

The other main function of the infer.py script is to append model alternatives to the input survey data. For example, the output of ActivitySim's tour scheduling model is an alternative number that corresponds to a tour start, end, and duration. Rather than expecting the user to assign the alternative number to each survey record, the infer.py script will determine the alternative number based on the tour time-of-day information and the ActivitySim configuration files for that region. This process occurs for many models, including non-mandatory tour frequency, tour scheduling, school escorting, joint tour frequency, atwork subtour frequency, joint tour composition, etc.

The infer.py module takes the human-readable survey data values and derives the necessary codes that ActivitySim uses under-the-hood. The output of the infer.py script is the same household, person, tour, and trip files, but with appended columns containing the required information for ActivitySim's estimation mode, particularly the tour and trip IDs and the chosen alternative numbers for specific models.

Creating Estimation Data Bundles

Estimation Data Bundles (EDBs) are created by running ActivitySim in estimation mode and are used as input into Larch for estimating model coefficients. The process of running ActivitySim in estimation mode is much like running ActivitySim in non-estimation mode with the exception of needing to provide additional estimation config files. An example run command would look like the following:

activitysim run -c configs_estimation -c configs -d data -o output

where the run directory would contain a *configs_estimation* folder with the estimation and run setting yaml files, a normal *config* folder with the typical model configs used in non-estimation mode, a *data* folder containing the override tables from the infer.py module, land use, and skims, and an *output* folder where the estimation data bundles would be written.

The *estimation.yaml* file *is located in the configs_estimation* directory. This file contains a list for all the models for which estimation data bundles are written. At the time of this writing, multi-threading is not supported for ActivitySim's estimation mode. Thus, run times for estimation are often longer than running ActivitySim in production mode. Run time for full data samples take roughly 8 hours for ~30k households in a two zone model, but this can vary greatly depending on the zone system used and the size of the survey data.

The cleaning process can be time consuming considering a few fringe cases can cause estimation mode to crash. This means that while using a small sample is useful to test configurations and setup, problems are often not found until running the entire sample.

Debugging typically involves running the entire sample until an error arises. Large samples are needed to catch edge cases that may arise, especially how an edge case in one model can interplay with corresponding edge cases in downstream models (e.g. school escorting results interplaying with mode choice availabilities). The household(s) that crash can then be traced in a smaller sample to uncover the error. Typically, the error is a result of unclean survey data. The full survey sample is run again after fixing the issue and tested with the traced household with the test sample.

However, the full survey sample must be run from the start since the input data is changing. This means that ActivitySim's restart functionality cannot be fully taken advantage of for most problems. Without multi-threading and hours long run times to get to later models, this process can be slow. Having the necessary cleaning outlined in this design should reduce the time it takes to get estimation data bundles written out for different regions.

Estimation Data Bundles are the output created by ActivitySim's estimation mode and contain everything needed to estimate a model in Larch. They consist of the following files:

- **Model Settings:** The model settings are specified in a yaml file and detail model specific settings. For example, the tour mode choice nesting structure would be listed in the *tour_mode_choice.yaml* model setting file.
- **Model Specification:** The model specification csv file specifies the utility structure. Each row in this file contains a term in the utility specification. This file is the main file that will be modified when trying new utility equations in model estimation to obtain the best model with the most explanatory power.
- **Model Coefficients:** The model coefficients csv file lists the name of the coefficients used in the model specification as well as their starting value and whether that value should be fixed in the model estimation procedure. Fixed coefficients often correspond to a reference alternative or an availability condition.
- **Model Choosers:** The model choosers csv file contains a row for each record in the survey that is going to make a model choice. For example, the chooser file would contain a row for each household for the auto ownership estimation data bundle and the chooser file for tour mode choice would list a tour in each row. The columns of the model chooser data correspond to the rows in the model specification, i.e. each utility expression is a column in the chooser data.
- **Model Alternatives:** The model alternatives csv file is only used for the “interaction simulate” models in ActivitySim that have many different alternatives. These models include the destination choice models, the scheduling models, and school escorting. This file would have the alternative listed as the columns and the rows would be a list of the utility expressions evaluated for the alternative for each chooser. So, if there were 10 tours, 5 utility expression terms, and 100 destination alternatives, the total number of rows would be 10*5 = 50 and the number of columns would be 100. As such, the size of this table can grow to be quite large (up to 10 GBs) for some models.

Estimation Data Bundles are written out for each individual model for ActivitySim as well as for each model segmentation. For example, non-mandatory tour scheduling models are often broken down by purpose. Each tour purpose would have its own EDB and contain just the observed survey records for that purpose so they can be estimated independently of each other.

Included in the model choosers table of the EDB are two columns: *override_choice* and *model_choice*. The *model_choice* column contains what the model would have chosen with the provided specification file and the *override_choice* column contains the observed value from the survey data. Thus, comparisons of how well a model estimation fits the survey data can be performed by comparing these two columns and can help validate the estimated model.

Model Estimation in Larch

The actual estimation of model coefficients is performed outside of the main ActivitySim application through the Larch python software package which is built on top of the SciPy python package. The user interacts with Larch through a Jupyter estimation notebook. The estimation notebooks load the EDB into Larch, allow the user to review the loaded data and specification, run the model estimation, and observe the estimated model fit and coefficient values and associated statistics. Many example estimation notebooks exist on the ActivitySim repository[50].

The final report out of the estimation notebooks is written to an easy-to-digest excel spreadsheet that lists the coefficient names, estimated values, standard error, t-statistic, and significance. These estimated coefficient values are then transferred back into the ActivitySim configuration files for use in model production.

The process of getting a valid model estimation specification is often more art than science. The user must ensure that the specification is functional, including proper reference coefficients, bug-free expressions, and utility equations that are not over-specified. Users change the utility equation by editing the model specification csv file in the estimation data bundle that is loaded into Larch.

As mentioned in the introduction section, getting the best estimated model is often an iterative procedure where the user will modify the specification, run ActivitySim to generate the EDB, and see how the estimation responds. The goal is to produce model estimations that predict the observed distributions as well as possible while producing coefficient values that can be explained, are statistically significant, and are temporally stable for forecasting.

Handling Multiple Survey Days

ActivitySim models a single 24-hour day, but many surveys now will record a single household for multiple days. This means that the survey data processing and estimation needs to be very careful in how to handle households that record data for more than one day.

Households that are surveyed for multiple days must be copied to produce unique household records when input into ActivitySim's estimation mode. This procedure is typically performed by duplicating rows in the survey household table by the number of days surveyed and assigning a new household id column. The same goes for persons. The new household and person ids are then propagated forward into the person, tour, and trip files. The result of this process is that each survey day will show up in the ActivitySim files as a unique household that performs one day's worth of travel. Flags are added to the data to keep track of whether a household or

---

50 https://github.com/ActivitySim/activitysim/tree/main/activitysim/examples/example_estimation/notebooks

person record was duplicated in this process. Tours and trips are day-level attributes and are assigned to the appropriate household and person.

While the procedure of copying household and person attributes to create self-contained travel days solves the problem of running the survey data through ActivitySim's estimation mode, further care needs to be taken in the model estimation. Because records are copied, those copied records need to be removed from the estimation data set before loading the data into Larch. Improper estimation results would arise for instance when estimating auto ownership, because household records that were surveyed using the smartphone app would appear seven times (the week long duration of the survey), whereas households filing out a single day of the online survey appear only once. The user would use the flags created in the data duplication procedure to remove the duplicated records as appropriate.

There can be additional nuance to this procedure for some models that are predicting day-level attributes but at a household or person level. For example, the CDAP model is coordinated across all household members, but if children are only surveyed on a single day, only that day can be used to determine CDAP, even if the household has records for adults on all days. Similarly, the school escorting and joint tour models should only be estimated with household-days where the whole household participated in the survey.

Joint Estimation Across Multiple Regions

This project proposes using data from different regions to create a set of joint model estimations that can then be applied to all regions.

The first step in this process is for all participating regions to run their specific survey data through the survey data processing and estimation data bundle creation procedures independently. It is recommended that all regions use and contribute to a single data processing pipeline so that work is not duplicated, and bugs are easier to manage across multiple regions. However, since each region will have their own set of geographic features including land-use and network level-of-service, each region will have to run ActivitySim independently in estimation mode to create separate estimation data bundles.

Once the estimation data bundles are created for all regions, they will be combined at the start of the estimation notebooks to create a single joined EDB. This will necessitate that all regions work with model configurations that are as close as possible for the merge process to be performed smoothly. As part of the merge process, a flag indicating the geographic region of the model should be added such that region specific constants can be estimated and evaluated for significance. . These constants can be applied in production mode to the corresponding region. In this way, subtle regional differences can be captured as part of the estimation while still producing one joint model estimation across all regions.

Auto ownership can be explored as a simple example. Each region will need to processe household level survey data to cap the number of autos at the maximum allowed by the model design (likely four). A common set of auto ownership configuration files are used across all regions in ActivitySim's estimation mode to generate EDBs for each region. Each region's EDB's chooser table will contain rows for each of the household in the survey data and columns that correspond to the utility expressions specified in the ActivitySim config files.

At the start of the estimation, the Jupyter notebook will append all the EDBs into a single large EDB where the rows will contain household records from all regions and the columns will still be the same expressions since all regions used the same config files. This single EDB would then be loaded into Larch and coefficients would be estimated on the full entire dataset. The estimated model coefficients would then be used in production across all regions with any region-specific constants from the estimation applied appropriately.

Auto ownership is a simple example because the alternatives are the same across all regions. The merge of the destination choice models will be more complicated because the zone system is not the same across regions. It is expected that the merge for destination choice models can either be performed through a re-mapping of zone indexes across regions or concatenating both rows and columns to expand the alternatives table in the EDB to create the estimation. The final merge strategy will likely be determined by a balance between usability and performance where the first re-mapping solution is not as intuitive and produces more code overhead, vs the more explicit concatenation approach which would be simpler from an implementation perspective but may be untenable in terms of table size and estimation speed.

At the end of the estimation process, there will be one singular estimation used across all regions for each ActivitySim submodel.

5.5 MODEL CALIBRATION

The development and use of transportation system planning models can be divided into four stages: estimation, calibration, validation, and application. We refer to, the process of using observations of individual travel behavior (primarily taken from surveys such as household travel surveys and transit on-board surveys) to fit statistical models that relate characteristics of the persons, transportation system, and land uses to the likelihood of observing the choices that were made as model estimation. Every statistical model estimated using a sample of the population contains some error, which can be random if arising from chance or systematic if arising due to a mismatch between the sample and the population of interest or between the measurement of variables and their true values.

In model calibration, we iteratively adjust the travel model until it generates a demand that reasonably fits travel patterns in the observed data. The process of calibration adjusts for error in the sample used to estimate model components and error introduced by using the results of upstream models in downstream decisions. Calibrating an ABM involves adjustments to different sub-components. Most models in the ActivitySim framework are modeled as a Multi-Nominal Logit (MNL) or Nested Logit (NL) Model. Calibrating an MNL or NL model involves updating the alternative-specific constants (ASCs) and/or other model parameters to move the aggregate model predictions in the desired direction.

Generally, the process starts with comparing observed distributions of a given travel attribute (for example, the number of tours by person type and purpose) against the predicted outputs. The ABM Visualizer is used to make this comparison. The following steps are implemented if the model distributions do not match the target distributions.

1. ASC adjustments are calculated in a spreadsheet or using a Jupyter notebook as follows for each alternative:
    a. $New\ ASC = Old\ ASC + \ln\left(\frac{Target\ Proportion}{Model\ Proportion}\right)$
2. The ActivitySim specification and coefficients file for the appropriate model is updated with the new coefficients.
3. The model is run with the updated coefficients.
4. If the model reasonably matches the observed data, stop; otherwise, go to step 1.

Each of the series of models in ActivitySim influences other models in the system. Not only are long-term decisions made in early models used to determine short-term decisions in later models, but the parameters in the later short-term decisions also influence long-term decisions through the inclusion of the maximum expected utility, or “logsum”, from key short-term models as an explanatory variable in some of the long-term models. Therefore, several models must be calibrated in tandem.

Because of these influences between model components, calibration of ActivitySim should be performed in an iterative, cyclic fashion. The calibration team progresses through each step in the model system in order, adjusting or re-estimating model parameters until the aggregate outcomes from the model matches the targets. Each cycle of re-calibration achieves increasingly precise matches between the modeled outcomes and the targets as the influences between model components reached an equilibrium.

Table 48 lists anticipated calibration summaries and data sources for each model component. Most of the calibration summaries rely on household travel survey (HTS) data. Other summaries rely on census data; either 5-year American Community Survey (ACS) data, 5-year Public Use Microdata Sample (PUMS), or other tract level data. Mode choice targets are often assembled from both HTS data and transit on-board survey (OBS) data. These datasets are described in more detail above. Most of the calibration summaries are based on shares or percents of persons, person-days, tours, and trips by model choice and one or more segmentation variables. For example, person-based mobility model summaries are often segmented on person type. Tour level models are often segmented on purpose. Mode choice model summaries are segmented on purpose and auto parity (autos compared to drivers). Such segmentation follows closely the calibration constants used to adjust the model. We typically keep the segmentation fairly simple to ensure that the calibration process is efficient and focused on choice model outcomes themselves rather than the explanatory variables that influence the choice, whose coefficients were statistically estimated. It is possible to further summarize choice outcomes by explanatory variables to ensure that results are logical, but hopefully this process would be conducted as part of model estimation. If models are borrowed from another region, or conditions and/or travel behavior changes significantly between when models were estimated and the base-year for calibration, the calibration process may need to be more significant.

We refer to model validation as the process of comparing the reasonableness of model outcomes to independent estimates of travel not used in the estimation or calibration process. There are two types of model validation typically performed. Cross-sectional model validation is performed by comparing the results of the highway and transit assignments to highway counts and the number of boardings by transit system operator. Predicted model speeds can also be compared to observed speed data. Sometimes a ‘hold-back’ sample of travel survey data is used to compare against predicted model choices, though this is much less common than the other types of cross-sectional validation. Longitudinal model validation involves sensitivity testing, where input variables are systematically modified and the model is run to evaluate the reasonableness of changes in model outcomes compared to changes in its inputs. Measures of elasticity are sometimes used as a way to evaluate the reasonableness of model sensitivities.

Model calibration and validation is often an iterative process because assignment results can influence calibration of certain components; for example, comparison of assignment results to traffic counts on screenlines might indicate the need to improve the quality of the input network or land-use data, modify the skimming process, or introduce a an adjustment in destination choice. A boarding rate comparison in transit assignment might indicate the need for a lower or higher transfer penalty in mode choice. It is typically the case that not all calibration and validation targets can be achieved simultaneously. Therefore model calibration and validation requires judgement on the part of the model developer to weight the tradeoffs in matching different targets and concluding the effort when the models are reasonably calibrated and validated and ready for applications.

Part of model validation is testing the model to ensure the model is working properly and has reasonable sensitivities to changes in model inputs. TMIP-EMAT has the functionality to automatically generate a design of univariate sensitivity tests, given an exploratory scope. This tool systematically generates experiments where all of the inputs but one are set to their default values, and the holdout input is then set to alternative values. Since TMIP-EMAT enables inputs to take on a range of values rather than just a discrete data point, the use of TMIP-EMAT promotes a robust sensitivity testing process.

TABLE 48: ACTIVITYSIM CALIBRATION SUMMARIES

Model Type	Model	Summaries
Mobility models	Driver license holding	Share of persons by driver license holding status and person type (HTS)
	Bike comfort level	Share of persons by bike comfort level and person type (HTS)
	Autonomous vehicle ownership	N.A.
	Auto Ownership	Share of households by vehicles available and workers (Census, ACS PUMS) Share of households by vehicles available and household income (Census, ACS PUMS) Share of households by vehicles available and district (CTPP, ACS 5-year summaries) Share of 0-auto households by Census Tract (Census)
	Work from home	Share of workers by work from home status and district (Census, ACS 5-year summaries)
	External worker identification	Share of workers by external worker indicator and district (HTS)
	External worker location choice	External worker work tour length distribution and average (HTS) Share of external worker attractions at each external station (HTS)
	Internal worker location choice	Home to work average distance and trip length frequency distribution (HTS) Share of workers by place of residence and place of work, district level (HTS, CTPP, ACS 3 or 5-year summaries)
	School location choice	Home to school average distance and trip length frequency distribution by school purpose (K-12, college/university) (HTS) Students by place of residence and place of school by school purpose (K-12, college/university), district level (HTS)
	Vehicle type choice	Share of vehicles owned by body type, fuel type, and age bin (HTS)
	Transit subsidy	Share of persons by transit subsidy indicator and person type (HTS)
	Transit pass ownership	Share of persons by transit pass ownership indicator and person type (HTS)
	Free parking eligibility	Share of workers by free parking available and district, if available (HTS)
	Telecommute frequency	Share of workers by telecommute frequency (HTS)
Day and tour level models	Coordinated Daily Activity Pattern	Share of persons by person type and daily activity pattern (HTS) Share of households by presence of fully joint tours and household size (HTS)
	Mandatory Tour Frequency	Share of mandatory tour generation model alternatives by person type (HTS)
	Fully joint tour frequency/composition and participation	Share of fully joint tour generation/composition alternatives (HTS) Share of fully joint tours by number of persons participating (HTS)

	Individual non-mandatory tour frequency	Share of non-mandatory tours by purpose, number, and person type (HTS) Total number of individual non-mandatory tours by person type (HTS)
	At-work subtour frequency	Share of work tours by at-work subtours
	Non-mandatory tour destination	Home to primary destination average distance and trip length frequency distribution (HTS) Share of tours by origin and primary destination district (HTS)
	Tour scheduling	Share of tours by departure, arrival, and duration half-hour period and purpose (HTS)
	School pickup/dropoff	Share of work tours with outbound stops for dropoff (HTS) Share of work tours with inbound stops for pickup (HTS) Share of students dropped off (HTS) Share of students picked up (HTS)
	Vehicle availability	Share of vehicles made available by body type and occupancy (HTS)
	Tour mode	Share of tours by tour purpose, mode and auto parity (HTS and transit on-board survey) Share of transit tours by mode of and origin/destination district (transit on-board surveys and/or other data)
	Intermediate stop frequency	Share of tours by number of outbound and inbound intermediate stops and tour purpose (HTS) Share of tours by number of trips per tour by tour purpose (HTS)
Stop and trip level models	Stop purpose	Share of stops by tour purpose and stop purpose (HTS)
	Stop destination	Intermediate stops by tour purpose and out-of-direction distance distribution (HTS)
	Stop scheduling	Share of intermediate stops in outbound direction by half-hour departure period and tour purpose (HTS) Share of intermediate stops in inbound direction by half-hour arrival period and tour purpose (HTS)
	Trip mode	Share of trips by tour purpose, tour mode, and trip mode (HTS and transit on-board survey) Transit trips by access mode and trip distance (HTS and transit on-board survey)
Overall Summaries		Total tours Total stops Total trips Average tour distance Average trip distance Total vehicle miles of travel

Validation Summaries	Traffic counts and VMT by volume group Traffic counts and VMT by facility type Traffic counts by screenline Transit boardings by operator Transit boardings by transit mode Number of transfers

5.6 MODEL APPLICATION SCENARIOS

The following section describes a few model application scenarios and how the user would test them in the current model design. We also posit how we expect the model to react to each scenario.

Congestion pricing scenario

In this test, tolls are imposed on facilities to manage congestion. The tolls are coded such that they are unavoidable for any trip entering or exiting the priced area; this is often referred to as “cordon pricing”. Real-world examples of cordon pricing include Singapore, Central London, Stockholm, and Milan. The New York City Metropolitan Transportation Authority recently voted in favor of a congestion pricing plan that imposes a $15 base fare for cars entering the southern part of Manhattan.

There are a few considerations for coding the toll and its treatment in the model. These include:

1) Time-of-day considerations. Higher toll rates might be specified during peak periods to attempt to shift travel out of peaks and into off-peak periods. Or tolls may be higher during the day compared to night in order to encourage shopping and recreational activities in the priced area. In Section 4.1 we recommend specifying separate toll fields by skimmed time period in order to give the user the option to represent tolls by time-of-day.

2) Subsidies offered and what types of households, persons, or vehicles they might apply to. For example, subsidies might be offered to low-income households, persons with disabilities, and/or low-emission/zero emission vehicles. These subsidies can largely be handled in ActivitySim configuration files.

3) The extent to which revenue raised by the tolls would be spent on improving transit service to the priced area. This has implications for network coding.

4) Whether there are differential tolls offered for trucks, taxis, or other commercial vehicles. This will affect the commercial vehicle model results and also may require adjustments to toll treatment for ride-hail modes in ActivitySim.

Tolls should be coded on the auto network according to the recommended network representation for toll costs in Section 4.1. We recommend coding the full toll cost in the network by vehicle class. Coding tolls in the network will affect not just trips with an origin and destination in the cordon area but also potential cut-through trips. Zone pairs with an origin or destination in the cordon area will have a toll cost in the toll skim matrix equal to the sum cost of all tolled links in the path. Other zone-pairs and user classes may have a toll cost if the tolled path is included in the equilibrium solution for auto assignment according to the value-of-time specified for the user class.

Subsidies and discounts can be applied in tour and trip mode choice pre-processor files. For example, low-income household discounts can be applied based on the income of the chooser. Similar discounts can be specified for disabled persons (we suggest including disability status indicators from PUMS data as person attributes in the synthetic population in Section 4.5). Toll discounts for low emission/zero emission vehicles may affect not just the mode choice decision but also the vehicle type choice model and the availability model; households may be more likely to purchase more fuel efficient cars and drivers traveling to the priced zone may choose to use more fuel efficient cars for travel to the priced zone.

Introduction of a cost term to the vehicle availability model would address the latter behavior. The model currently considers distance in the utility equation. The model can be modified to consider toll cost as well. We suggest using the same cost coefficient and sensitivity as the tour mode choice model as a starting point and adjusting those sensitivities if necessary in scenario testing. Consideration of toll cost in the vehicle type choice model is a bit more challenging because there are no destination choice logsum terms in the utility equations. Such terms could be introduced but would require asserting the coefficient on the term since estimating such terms would likely not be successful.

We suggest that whenever such discounts are introduced in the model, the user specify the discount as a percent to be applied to the toll for a specific segment in the constants.yaml file so that these can be easily adjusted for different scenarios or turned off completely. Discount policies can be challenging to define in advance of the specific policies and segments so we expect the user to define these within a particular model application. Training on how to introduce such terms into the model is essential.

We also note that it is possible that there would be some inconsistency between the toll specified in the assignment model and the toll considered in mode choice due to these subsidies; in the case of cordon pricing this would be true for pass-through trips but not trips with an origin or destination in the priced area. For other toll policies the difference may affect more trips.

We expect the following outcomes in the case of a toll pricing policy.

- Mode choice will shift based on the value of the toll, the value of time for the traveler, and the availability and utility of other modal options. We would generally expect more transit, non-motorized, and micromobility travel to and from the priced area. We would expect more shared-ride travel due to cost sharing among vehicle occupants. We would also expect more fuel efficient vehicles to be used for auto travel to and from the priced area if discounts are made available.

- Destination choice will shift based on the value of the toll, the value of time for the traveler, and the utility of other destinations. Since the work and school location choice models are constrained to match the percent of employment and enrollment by MAZ, we would not expect total work or school tours to the priced area to change unless employment and/or enrollment is changed. However, we would expect the distribution to change such that tours and trips in origins outside the priced area with better transit service, non-motorized accessibility, or micromobility accessibility will increase and other origins would decrease, and vice-versa. Destination choice for non-mandatory models will change more significantly since these models are not doubly-constrained. We would expect less tours and stops in the priced area for non-mandatory models. We would expect that overall tour and trip lengths to and from the priced area would decrease. This may increase the tour and trip lengths for origins and destinations that are not in the priced area, depending upon the spatial distribution of activities in the region.
- If differential tolls by time-of-day were applied, we would expect shifts in travel by time of day, out of the higher priced periods and into the lower priced periods. These would be based on changes in mode choice logsums by time of day.
- We expect small shifts in the magnitude of tours and stops generated, depending upon the household location. These would be based on changes in origin-based accessibilities calculated from the disaggregate accessibility calculator.
- We expect some shift in auto ownership from more to less autos owned, especially for households living close to or within the priced area. These would be based on changes in origin-based accessibilities. We would expect more fuel efficient vehicles to be owned for those households if discounts are made available for such vehicles.
- We would expect higher transit pass ownership for travelers living close to or within the priced area.
- We would expect diversion around the priced area for through-trips.
- We would expect VMT and emissions to decrease during priced time periods.

All-Road Use Fee

An all-road use fee such as a Vehicle Miles Travelled (VMT) tax has been discussed in Oregon as a replacement for the fuel tax. A number of alternatives have been proposed for its implementation, including odometer checks and GPS technology. Regardless of its collection method, modeling such policies in the model system should be relatively straightforward.

If the vehicle tax applies to all vehicles regardless of vehicle type, it can be represented as an additive fee to auto operating cost in mode choice and auto assignment. The mode choice calculation can be implemented in tour and trip mode choice pre-processors, in which the auto operating cost is set from the vehicle chosen for each auto occupancy level based on the vehicle availability model choices. One would assume those costs would also be passed on to users by ride-hail vehicles, so these costs may also need to be adjusted.

If the costs vary by time of day, that would also need to be addressed in the pre-processors – such that changes are in Python, versioned, and replicable. If the costs vary by vehicle type, we suggest modifying the auto operating cost in the vehicle alternatives file to represent the base cost, and adding a period-specific cost or multiplier in the pre-processor. That way, the vehicle type model will react to the cost difference between vehicle types and the mode choice model (and other upstream models) will react to the different costs by time of day. Note that if the fee is a replacement for the gas tax, then the fuel portion of the auto operating cost should be eliminated before adding the fee.

As mentioned, the cost also needs to be represented in the generalized cost equation used in auto assignment (see Section 5.1). If the fee varies by vehicle type, we suggest that the average per-mile fee set in the auto generalized cost equations be calculated from the output trip file for each user class (occupancy and value-of-time combination). These can also be defined for each combination of class and time period if the fee varies by time of day.

The expected results of an all-road use fee depends on the difference between the auto operating cost in the baseline scenario and the auto operating cost plus fee in the road fee scenario.

- For travelers with a higher fee, we expect less auto use, including a shift from single-occupant to shared-ride vehicles, transit, non-motorized, and micromobility. For travelers with a lower fee, we expect more auto use. Shift to ride-hail use changes might occur as well depending upon comparative cost between baseline and build scenarios and whether costs are passed on to riders via higher fees.
- We expect time of day shifts to occur based on differences in auto costs if the fee varies by time of day. Travelers whose cost increases during peak periods would be shifted out of those periods into lower cost periods. This would be captured by mode choice logsums in time of day choice models.
- There could be destination shifts towards closer destinations for travelers whose costs increase, and towards further destinations if costs decrease. This effect would be captured by mode choice logsums in destination choice models.
- Tour frequency may be affected by a road user charge. Travelers whose costs increase would be expected to make less tours due to decreases in origin-based accessibilities.
- If costs vary by vehicle type, we expect more vehicles to be owned whose cost decreases between baseline and build scenarios and less vehicles to be owned whose

cost increases. We also expect less vehicles to be used whose cost increases and more vehicles to be used whose cost decreases.

- We would expect vehicle miles of travel to decrease for travelers whose cost increases and to increase for travelers whose cost decreases.

High Occupancy Toll Lanes

High occupancy toll lanes, also referred to as HOT lanes, are high occupancy vehicle (HOV) lanes that allow vehicles that don't meet occupancy requirements to pay a toll to use the lane. Single occupant vehicles pay to use the facility, while multi-occupant vehicles use the facility for free or at a reduced cost. Variable pricing is used to manage the lane so that reliable performance is maintained. The concept behind the lanes is to sell the excess capacity where high-occupancy vehicle demand is below a certain threshold, raising revenue while encouraging ridesharing and transit. There are a number of different types of HOT lanes. These include HOT2+, where single occupant vehicles pay a toll and shared-ride vehicles are free, and HOT3+, where single occupant vehicles and shared-ride 2 vehicles pay a toll and shared-ride 3+ vehicles are free. Examples of HOT lanes include I-15 FasTrak in San Diego (HOT2+), I-25, I-36 and I-70 Express Lanes in Denver (HOT3+), I-15 Express Lanes in Salt Lake City (HOT2+), and I-85 Express Lanes in Atlanta (HOT3+). There are a number of others in the US and abroad.

Several facilities started as HOV lanes and were then converted to HOT lanes due to excess capacity and public concern. Some of today's HOT3+ facilities started as HOT2+ lanes but were re-classified due to increased demand (e.g. Katy Freeway in Houston). It should be noted that when one considers the amount of capacity required to maintain reliable free-flow travel times (around 1600 vehicles per lane hour), relatively small changes in HOV demand can significantly affect the occupancy requirements and/or toll rates charged to drivers. If HOV demand is a significant portion of capacity, the amount of excess capacity to sell to single occupant vehicles is limited. This will require either a significant price (to ensure a reasonable level-of-service on the facility) and/or a high occupancy restriction. The ability of the travel demand model to differentiate between shared 2 and shared 3+ vehicles is therefore an important consideration for planning these facilities.

The following considerations need to be defined in order to code and model HOT lanes

- The access and egress points for the facility. Some lanes have continuous access, other facilities have very limited access and egress points.
- The occupancy restrictions and pricing regime for the lane. Will trucks be allowed to use the facility? Will HOV2 pay no toll, a reduced toll, or the full toll?
- The price for each user class and the way in which the price will be set. Some facilities have variable costs for each access/egress combination. Other facilities charge a fee at

a series of points along the corridor. What is the time period over which fees will be adjusted?

- Discounts and subsidies offered. These could include discounts for fuel efficient vehicles, or subsidies for low-income households.
- Whether and which transit routes would be re-routed to use the facility.

HOT lanes are coded as separate, parallel links to the general-purpose lanes. Unlike HOV lanes, SOVs (and HOV2 vehicles) are not prohibited from using the lanes. However, the link toll cost fields should be set according to the occupancy regime. For this reason, we recommend specifying toll cost fields by auto occupancy as described in Section 4.1. For general planning purposes, toll costs can be calculated in terms of an average cost per mile for each link by occupancy/class and time period. Demand weighted averaging (using baseline traffic counts in the corridor) can be used to convert hourly cost to multi-period costs. For more complicated entry\exit tolling regimes, costs can be approximated with link costs using linear programming. An example of how to perform these calculations is described in the SANDAG SHRP2 C04 report on pricing and reliability[51].

Network procedures may also be implemented to dynamically adjust the price for each occupancy level based on demand. Such methods typically iteratively adjust price and re-run assignment based on some maximum threshold of demand (e.g. 1600 vehicles per lane hour) on the HOT lane. This implies both an inner loop of price adjustment with traffic assignment and an outer loop iterating the demand model with the updated price. Method of successive averages on price for each occupancy level can be used to ensure that convergence (measured by reducing demand oscillations between iterations to a maximum preferred level) is reached. If such dynamic procedures are not implemented, manual adjustment of prices are typically required to ensure a reasonable level of service in managed lanes. This process can be time consuming and tedious.

If discounts are provided by vehicle type, these can be considered in ActivitySim, but would introduce differences in assignment if not also considered in assignment via the introduction of new trip tables. For example, if single-occupant drivers of hybrid and/or electric vehicles are allowed to use the facility for free, this can be considered by adding rules to the tour and trip mode choice pre-processors to account for the discount. However, if single occupant electric vehicles are assigned with other single occupant vehicles, they would see the SOV toll cost on the HOT lane in traffic assignment and therefore would tend to avoid the facility. Therefore we suggest to add a set of SOV-free trip tables to the assignment to ensure that the toll cost they are exposed to in assignment is consistent with the subsidy.

[51] RSG, Pricing and Travel Time Reliability Enhancements in the SANDAG Activity-Based Travel Model: Final Report, San Diego Association of Governments, June 2016.

The effects of HOT lanes on demand depend on the value of time for the traveler, the size of the toll by occupancy level, and the time savings offered by the managed lane, as well as the lane configuration in the baseline alternative. For example, we would expect a HOT lane that adds capacity to have different outcomes than an alternative where single-occupant vehicles can pay to use an existing HOV lane. For these reasons, the analysis of HOT lane results requires a solid understanding of the demand model and the alternative(s) being modeled. In general, we expect the following outcomes:

- If the HOT lane is adding capacity to the corridor (as opposed to converting an existing HOV lane), we may observe an increase in HOV and transit usage (to the extent that transit can use the facility). We may see some shifting to SOV in the corridor for higher value of time travelers who are willing to pay the price to use the facility.
- If the HOT lane is replacing HOV capacity with HOT lane, we may see some reduction in HOV demand in the corridor as the SOVs buying into the facility may cause some reduction in travel time on the HOV lane. We may even see more shifting to SOVs depending upon the price and time savings offered by the facility. We would not expect to see significant changes in transit demand if there is no additional transit capacity added to the corridor.
- Changes in upstream models are dependent on the extent to which travelers value the time savings and their cost sensitivity. Travelers with a high value of time may shift into peak periods of travel, and changes in destination choice may also occur.
- We do not expect significant changes in tour frequency or auto ownership.
- We may observe changes in vehicle type and use in the HOT lane corridor depending on whether subsidies for such vehicles are offered. This would require changes to the vehicle type choice and vehicle availability models as discussed above in the All-Road Use Fee description.

Car Free/Car Limited Downtowns

In the car free downtown scenario, vehicles are prohibited from entering a central business district or employment campus that is several zones in size. In a car limited scenario, parking zoning laws are changed for certain districts such that buildings no longer need to supply parking for residence and customers, resulting in a reduction or complete elimination of parking supply.

We assume that in the baseline scenario, the parking area calculations identify the MAZs in the car free downtown as parking constrained area (see Section 4.3). Starting from the baseline scenario output MAZ file, the user would modify the number of parking spaces field for corresponding MAZs to zero. The user would turn off the code that calculates parking area in the build scenario so that the same parking constraint applies in the build scenario. This will cause the discounted price of parking for the parking constrained area to change to only those zones outside the parking constrained area since there would be no parking supply within the area. The user would also change the terminal time specified in the MAZ file to the expected walk time to/from the parking constrained area from outside the area, since travelers would have to walk from their parked vehicle into and out of the area. Finally, auto links in parking-free areas would need to be disabled to disallow vehicles into, out of, and through the area. We suggest connecting zonal centroid connectors to non-disabled links representing parking options outside the parking-free area to ensure that an auto path can still be built to the area, but the auto time and cost would represent the auto parking destination outside the area.

We would expect to observe the following model outcomes from auto-free/auto-limited downtowns:

- We expect a shift from auto to non-auto modes because of the increase in terminal time to the parking constrained area. If parking outside the area is priced, the shift will be more pronounced.
- We expect a decrease in non-mandatory destinations to the parking constrained area due to a lower utility of travel, and a shift to non-parking constrained areas. We expect the same number of mandatory destinations due to the double-constraint mechanism for work and school travel.
- We expect no traveler to choose to park in the area and increased parking demand outside the area where parking is available.

This scenario may also impact employment in the car-free zone; iteration with a land-use model may be warranted.

Micromobility (Electric bikes\scooters)

The Oregon design includes both privately-owned and shared e-bikes and e-scooters as discrete alternatives in mode choice. Policies in which these modes are made more available, as well as changes in speed and range, will be directly modeled via changes in the MAZ data (which specifies availability of shared micromobility), or changes in the constants.yaml file (which specifies vehicle maximum range, speed, and shared micromobility cost).

We expect to observe the following shifts in response to these settings

- We expect a mode shift from active transport, transit, and auto modes (in that order of percentage) to e-bikes/e-scooters if increased ownership of these modes is assumed.

- We expect similar increases if the costs of shared e-bikes/e-scooters is decreased or availability is increased.

The model design does not currently explicitly account for micromobility as an access or egress mode to/from transit. If such policies were of interest, one could modify the tour and trip mode choice utility equations for walk access and egress to assume that micromobility might be used under certain circumstances (e.g. when the walk to/from transit distance is greater than 0.5 miles and the origin or destination is within a mobility hub with relatively easy access to shared micromobility, for example). Such a scenario could also be modeled within the transit path-finder, depending upon the commercial software package, though this would also require changes to the assumptions regarding which transit stop the user accesses.

5.7 SCENARIO MANAGER

ActivitySim models generally rely on a slew of model parameters in order to forecast the demand in the current or future year scenario. A number of these parameters including auto operating cost, taxi and TNC fare, micromobility cost, and AV ownership penetration are usually assumed to change by forecast year or scenario. Manually changing these parameters requires the model user to know where each parameter is located, and individually changing them according to the scenario forecast values. A scenario manager, therefore, can be a convenient and efficient tool to automate this process. Below we describe one potential solution to scenario management. This solution is implemented in the SANDAG activity-based model and can be adopted for use in Oregon deployments.

An ActivitySim Scenario Manager is a python script that reads in a CSV input file containing the parameter values for each scenario, and updates the associated parameters in the ActivitySim config files. An example of the input parameter CSV file is shown in Table 49, where each row is associated with a specific scenario year/name. The parameter names used here can either be identical to the parameter names used in ActivitySim, or different. In case the parameter names are different, a separate file is used to map the parameters names between the input CSV and ActivitySim config files.

TABLE 49 AN EXAMPLE OF A SCENARIO MANAGER INPUT FILE

Scenario Year	AOC fuel	AOC maintenance	Taxi baseFare	Taxi costPerMile	Taxi costPerMinute
2012	13.5	6.3	1.78	1.87	0.08
2014	12.9	6.3	1.78	1.87	0.08
2015	19.5	6.2	1.78	1.87	0.08
2016	10.7	5.6	1.78	1.87	0.08
2017	10.8	5.5	1.78	1.87	0.08

The bulk of parameters in ActivitySim are stored in a configuration file named constants.yaml, although depending on the ActivitySim setup, some models' specific parameters may also be set in that model step's YAML file directly. An example of the setup of a constants.yaml is shown in Figure 1920.

```yaml
## ActivitySim
## See full license in LICENSE.txt.

scenarioYear: 2022
NO_EXTERNAL: 0
PRE_COVID: 0
max_local_walk_dist: 0.85
max_prm_walk_dist: 1.2
max_mix_walk_dist: 1.2
walk_speed: 3
max_local_walk_time: 17
max_prm_walk_time: 24
max_mix_walk_time: 24
HHT_NONFAMILY: [4, 5, 6, 7]
HHT_FAMILY: [1, 2, 3]
PSTUDENT_GRADE_OR_HIGH: 1
```

FIGURE 1920 AN EXAMPLE OF AN ACTIVITYSIM YAML FILE CONTAINING MODEL PARAMETERS

In running the scenario manager, the script will receive the scenario year and model run directory from the user, read in the scenario parameters from the input CSV file, and updates the ActivitySim YAML files accordingly. This script is run at the beginning of each model run to ensure the correct parameter values throughout the full run.

This procedure would need to be extended to enable the user to apply the properties specified in Table 49 to the VISUM configuration files as well.

An extension of this scenario manager approach would be to utilize TMIP-EMAT for running scenarios. Similar to the scenario manager approach discussed above, TMIP-EMAT also utilizes a set of scripts (i.e. YAML files) to specify the set of parameters to utilize for each model run. However, TMIP-EMAT utilizes an experimental design to determine which set of scenarios are run. The experimental design lays out a list of model experiments (i.e. scenarios) to be conducted, and can be constructed in a number of different ways: it can be a systematic design, a random design, or something in between. Developing an experimental design using TMIP-EMAT requires that an exploratory scope is defined, to identify the model inputs and their potential values. The scope provides a high-level group of instructions for what inputs and outputs a model provides, and what ranges and/or distributions of these inputs will be considered in the analysis. This allows for inputs to take on a range of values rather than just a discrete data point.

5.8 POTENTIAL SIMPLIFICATIONS FOR SPECIFIC IMPLEMENTATIONS

There are a number of steps that agencies can take to simplify the model system for specific implementations if desired. These include the following:

Use the existing TAZ system for MAZs. The two-zone version of ActivitySim could be applied with a one-to-one correspondence of MAZs to TAZs. This would speed up deployment and potentially improve runtime. If this is done, the user would probably want to use the walk time from the transit network for walk transit access/egress times rather than use an MAZ level override as specified in the design. The user may also want to use the planning network for walk and bike times rather than an all-streets network (see below) since the planning network is likely to be consistent with the spatial aggregation in the TAZ layer. The disadvantage of this approach is aggregation bias with respect to land-use variables and non-motorized times.

Use the Planning Network for Non-Motorized and Micromobility Times. The current design specifies the development and maintenance of an all-streets network which would be used for walk and bike times. A potential simplification would be to use the planning network for non-motorized and micromobility times instead. This would speed up deployment and reduce maintenance costs for the model, but at the disadvantage of less accurate walk and bike times and potentially less opportunity to measure the benefits of bike infrastructure investments. If the planning level network is used instead of an all-streets network, it is probably unnecessary to maintain a separate MAZ system as well (see above).

Use a volume-delay function that does not consider intersection delay. The current design calls for using a volume-delay function that includes measurement of delay at controlled intersections. It may take some time to develop data to implement that function. If data is not available, the agency may wish to use a more traditional function that does not consider intersection delay. The disadvantage of doing so is that the model would be less accurate in terms of representing delay at controlled intersections and the model system would not be sensitive to certain types of intersection improvements such as adding turn lanes.

Use aggregate value of time calculations. The current design recommends a heterogeneous representation of travel time sensitivity which results in distributed values of time across the population. It also recommends value of time bins in auto skimming and assignment. A potential simplification would be to revise the ActivitySim configuration files to remove the distributed value of time calculations and skim expectations. The auto skimming process would generate just one set of skims for each period and occupancy level and ActivitySim would output just one set of trip tables for each period and occupancy level. This would reduce memory requirements and potentially runtime, but would also reduce the sensitivity of the model system to pricing.

Turn off vehicle type model. The current design includes a vehicle type model that predicts body type, fuel type, and age for every vehicle owned by a household. These vehicles are also modeled as choices for auto modes in mode choice; every tour and trip is assigned to either a household vehicle or a non-household vehicle. The auto operating cost from the vehicle is used in mode choice utilities. The model also considers autonomous vehicles. It would be possible to turn this model off in ActivitySim and replace the vehicle-specific auto operating cost with a generic auto operating cost. This would reduce model runtime but would remove the ability of the model to test the impacts of a more fuel-efficient fleet or policy levels (like investment in public charging stations or tax incentives) that could be used to reach efficiency targets. It would also make the model less able to measure greenhouse gas emissions and energy consumption. Additionally, it would make the model insensitive to pricing rebates offered to fuel efficient vehicles such as managed lane discounts, and would make the model insensitive to scenarios involving autonomous vehicles.

Other network simplifications. If the region will never test any pricing alternatives, the fields set up to skim tolls can be dropped from the auto network. Occupancy restrictions could also be removed from the network, and skims could be simplified such that only one set of occupancy skims are built. Utility calculations would need to be modified to utilize the simplified skim representation and output fewer trip tables for assignment. Similarly, if the region will never test a premium transit line, configurations that skim modes such as bus rapid transit, light-rail transit, and commuter rail can be dropped from the configuration. In fact the model system could be set up without any transit representation if necessary, either by setting these utilities to -999 in configuration files or by creating ‘dummy’ skims with zero in-vehicle time. This would likely reduce runtime – both in skimming\assignment as well as in ActivitySim. The memory requirements to run ActivitySim would also be reduced.

6.0 NEXT STEPS

This section of the design document describes the activities required to move forward with the implementation, estimation, and calibration of the proposed design. It covers the following topics:

- the processing of survey data for specific model inputs or parameters,
- the recommended donor model for implementation,
- the process of deploying the donor model,
- enhancements required to implement the proposed design in the donor model,
- limitations of the proposed design and potential future activities to address those limitations.

The section also describes the general level of effort required to estimate the proposed design and tracks decisions that need to be made to finalize either the design or the implemented models based on the comments received on this document to date.

6.1 SURVEY DATA PROCESSING AND AUGMENTATION

Section 5.3 describes data processing to create estimation data files. This section describes creation of some auxiliary variables that may be useful for model estimation and/or survey metrics beyond development of the model.

Bike comfort level: Section 3.9 describes the bike comfort level model which can be used to segment bicycle mode choice utility calculations in path choice and mode choice, and potentially decrease the aggregation bias in the model system. Bike comfort level would be calculated based on responses to a number of questions including stated comfort level with biking on various types of facilities, stated bicycle trip frequency, and revealed data on number of bicycle trips taken.

Average walk and bike speeds: It may be possible to calculate average walk and bike speeds from route trace data in rMove, for different urban environments and traveler characteristics such as age. This data could be used to improve the accuracy of non-motorized calculations in the model, though consistently modeling walk and bike speeds that vary by age may complicate network procedures and the data produced by those procedures.

Vehicle attributes: rMove and rMove for the web record the make and model of each vehicle, as well as the fuel type. If this data is to be used to estimate and/or calibrate the vehicle type choice and vehicle availability models, the body type and fuel type of each household vehicle needs to be coded consistently with the model. The vehicle alternatives table was developed for 2017; given that the calibration year is 2023, it would be helpful to update the vehicle table with makes/models through 2023.

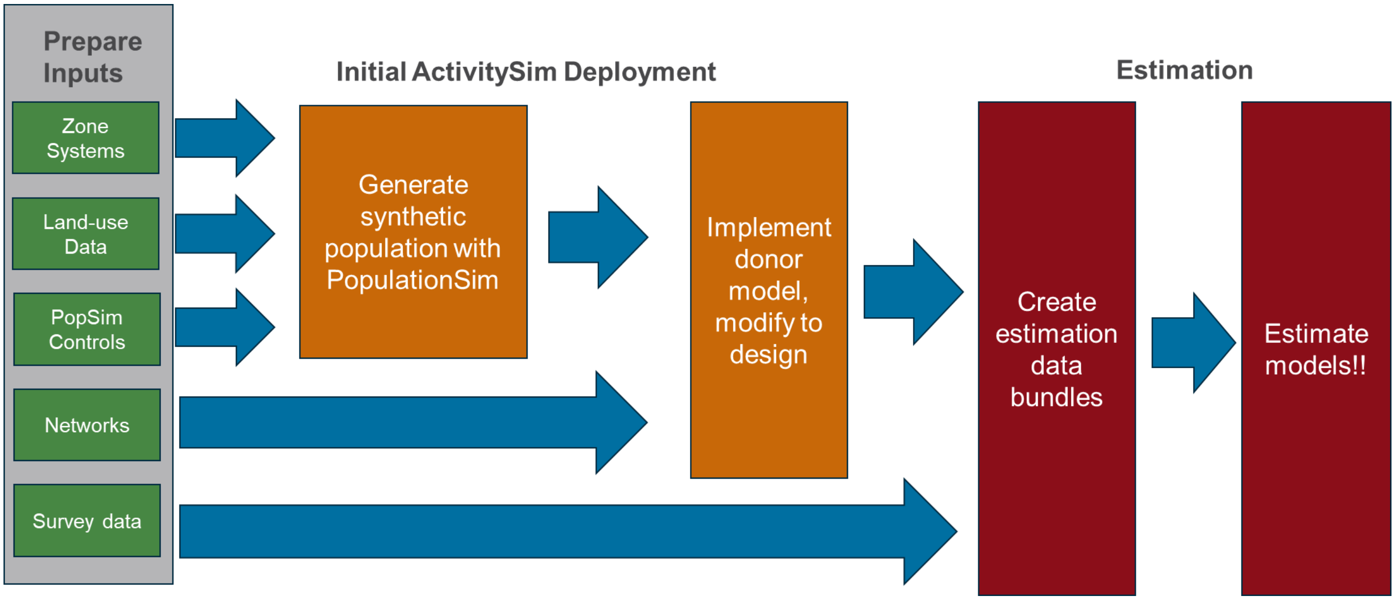
Situational variables: Certain situational variables may be useful to explain monthly and or daily variations in travel. These include the following:

- Average gas prices for the day, week, or month that travel occurs
- Weather including temperature and other conditions (rain, snow, ice)
- Indication if school was in session on survey day

6.2 PROPOSED DONOR MODEL, DEPLOYMENT, AND REQUIRED ENHANCEMENTS

The model development process is shown graphically in Figure 21. The process starts with the preparation of inputs. The zone system including microzones (MAZs) and transportation analysis zones (TAZs) must be defined and the MAZ land-use data file (shown in Table 26) must be developed for each region. Population synthesis controls must be defined and generated for each region. These data must be developed prior to running PopulationSim to generate the synthetic population for each region. The zone systems and land-use data are also inputs to ActivitySim, skimming and assignment procedures, and auxiliary models such as commercial vehicle models. Networks must be developed for each region including planning level networks for auto and transit skimming and assignment and all-streets networks for non-motorized calculations.

FIGURE 21: MODEL DEVELOPMENT PROCESS FLOWCHART



Upon development of networks, skimming and assignment procedures can be finalized and with that, the donor model can be implemented for each region. Once the donor model is up and running, it can be modified to the agreed-upon design. As mentioned previously, we recommend starting from the SANDAG ActivitySim model for Oregon's Joint Estimation effort. The SANDAG model is the closest available model to the model design described above. It includes many of the proposed mobility models, internal-external models, the vehicle type choice model, and the school pickup/dropoff model. PopulationSim will need to be implemented for each region because it is useful to deploy the model before starting to estimate components. Land-use data and skims will need to be developed for each region according to the above specifications.

The following changes will need to be made to the donor model according to the above specification:

- There are several new model components that must be implemented in the model system. These include a driver's license holding status model, a bike comfort level model, and an auto driver identification model. Placeholder versions of these models will need to be implemented in the model system and the choices from the models will need to be introduced into downstream models, particularly mode choice.
- The full vehicle type choice model ("Option 4") in the donor model must be replaced with the partial vehicle type choice model ("Option 2"). The Option 4 model predicts vehicle age, body type, and fuel type simultaneously. The Option 2 model predicts vehicle age and body type; fuel type is selected according to an input distribution of vehicles by fuel type for each combination of body type and age. There is a version of the Option 2 model in the "mtc_extended" model implementation that must be copied over and

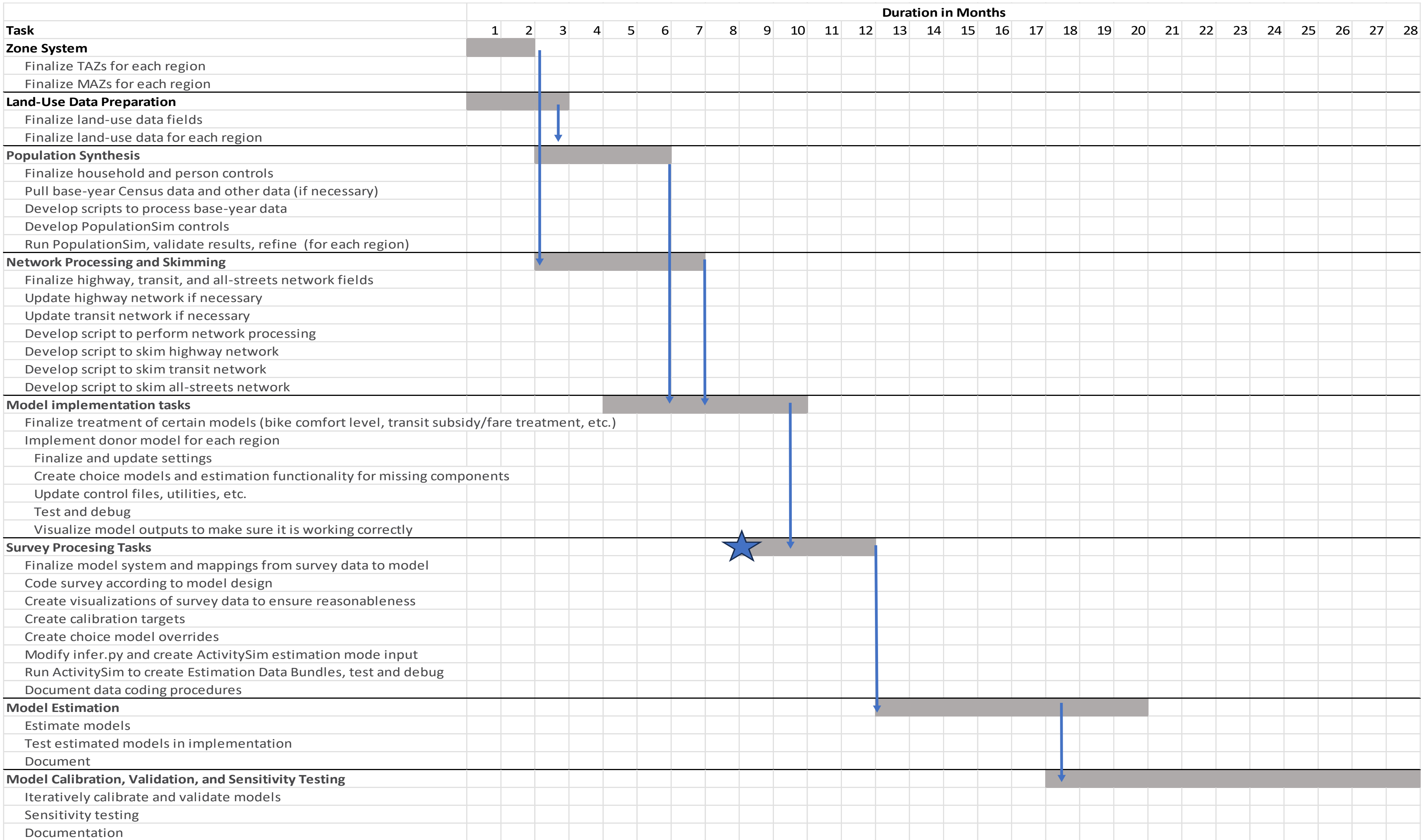
implemented in the donor model. This is probably not a significant effort but must be considered in the transfer scope.

- Person types, purposes, and modes in the model system will need to be made consistent with the design specification.
- The donor model pre-processors and utility terms will need to be revised to be consistent with the fields from population synthesis, land-use data fields and skim tables.
- Outputs will need to be created where the SANDAG model varies from the above specification.

Once the model is modified according to the design, processed survey data can be run through ActivitySim to create estimation data bundles for each region. This is typically an iterative process, as described in Section 5.4. We suggest deployment of a visualizer to summarize model outputs and ensure that the model is functioning correctly prior to estimation. Once data is processed and calibration targets are available (see below), the model results can be compared to targets for each region in the visualizer. After the model is implemented, estimation data bundles can be prepared and models can be estimated.

A detailed list of tasks and an overall schedule for model development is shown in Figure 22. Dependencies between tasks are shown with blue arrows; the delivery of cleaned and processed household travel survey data is denoted by a blue star. The overall development schedule is currently planned for just over two years but could be extended if necessary or desired, or if model development tasks are delayed. Expected availability of household survey data makes condensing the schedule more than currently shown difficult.

FIGURE 22: MODEL DEVELOPMENT TASK LIST AND OVERALL SCHEDULE



The model design must be finalized before the Model Implementation Task can be completed, and it must be determined which models to estimate and which models to adopt from the donor model and calibrate. To aid in the decision-making process, we provide a summary of each model component in Table 50. The table includes information on whether the model component is a “core” model component or not, whether it exists in the SANDAG donor model or would need to be implemented, the source of the data used for estimation of coefficients in the donor model if it does exist, the number of market segments in the model (more segments generally means greater effort for estimation of the component), a general indication of level of effort required for estimation of the model (low, medium, or high), and the order that the component must be estimated in, considering the use of logsums from the model in upper level model choices. Generally, tour and trip mode choice models must be estimated first since their logsums are used in destination choice and time-of-day choice models; next destination and location choice models must be estimated because destination choice logsums are disaggregate accessibilities used in upstream mobility and tour frequency models. All other models can then be estimated in any order. Refer to Figure 6, Figure 7, Figure 8, and Figure 9 for graphics that describe the model system.

Each component is grouped into broad categories describing the type of model -Mobility Models, Day Pattern and Primary Activity\Tour Models, Stop and Trip Level Models - corresponding to Figure 7, Figure 8, and Figure 9 respectively. The row for each model component is colored according to the estimated level of effort for model estimation (green = low level of effort, orange = medium level of effort, red = high level of effort).

The following acronyms are used describe estimation data in the table:

- SANDAG: San Diego Association of Governments
- MAG: Maricopa Association of Governments
- SEMCOG: Southeast Michigan Council of Governments
- HTS/HIS: Household Travel/Interview Survey
- NHTS: National Household Travel Survey
- CMAP: Chicago Metropolitan Agency for Planning
- OBS: On-Board Survey

Each non-core model component is listed again in Table 51 along with the purpose for inclusion of the component in the proposed design and the primary linkages between the model component and other upstream and downstream models.

TABLE 50: MODEL COMPONENTS, ESTIMATION DATA IN DONOR MODEL, AND RELATED INFORMATION FOR ESTIMATION

Model Type	Model component	New model or existing	Core Model?	Data year estimated/region	Segments	LOE for re-estimation	Estimation order	Notes
Mobility Models	License holding status	New	No	NA	1	Low		
	Bike comfort level	New	No	NA	1	Low		
	Autonomous vehicle ownership	Existing	No	NA	1	NA	NA	Asserted model since AVs do not exist
	Auto ownership	Existing	Yes	2016 & 2022 HTS/SANDAG	1	Medium		
	Work from home	Existing	Yes	2016 & 2022 HTS/SANDAG	1	Low		SANDAG models use SANDAG-specific industry codes as explanatory variables
	External worker identification	Existing	No	2016 & 2022 HTS/SANDAG	1	Low		
	External worker location	Existing	No	2016 & 2022 HTS/SANDAG	1	Low		
	Internal work location	Existing	Yes	2016 & 2022 HTS/SANDAG	1	Medium	2 - needed for disaggregate accessibilities	

Model Type	Model component	New model or existing	Core Model?	Data year estimated/region	Segments	LOE for re-estimation	Estimation order	Notes
	School location	Existing	Yes	2004 HTS/SANDAG	3 (K-8, 9-12, college/university)	Medium	2 - needed for disaggregate accessibilities	
	Vehicle type	Existing	No	2017 NHTS	1	Medium		Estimated with Oregon constants; do not recommend re-estimation
	Transit subsidy	Existing	No	2016 & 2022 HTS/SANDAG	1	Low		
	Transit pass ownership	Existing	No	2016 & 2022 HTS/SANDAG	1	Low		
	Free parking eligibility	Existing	No	2004 HTS/SANDAG	1	Low		
	Telecommute frequency	Existing	Yes	2016 & 2022 HTS/SANDAG	1	Low		
Day Pattern and Primary Activity/Tour Models	Coordinated daily activity pattern	Existing	Yes	2016 & 2022 HTS/SANDAG	1	High		Estimation functionality needed for extended model (joint tour extension). Estimated model uses 2004 coefficients for joint tours
	Mandatory tour frequency	Existing	Yes	2004 HTS/SANDAG	1	Low		

Model Type	Model component	New model or existing	Core Model?	Data year estimated/region	Segments	LOE for re-estimation	Estimation order	Notes
	Fully joint tour frequency and composition	Existing	Yes	2004 HTS/SANDAG	1	Medium		Estimation functionality needed for extended model (combined frequency and composition)
	Fully joint tour participation	Existing	Yes	2004 HTS/SANDAG	1	Low		
	Individual non-mandatory tour frequency	Existing	Yes	2004 HTS/SANDAG	7-8 (person type)	Medium		
	Work subtour frequency	Existing	Yes	2004 HTS/SANDAG	1	Low		
	External non-mandatory tour identification	Existing	No	2016 & 2022 HTS/SANDAG	1	Low		
	External non-mandatory location	Existing	No	2016 & 2022 HTS/SANDAG	1	Low		
	Internal non-mandatory tour destination	Existing	Yes	2004 HTS/SANDAG	7 (escort, shop, maintenance, social/visit, eat out, discretionary, atwork subtour)	High	2 - needed for disaggregate accessibilities	
	Tour scheduling	Existing	Yes	2015 HIS/SEMCOG	10 (work, K-12, college/university, escort, shop,	High		

Model Type	Model component	New model or existing	Core Model?	Data year estimated/region	Segments	LOE for re-estimation	Estimation order	Notes
					maintenance, social/visit, eat out, discretionary, atwork subtour)			
	School pickup/dropoff	Existing	No	2008 NHTS Add-on/MAG	1	High		Estimation functionality needed.
	Vehicle availability	Existing	No	2017 NHTS	1	Low		Estimated recently with much bigger dataset; do not recommend re-estimation.
	Tour mode	Existing	Yes	2004 HTS & OBS/SANDAG	6 (Work, K-12, college/university, maintenance, discretionary, atwork)	High	1 - needed for location choice and destination choice	
	Driver/passenger	New	No	NA	2 (fully joint, individual)	Medium		
	Intermediate stop frequency	Existing	Yes	2004 HTS/SANDAG	6 (Work, K-12, college/university, maintenance, discretionary, atwork)	Medium		

Model Type	Model component	New model or existing	Core Model?	Data year estimated/region	Segments	LOE for re-estimation	Estimation order	Notes
Stop and Trip Level Models	Stop purpose	Existing	Yes	2016 HTS/SANDAG	NA	NA	NA	Distribution from observed data
	Stop destination	Existing	Yes	2004 HTS/SANDAG	6 (Work, K-12, college/university, maintenance, discretionary, atwork)	High		
	Stop departure/arrival period	Existing	Yes	2022 HTS/SANDAG, 2015 HIS/SEMCOG, 2019 HTS/CMAP	NA	NA		Distribution from observed data
	Trip mode	Existing	Yes	2004 HTS & OBS/SANDAG	6 (Work, K-12, college/university, maintenance, discretionary, atwork)	High	1 - needed for stop destination	Helpful to estimate tour and trip mode choice together so consistency can be ensured
	Parking location	Existing	No	2004 HTS/SANDAG	1	Low		

TABLE 51: PURPOSE AND PRIMARY LINKAGES FOR NON-CORE MODEL COMPONENTS

Model Type	Model component	Purpose	Primary model linkages
Mobility Models	License holding status	Constraint in driver identification and/or mode choice models, instead of assuming everyone of a certain age can drive/drive alone. Possible explanatory variable in other downstream models.	Mode choice; upstream models via logsums
	Bike comfort level	Segmentation variable for bicycle utilities/logsums, reduce aggregation bias for bicycle choice	Mode choice; upstream models via logsums
	Autonomous vehicle ownership	For modeling future-year scenarios with autonomous vehicles; impacts auto ownership (AV ownership leads to fewer cars) and mode choice (lower parking cost, potentially less onerous in-vehicle time due to productivity), can be considered in assignment (less congestion on freeways due to vehicle platooning, less wait time at intersections due to vehicle-signal technology, etc).	Auto ownership, vehicle type choice, mode choice; upstream models via logsums
	External worker identification	To explicitly represent workers working outside the region (I-5 corridor) and not double-count travel near region boundary.	Work location choice, external worker location, possibly telecommute frequency, activity patterns, tour frequency, downstream models
	External worker location	To explicitly represent workers working outside the region (I-5 corridor) and not double-count travel near region boundary.	Possibly telecommute frequency, activity patterns, tour frequency, downstream models
	Vehicle type	Calculate greenhouse gas emissions and energy consumption; test policies to encourage electric vehicle acquistion such as tax rebates and public charging stations (Option 4), use vehicle-specific auto operating costs in mode choice	Vehicle allocation, mode choice
	Transit subsidy	Reduce transit cost aggregation bias in mode choice; test impacts of increasing transit subsidies.	Transit pass ownership, mode choice, upstream models via logsums

Model Type	Model component	Purpose	Primary model linkages
	Transit pass ownership	Reduce transit cost aggregation bias in mode choice; test impacts of increasing transit pass ownership, reducing transit pass costs.	Mode choice; upstream models via logsums
	Free parking eligibility	Reduce parking cost aggregation bias in mode choice.	Mode choice; upstream models via logsums
	External non-mandatory tour identification	To explicitly represent non-mandatory resident travel to destinations outside the region (I-5 corridor) and not double-count travel near region boundary.	External non-mandatory location choice, internal non-mandatory tour destination
	External non-mandatory location	To explicitly represent non-mandatory resident travel to destinations outside the region (I-5 corridor) and not double-count travel near region boundary.	Internal non-mandatory tour destination
	School pickup/dropoff	Increase accuracy in predicting origin and destination of shared-ride travel and sequence of changes in vehicle occupancy on tours. Precursor to internally consistent vehicle availability and driver/passenger modeling.	All downstream models
	Vehicle availability	Calculate greenhouse gas emissions and energy consumption; test policies to encourage electric vehicle acquistion such as tax rebates and public charging stations (Option 4), use vehicle-specific auto operating costs in mode choice	Mode choice; upstream models via logsums
	Driver/passenger	Simplify conversion of trip lists to trip tables for assignment; potentially increase realism of model.	Trip mode choice
	Parking location	More accurately model parking location of auto trips to parking constrained areas	Trip tables

6.3 LIMITATIONS OF THE PROPOSED DESIGN AND POTENTIAL ENHANCEMENTS

Several limitations of the current general structure and design of ActivitySim have been identified as part of this design project or directly related to the ability of the model system to generate outputs and metrics described as important by the Oregon Modeling Statewide Collaborative in the course of this project. These are outlined below, along with some potential solutions to these problems. We note here that the project sponsors have decided to avoid spending limited project resources on significant software improvements as part of this project. To that end, there is a proposed roadmap for the further development and evolution of the ActivitySim software platform and sample model implementations that may address these shortcomings in the future. We suggest that OMSC and partner agencies carefully monitor the ActivitySim consortium meetings and website for updates on progress towards implementing various aspects of the roadmap. This is described further in Section 6.4.

1. ***Activity schedules include travel time***: The scheduling mechanism in ActivitySim is based on tour departure (from tour anchor location) and arrival (back to tour anchor location) times first; this includes all travel time on the tour and the duration of all activities on the tour. As intermediate stops are generated on the tour, each stop is scheduled; for stops in the outbound direction, the scheduling model predicts departure time; for stops in the inbound direction, the model predicts arrival time. These times are constrained by the tour departure and arrival period - stops must be scheduled in the same or later period than the tour departure and before the tour arrival period and stops are scheduled in sequential order. The key limitation of this approach is that activity start and end times are ambiguous and activity durations may not match observed data when considering travel time requirements. Because hard constraints are not considered in the model, it’s possible that subtracting travel time from total tour duration may even result in negative activity durations.

There are several potential ways to address this shortcoming. One way would be to implement the activity scheduling submodel originally developed for Atlanta Regional Commission’s CT-RAMP model and recently implemented in ActivitySim. The model works by first dividing the tour duration into three segments: the outbound segment, the inbound segment, and the primary activity segment. The outbound segment includes the duration of all intermediate stops in the outbound direction and the travel time to all outbound stops as well as to the primary activity. The inbound segment includes the travel time to all inbound stops as well as time back to the anchor location. The primary activity segment includes just the duration of the primary activity. After stops are generated, the secondary scheduling model runs on the outbound segment and the inbound segment to predict travel time and activity duration episodes, instead of running a separate stop scheduling model.

The model has the advantage that it is already implemented in ActivitySim and that it works to separate travel times from activity durations as well as schedule stops. However, it still works in half-hourly periods, and because trip mode is determined after stops are scheduled, hard travel time constraints cannot be considered in the model. Other options would require more significant changes to the structure of the models implemented in ActivitySim. For example, the order in which activities are generated and scheduled could be modified such that primary activities would be scheduled first by start and end time, then intermediate stops could be generated, located, and scheduled after considering time between activities. This is akin to the way DaySim works, but it requires significant modifications to the overall model structure and corollary changes to the code base. It has implications for nearly every component of the model system – mode choice, scheduling, location choice, activity generation, modeling joint travel episodes, etc.

2. ***ActivitySim does not explicitly model all joint travel.*** There are two aspects of joint travel that are considered explicitly by the model; partially joint travel for the purposes of dropping off and picking up children at school, and fully joint travel episodes in which two or more household members travel together to all out-of-home activities on the tour. Travel for the purposes of picking up or dropping off household members at locations other than school, travel with household members for the purposes of joint activity participation followed by separate travel episodes, and travel with non-household members, are all treated implicitly through ride-sharing in mode choice. This limitation constrains the ability of the model to explicitly allocate vehicles to travel episodes, and to explicitly identify the driver of shared-ride tours. A key aspect of this limitation is that survey data is also limited in the ability to explicitly identify all occupants of shared-ride vehicles. This is particularly true in the case of data collected via smartphone apps, because the data is self-reported and cannot be cross-checked in real time; it is common that person A reports traveling with person B but person B does not report traveling with person A, or can even have a reported activity\travel episode which conflicts with the information provided by person A. To some extent, survey post-processing procedures can attempt to correct for these reporting inconsistencies, but certain intra-household joint travel episodes cannot be resolved via post-processing procedures due to reporting conflicts, and inter-household travel episodes are not explicitly recorded. Even if survey data perfectly recorded all joint travel and activity participation episodes, explicitly modeling all situations would significantly complicate the model system.

3. ***ActivitySim does not allocate vehicles to tours/trips in a fully constrained, explicit manner, or identify drivers on auto trips.*** ActivitySim has a vehicle availability model but does not prevent vehicles from being allocated to multiple overlapping tours. This is due to both the mechanical implementation as well as issues noted above with respect to modeling joint travel. From a mechanical perspective, ActivitySim is currently designed to model mode choice by tour purpose; it is possible that within a single run of mode choice, tour mode could be

simultaneously chosen for multiple overlapping tours within a household. This precludes the explicit allocation of vehicles to tours since conflicting allocations could be made for the same set of activity\travel periods. ActivitySim would have to be restructured to prevent simultaneous vehicle allocations from occurring. Second, because all shared ride is not modeled explicitly, it is not possible to allocate vehicles explicitly to travel parties or to represent who is the driver for each tour. Complications and solutions to modeling all joint travel explicitly are described above.

4. ***ActivitySim outcomes are inconsistent with economic theory.*** ActivitySim uses Monte Carlo simulation from probability distributions as the method of choosing an alternative. This procedure leads to inconsistent results between baseline and build alternatives because the change in utility and probability between the baseline model and the build model for any given decision-maker can lead to a decision-maker switching their choice to an alternative whose utility either stayed the same or got worse in the build case. Consider an example such as mode choice where the utility of bike improves in the baseline due to availability of a bike lane for a work tour. The decision-maker chose shared 3+ in the baseline. Their random number draw was 0.68 and the shared 3+ probability range in the baseline was 0.5 to 0.7. In the build alternative, the shared 3+ probability range changes to 0.4 to 0.6 because of the increase in the bike probability. Their random number in the build stay the same so they choose walk whose probability range is 0.6 to 0.7 in the build. Note that walk utility did not improve but decision-makers switched from shared 3+ to walk because of the order of the alternatives and the change in utility. Such math makes the changes in travel behavior of specific decision-makers between baseline and build alternatives inconsistent with economic theory; the aggregate outcomes of each alternative are reasonable but the changes in decisions are not necessarily reasonable. There are methods to address this issue which involve sampling the error term for each decision-maker and alternative, and choosing the alternative with the highest utility, rather than sampling a random number and choosing an alternative according to the probability distribution. This is referred to as 'inverse cumulative distribution function sampling'. An example of this method which is applicable to multinomial logit models was implemented in a test version of ActivitySim[52] but was not very performant and was not pulled into the current release. Further analysis and testing of the approach, performance improvements, and extension to nested logit models, will be required in order to incorporate it into future ActivitySim releases.

5. ***ActivitySim lacks a formal data model***. A data model specifies the variable type (integer, float, double, etc.), the allowed ranges of each variable (including enumerations of categorical variables instead of strings) and the relationships between variables for each variable used in the software system. Formal data models eliminate ambiguity about the variables used in the model system and reduce bugs, especially when transferring models between regions and

[52] For more information, see Zill and Veitch, Frozen randomness at the individual utility level, Australasian Transport Research Forum 2022 Proceedings 28-30 September, Adelaide, Australia.

when modifying the model system. A formal data model would also likely reduce ActivitySim's memory footprint because ActivitySim currently uses string variables in a number of places which consume significant memory. A formal data model can also be used as a self-documenting system when coupled with tools that automatically generate documentation from software code. However, formal data models take time to implement and can have performance implications. For example, Pandera was recently tested by the ActivitySim consortium as a data model for the ActivitySim input checker and found to be much slower than Pydantic which is a data validator rather than a data model. Due to the slow performance of Pandera, the consortium decided to forgo a formal data model and use Pydantic. We expect the consortium and other ActivitySim agencies to continue to explore the incorporation of these tools in the ActivitySim software and suggest the OMSC continue to monitor this space.

6.4 COORDINATION WITH ACTIVITYSIM CONSORTIUM

Adoption of the ActivitySim software for implementation in Oregon offers many advantages in terms of leveraging software investments made by the ActivitySim consortium. It also requires coordination in order to appropriately plan for adoption of those enhancements and outreach to the consortium in the case that enhancements made by this project can and should be pulled into the ActivitySim code base so that other members can benefit from Oregon's investment. This requires following closely ActivitySim's scope of work for upcoming improvements[53]. The joint estimation effort will benefit from some of the enhancements proposed for Phase 9 (2024), including:

- Updating ActivitySim with some of the models and features developed for the donor San Diego Association of Governments model
- Various memory and performance enhancements
- Estimation mode enhancements to improve performance, reduce runtime, and additional features (currently not funded). There are also a number of issue requests related to estimation bugs that should be addressed prior to proceeding with estimation of affected model components.

We suggest not only following the work as it progresses, but also being proactive in sharing feature improvements, bug identification\fixes, and proposed enhancements with the consortium early and often. Github Issues should be used to communicate feature requests and bugs[54],

[53] https://github.com/ActivitySim/activitysim/wiki/Current-Development
[54] https://github.com/ActivitySim/activitysim/issues

whilst we suggest getting on the agenda for one of the weekly consortium meeting calls to discuss proposed new features or models under development by OMSC.

6.5 SAMPLE LIST OF MODEL OUTPUTS AND METRICS

Section 4.7 lists model outputs; these can be summarized in various ways to support metrics needed for analysis of model results to inform decision making. Table 52 provides a sample list of metrics and the calculations from model outputs required to create them. We recommend automated procedures that would run after each model run to produce the required summaries from the run; an additional procedure would be run to compare the results to a baseline alternative.

TABLE 52: METRICS FROM TRAVEL MODEL AND CALCULATIONS REQUIRED TO SUPPORT THEM

Broad Category	Metric	Calculation
Traveler Metrics	Accessibility metrics (destination choice logsum) by purpose	Disaggregate accessibility calculator can be run for relevant purposes and market segments. Logsums would be multiplied by relevant persons/households in actual population and converted to expenditures in monetary or time units. They can be compared to baseline alternatives to measure benefits.
	Total personal travel (like PMT) and level of time spent traveling (PHT)	Total time spent traveling can be summed from output trip files and summarized for various market segments
	Total vehicle travel (VMT, VHT)	Can be summarized from highway assignment
	Road reliability	Current design does not include measures of auto or transit reliability. Simple methods such as difference between congested and free flow travel time can be quantified from assignment results by period as a proxy for auto reliability, but this is not really measuring reliability per se
	Cost of travel, both direct (fares, operating costs) and indirect (vehicle ownership, etc.)	Travel expenditures can be summarized from trip file
	Mode share	Mode share can be summarized from trip file
	Transit boardings	Transit boardings can be summarized from assignment results
Health, Safety, and Environment Metrics	Accessibility in terms of the safety or stress of active mode options	Destination choice logsums could be calculated for only active modes to measure changes in accessibility for active transport.
	Physical activity (hours spent being active, hours per	Total hours spent walking and biking can be summarized from trip file. Note that purely recreational walking and biking trips are not represented

Broad Category	Metric	Calculation
	capita, hours by traveler)	
	Air Quality (GHGs and criteria pollutants)	Air quality can be summarized from AQ software run on model outputs (such as MOVES). Greenhouse gas emissions and energy consumption can be summarized from trip file and vehicle file.
Economic and Freight Metrics	Total number of jobs accessible within selected travel times (by mode) weighted by origin TAZ.	The measure is an aggregate origin-zone-population-weighted sum of total jobs accessible with defined travel sheds. The travel sheds are set individually by mode (e.g. transit 45 minutes, auto 30 minutes, bike and walk 20 minutes). This is based on an MTC implementation, see details in Chapter 5 of: https://www.planbayarea.org/sites/default/files/documents/Plan_Bay_Area_2050_Performance_Report_October_2021.pdf.
	Destination choice logsum from job locations to workers' residences.	Aggregate accessibility calculations includes an origin-based accessibility counting access to households from each origin zone. This could be modified to calculate access to workers from each origin zone and multiplied by number of jobs in origin zone to measure accessibility of workers to jobs.
	Truck travel time, miles traveled, and delay (truck VHT, VMT, and VHD)	Summarized from highway assignment. Taken together these measure the amount of truck travel, the time trucks spent traveling, and the delay trucks encountered.

7.0 CONCLUDING REMARKS

This document describes a new activity-based travel modeling system for Oregon's MPOs. The new model system, referred to as SimOR, will provide a range of capabilities that exceed the current state of modeling practice in the state. This new paradigm of modeling provides more robust methods to analyze key policies and needs in today's rapidly evolving world including active transportation, micromobility, priced infrastructure, transit, energy consumption, and greenhouse gas emissions. The models are sensitive to a wide range of socio-economic variables and are capable of analyzing the impacts of an aging population on transportation demand. The model system provides more detailed data for performing equity analysis and impacts of transportation investments on communities of concern. The software system that the new models are implemented in is flexible and under constant development by a consortium of agencies. Oregon will benefit from the investments made by the consortium as the software performance and functionality continues to improve into the future. In short, the models represent a significant advancement in the state of modeling practice that should serve Oregon well in the coming years.

This model design document was created over a roughly one-year period beginning in Spring 2023 and ending in Spring 2024. This period coincides with the collection of statewide household travel survey data, which is expected to be delivered in November 2024. The model design project included a series of virtual meetings with the OMSC to cover each section of the document. Each section was covered in two, two-hour sessions; a draft of the chapter was made available prior to the initial session and a complete version of the chapter was made available prior to the final session that incorporates comments received by OMSC members. In addition, two contingency sessions were held; one session focused on network procedures such as creation of bike logsums from an all-streets network, and the other session focused on a few key aspects of the design, such as treatment of value-of-time, that required a bit more explanation and discussion. Recordings were made of each session and meeting notes and slides were made available as deliverables. Finally, bi-weekly meetings were held throughout the project to communicate on work-in-progress, plan for upcoming meetings, and discuss administrative and technical issues.

We fully anticipate that this document will live on and continue to be modified as OMSC continues to debate various aspects of the model system and as model development proceeds. Some aspects of the design, such as potential segmentation of bicycle riders, could not be finalized until the household survey is made available and data is analyzed to determine how best to define the segments. Other aspects of the design, such as the method used to calculate the walk time to the nearest set of transit stops, may require experimentation. Finally, we recognize that some aspects of the design, such as treatment of special travel markets such as university student travel, and certain model features such as highway and transit reliability, are somewhat aspirational in nature and may be more fully flushed out and implemented in subsequent phases of work. This is particularly the case for aspects of the model design that do not affect model estimation, such as parking capacity constraints.

To that end, we have created a spreadsheet to track the decisions made by the OMSC with respect to the model design. The spreadsheet is available online. We also expect the information listed in the input and output tables to be further refined. These include the following:

- Table 21: Auto Skim Tables
- Table 22: Truck Skim Tables
- Table 23: Transit Skim Tables
- Table 26: MAZ Land-use Data File
- Table 27: Synthetic households (synthetic_hhs.csv) file Fields
- Table 28: Synthetic persons (synthetic_persons.csv) file Fields
- Table 34: ActivitySim Output Household File fields
- Table 35: ActivitySim Output Person File fields
- Table 37: ActivitySim Output Tour File fields
- Table 38: ActivitySim Output Trip File fields

We have also begun an online wiki document[55] that describes the model system, which will eventually become a fully-fledged model users guide. Managing the development of a complex model development project such as this is no small challenge, and made even more complex by the participation of multiple agencies and two consultant teams who will be working on implementing, estimating, calibrating, validating, and documenting the models. Communication among team members will be essential. We recommend planning and budgeting for continued bi-weekly meetings as the next round of work covering model implementation and estimation gets underway. We suggest sufficient time and budget be set aside for multiple rounds of review and model estimation given the breadth of agencies and stakeholder needs, and we recommend documentation of not just the final model design but also estimations and variables that were found to be insignificant or otherwise problematic. Although many of these meetings can be attended by technical staff, we suggest scheduling regular checkpoints (potentially every two months) with the broader group of stakeholders through OMSC to provide updates on progress. We have found the project Teams site to be invaluable for sharing files, version documents, and messaging, and suggest either continuing to maintain the site or transferring

[55] See https://github.com/OrMSC/SimOR/wiki

the information to a different site with similar capabilities. We also suggest that resources be set aside for communicating with the ActivitySim consortium throughout the project and maintaining software consistency with ActivitySim releases.