Explicit Error Terms Summary Report, Phase 10

September 26, 2025

RSG: Joel Freedman (primary author), Will Alexander, David Hensle
WSP: Sijia Wang, David Ory
CS: John Gliebe, Andrew Kay, Xuenan Ni
Outer Loop Consulting: Jan Zill
Driftless Labs: Jeff Newman

This report summarizes work completed under Phase 10 of the ActivitySim consortium support project to implement and test Explicit Error Terms (EET) in ActivitySim as a replacement for Monte Carlo simulation. It builds upon the work presented in the memo “Monte Carlo Simulation versus Explicit Error Terms”, dated May 4, 2025.

The previous memorandum described:

- The motivations for this work,
- How EET can be used to draw alternatives from multinomial and nested logit models,
- How random number seeds are generated and used in ActivitySim for both Monte Carlo (MC) simulation as well as EET,
- Some practical considerations around the random number draws and consistency between baseline and build scenarios,
- An outline for implementation and testing of EETs in ActivitySim
- Some possible ways in which model performance could be improved

This memorandum describes the actual implementation of Explicit Error Terms in the ActivitySim software, runtime performance of the code compared to model runs with MC simulation, the results of scenario testing, and provides suggestions for future work.

Implementation of Explicit Error Terms in ActivitySim

EET was implemented in ActivitySim largely as described in the previous memorandum (See ‘Random Number Seeding in ActivitySim’ in the Appendix). For each choice model, the user has the ability to select whether to use MC simulation or EET simulation. This can be set globally in the settings.yaml file (Figure 1) or in each individual model yaml (Figure 2). If use_explicit_error_terms is set to False, MC simulation will be used. In this case a cumulative probability distribution is created for the alternatives in the choice model, a single uniform random number is drawn, and a selection is made from the cumulative probability distribution according to the random number. If use_explicit_error_terms is set to True, then a random number is used to select an error term for each alternative. The error term is added to the systematic utility to calculate a total utility for each alternative, and the alternative with the highest utility is selected. The method for nested logit is similar to what is described above, except that a choice is made from each nest in the nesting structure until the lowest level alternative is selected.

Figure 1: Control simulation method globally via settings.yaml

```
use_explicit_error_terms: bool = False
"""
Make choice from random utility model by drawing from distribution of unobserved
part of utility and taking the maximum of total utility.

Defaults to standard Monte Carlo method, i.e., calculating probabilities and then
drawing a single uniform random number to draw from cumulative probabily.
```

Figure 2: Control simulation method for an individual model in model-specific yaml

```
# Make this more general compute settings and use for explicit error term overrides
# Default None work for sub-components defined in getter below (eet_subcomponent)
use_explicit_error_terms: None | bool | dict[str, bool] = None
```

After the initial implementation of EET in ActivitySim, we noticed an issue where destination choice models would crash because of the availability conditions that are set for alternatives in ActivitySim. The software currently treats all alternatives with a total systematic utility smaller than -999 as unavailable. This was set in CT-RAMP in order to prevent overflow errors when exponentiating the utility, which is a predecessor step to calculating the cumulative probability distribution in MC simulation. The same threshold was carried over to ActivitySim in order to maintain equivalent results. The initial implementation of EET did not include this constraint. The problem manifests in a crash for certain tours where an intermediate stop location is being selected, and there is no viable sampled destination for the tour. This is rare but it can happen - for example, in cases where the tour mode is walk, the maximum distance from a sampled stop location to either the tour anchor location or the primary destination is set to the maximum walk distance (e.g. ~3 miles) and there is no zone with a viable size term within that distance. In such cases, the MC simulation code would cause the software to resimulate the trip stop purpose and destination in the hopes of finding an acceptable solution. In the case of EET simulation, the code crashed because it did not recognize the sample of alternatives as infeasible. We fixed this problem by using the same infeasibility cutoff of -999 to make alternatives unavailable in EET. As future research, we suggest further investigating the problem of infeasible stop destinations with the hopes of a more elegant solution than resampling. However, this is no longer specifically a problem with EET implementation.

Runtime Performance

As discussed in further detail below, we tested MC simulation and EET for two different scenarios using the San Diego Association of Governments (SANDAG) ActivitySim model. We compared the runtimes for each simulation method, as shown in Table 1. In summary, model runtime with EET is 2.4 times the runtime of MC simulation in the SANDAG model for a 100% sample (the times shown in the table are for the transit scenario). The MC simulation runtime is 372 minutes versus 881minutes for EET simulation. The vast majority of the additional runtime is in destination choice models, which are responsible for 495.8 minutes or 96% of the increase. In the MC run, these models (highlighted in yellow in the table) account for 26% of total runtime. In the EET run, these models account for 74% of total runtime. There are slight increases in runtime for some of the other models but these are dwarfed by the increase in runtime associated with destination choice models. We know from further investigation that the majority of this runtime is spent sampling alternatives, likely because of the sheer number of alternatives - there are over 22k microzones in the SANDAG model. Drawing an explicit error term for each of the alternatives is computationally expensive. The optimization of EET code should prioritize improving performance for destination choice.

Table 1: Monte Carlo Simulation versus Explicit Error Term Simulation Runtime Comparison

model_name	**MC (min)**	**EET (min)**	**Diff (min)**	**%Diff**
mp_setup_skims	17	16.6	-0.4	-2%
initialize_proto_population	0.4	0.4	0	0%
mp_disaggregate_accessibility_apportion	0.4	0.4	0	0%
compute_disaggregate_accessibility	6.1	72.6	66.5	1090%
mp_disaggregate_accessibility_coalesce	0.4	0.7	0.3	75%
initialize_landuse	0.3	0.4	0.1	33%
initialize_households	2.1	2.2	0.1	5%
mp_accessibility_apportion	1.6	1.7	0.1	6%
compute_accessibility	2.8	3.1	0.3	11%
mp_accessibility_coalesce	0.5	0.5	0	0%
mp_households_apportion	0.9	0.9	0	0%
av_ownership	0.9	0.7	-0.2	-22%
auto_ownership_simulate	0.1	0.1	0	0%
work_from_home	0.1	0.1	0	0%
external_worker_identification	1.1	1.3	0.2	18%
external_workplace_location	1.1	1.1	0	0%

school_location	15.9	99	83.1	523%
workplace_location	18.4	77.6	59.2	322%
transit_pass_subsidy	0.1	0.1	0	0%
transit_pass_ownership	0.1	0.1	0	0%
vehicle_type_choice	5.6	5	-0.6	-11%
adjust_auto_operating_cost	0	0	0	#DIV/0!
transponder_ownership	0.1	0.1	0	0%
free_parking	0.1	0.1	0	0%
telecommute_frequency	0.1	0.1	0	0%
cdap_simulate	1.1	1.1	0	0%
mandatory_tour_frequency	0.1	0.1	0	0%
mandatory_tour_scheduling	17.9	16.8	-1.1	-6%
school_escorting	3.8	3.5	-0.3	-8%
joint_tour_frequency_composition	0.3	0.2	-0.1	-33%
external_joint_tour_identification	0.6	0.6	0	0%
joint_tour_participation	0.2	0.1	-0.1	-50%
joint_tour_destination	2.5	8.3	5.8	232%
external_joint_tour_destination	1	0.8	-0.2	-20%
joint_tour_scheduling	1	0.7	-0.3	-30%
non_mandatory_tour_frequency	1.8	1.4	-0.4	-22%
external_non_mandatory_identification	1.3	1.4	0.1	8%
non_mandatory_tour_destination	20	110.5	90.5	453%
external_non_mandatory_destination	1.2	1.2	0	0%
non_mandatory_tour_scheduling	61.6	60	-1.6	-3%
vehicle_allocation	1	1	0	0%
tour_mode_choice_simulate	3	5.4	2.4	80%
atwork_subtour_frequency	0.1	0.1	0	0%
atwork_subtour_destination	2.7	13.9	11.2	415%
atwork_subtour_scheduling	0.5	0.4	-0.1	-20%
atwork_subtour_mode_choice	0.3	0.4	0.1	33%
stop_frequency	0.8	0.8	0	0%
trip_purpose	0.2	0.2	0	0%
trip_destination	98.8	284.1	185.3	188%
trip_purpose_and_destination	3.2	3.1	-0.1	-3%
trip_scheduling	0.3	0.3	0	0%
trip_mode_choice	6.7	13.7	7	104%
parking_location	0.7	0.7	0	0%
mp_households_coalesce	5.8	6.1	0.3	5%
write_data_dictionary	4.9	5	0.1	2%
track_skim_usage	0.3	0.3	0	0%
write_trip_matrices	45.3	46.4	1.1	2%
write_tables	6.7	8	1.3	19%
total	371.9	881.5	509.6	137%

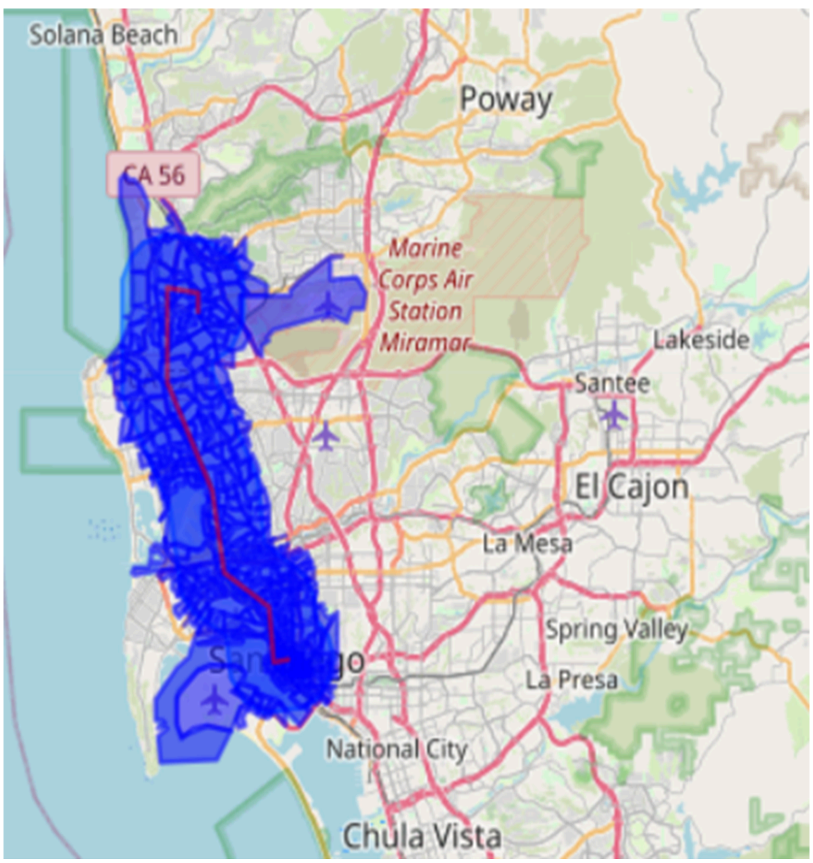
Scenario Tests and Outputs

As part of this task order, we compare and contrast ActivitySim outputs obtained via MC simulation versus EET simulation for two scenarios - a transit scenario and a land-use scenario. We test these results using the SANDAG ActivitySim model. This is a two-zone model[1]. It has been calibrated to observed household travel survey data for the region. The model is documented in an online wiki, https://sandag.github.io/ABM/. The scenarios we test are described below.

Transit Scenario

In this scenario, we created a five-mile wide buffer (2.5 miles per side) around the Metropolitan Transit System (MTS) Blue Line from downtown San Diego to the northern terminus of the line at University of California - San Diego (UCSD). We decrease transit in-vehicle time and first wait time in the A.M. and midday periods by 50% for all zone pairs, access\egress mode combinations, and transit path types, within that buffer (shown in Figure 3). Any tour or trip with *both* origin and destination in the buffer are defined as being in the 'affected area' for summaries of model results, as described in further detail below.

Figure 3: Transit Scenario Affected Area

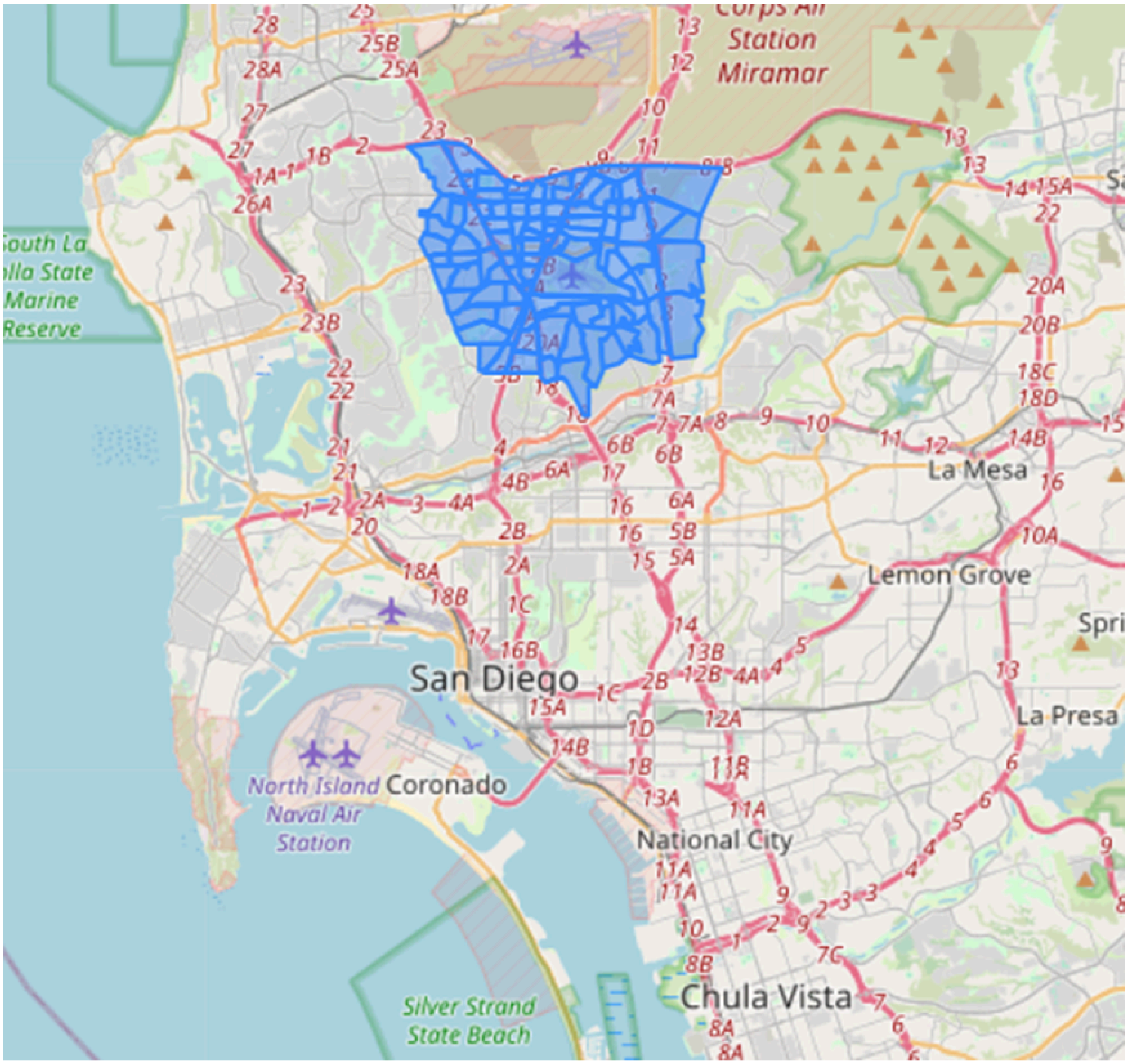


[1] A two-zone model uses Transportation Analysis Zones (TAZs) for auto and transit times and costs, and Micro-Analysis Zones (MAZs) as the spatial unit for origins and destinations and for calculating non-motorized times, typically across an all-streets network.

Employment Scenario

In this scenario, we create a five-mile buffer around an employment center in Kearney Mesa and double employment within the buffer. Any tour or trip with *either* end in the buffer are defined as being in the affected area for this scenario.

Figure 4: Employment Scenario



We ran the SANDAG model for the following six alternatives:

1. Base run with MC simulation
2. Base run with EET simulation
3. Transit scenario with MC simulation
4. Transit scenario with EET simulation
5. Employment scenario with MC simulation
6. Employment scenario with EET simulation

We compared model outputs for base versus build scenarios using MC simulation versus EET simulation. Model outcomes were compared for disaggregate outputs by cross-tabulating choosers by their base versus build choice. Aggregate model outcomes were analyzed by summarizing output trip tables by mode and time period. There are five time periods used for skimming and assignment in the model, as shown in Table 2.

Table 2: Time periods in SANDAG Model

Period	Name	Start Time	End Time
EA	Early Morning	3:00 AM	5:59 AM
AM	AM Peak	6:00 AM	8:59 AM
MD	Mid-Day	9:00 AM	3:29 PM
PM	PM Peak	3:30 PM	6:59 PM
EV	Evening	7:00 PM	2:59 AM

As noted, we break out summaries into ‘affected’ versus ‘unaffected’ area in order to evaluate the extent to which direct systematic changes in utility affect model outcomes versus indirect changes to utility, as might occur when a tour mode changes due to an improvement in transit accessibility, which causes a change to a trip on the tour whose destination is not in the affected area. Or, when the change in transit times leads to the retiming of a tour, which affects the timing, destination, or mode of another tour that is not in the affected area.

With respect to disaggregate summaries, because the same synthetic population is used in all scenarios, and household locations do not vary between the scenarios, we define any household or person whose residence zone is in the affected area as ‘affected’. However, for tours and trips, the definition of which choosers are in the affected area or not depends upon the scenario tested, as follows. For the transit scenario:

- Base tours with BOTH origin and primary destination in the zone list are tagged as "affected“
- Base trips with BOTH origin and destination in the zone list are tagged as "affected“
- Build tours that are not matched (via chooser ID, as explained in the Appendix) with Base tours (new tours) with BOTH origin and primary destination in the zone list are tagged as "affected“
- Build trips that are not matched with Base trips with BOTH origin and destination in the zone list are tagged as "affected"

For the employment scenario:

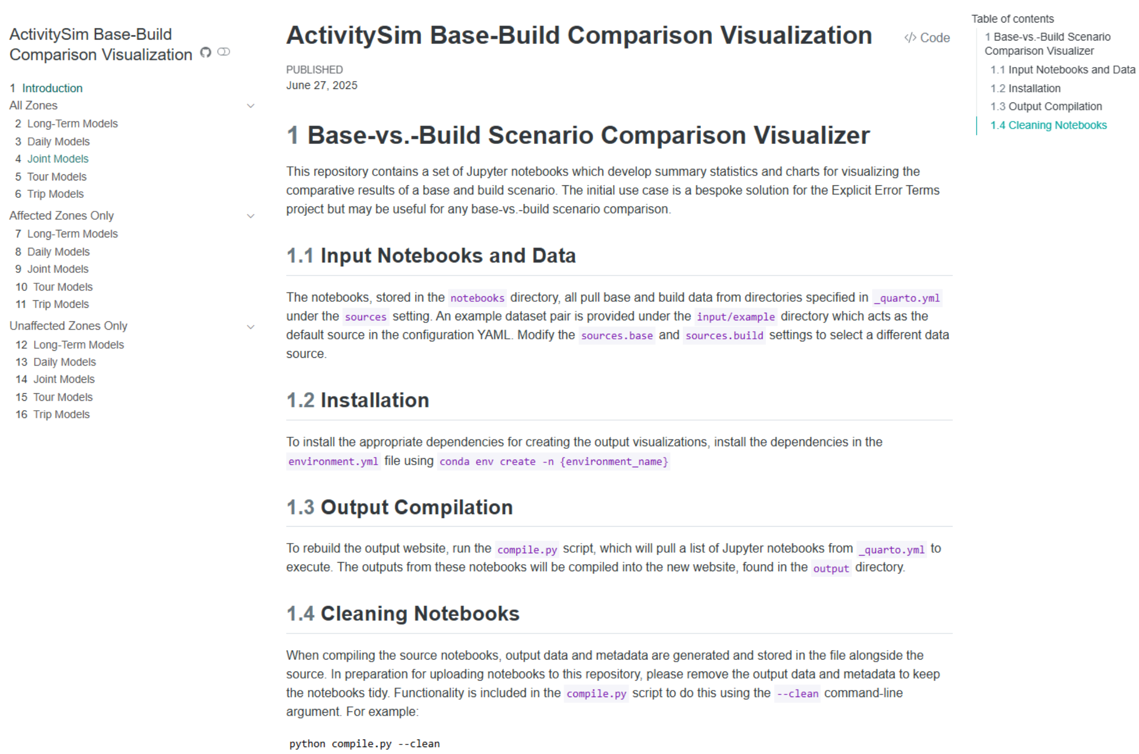
- Base tours with EITHER origin or primary destination in the zone list are tagged as "affected“
- Base trips with EITHER origin or destination in the zone list are tagged as "affected“
- Build tours that are not matched with Base tours (new tours) with EITHER origin or primary destination in the zone list are tagged as "affected“
- Build trips that are not matched with Base trips with EITHER origin or destination in the zone list are tagged as "affected"

We use EITHER origin or destination to identify tours and trips in the affected area for the employment scenario because we would expect significantly more change to destinations as a result of a change in the zonal inputs in the employment scenario as compared to the transit scenario in which the change in systematic utility is mostly restricted to tours and trips with both origin and destination in the affected area.

Summaries were implemented in Python but different technologies are used for disaggregate versus aggregate summaries. For disaggregate summaries, we developed a visualization tool using Quarto notebooks. We output HTMLs from this tool which can be viewed in a web browser (tested HTMLs in Firefox and Chrome). The HTML files are organized into summary pages for all outputs, outputs in the affected area, and outputs in the unaffected area. There are separate pages for long-term choices, daily models, joint models, tour models, and trip models for each of the areas. One visualizer is created for each of the four scenarios (MC transit, EET transit, MC employment, EET employment). The landing page for the visualizer, which can be opened by clicking on the index.html file in the directory, gives instructions on how to configure settings and build visuals for a given base versus build scenario (Figure 5).

Each page contains cross tabulations of base versus build choice for different models. The cross tabulations are created by merging the base and build chooser by chooser ID, consistent with the way in which random numbers are generated and held constant between model runs. Note that there are cases where the number of tours or trips changes between the base versus build run. In such cases, we identify base choosers who could not be matched to build choosers as an additional row in the cross tabulation, and build choosers who could not be matched to base choosers as an additional column in the cross tabulation.

Figure 5: Disaggregate Output Visualizer Landing Page



ActivitySim Base-Build Comparison Visualization

1 Introduction
All Zones
2 Long-Term Models
3 Daily Models
4 Joint Models
5 Tour Models
6 Trip Models
Affected Zones Only
7 Long-Term Models
8 Daily Models
9 Joint Models
10 Tour Models
11 Trip Models
Unaffected Zones Only
12 Long-Term Models
13 Daily Models
14 Joint Models
15 Tour Models
16 Trip Models

ActivitySim Base-Build Comparison Visualization

Code

PUBLISHED
June 27, 2025

1 Base-vs.-Build Scenario Comparison Visualizer

This repository contains a set of Jupyter notebooks which develop summary statistics and charts for visualizing the comparative results of a base and build scenario. The initial use case is a bespoke solution for the Explicit Error Terms project but may be useful for any base-vs.-build scenario comparison.

1.1 Input Notebooks and Data

The notebooks, stored in the `notebooks` directory, all pull base and build data from directories specified in `_quarto.yml` under the `sources` setting. An example dataset pair is provided under the `input/example` directory which acts as the default source in the configuration YAML. Modify the `sources.base` and `sources.build` settings to select a different data source.

1.2 Installation

To install the appropriate dependencies for creating the output visualizations, install the dependencies in the `environment.yml` file using `conda env create -n {environment_name}`

1.3 Output Compilation

To rebuild the output website, run the `compile.py` script, which will pull a list of Jupyter notebooks from `_quarto.yml` to execute. The outputs from these notebooks will be compiled into the new website, found in the `output` directory.

1.4 Cleaning Notebooks

When compiling the source notebooks, output data and metadata are generated and stored in the file alongside the source. In preparation for uploading notebooks to this repository, please remove the output data and metadata to keep the notebooks tidy. Functionality is included in the `compile.py` script to do this using the `--clean` command-line argument. For example:

```
python compile.py --clean
```


The aggregate summaries are contained in an excel file, which can be created by running the Python script. The following summaries comparing trip tables between Base vs. Employment by MC/EET and Base vs. Transit by MC/EET were created:

- Difference in total trips by mode and time period
    - Modes include all auto, SOV, SR2/3, Transit, Walk and Bike
    - Both absolute difference and percentage difference were provided
- Max difference across cells for each mode/period, RMSE and percent RMSE

- Additional statistics discounting small values to minimize the impact of zero or near-zero values on %RMSE calculations
    - Number of cells in the matrix that change from less than 1 trip (0 to 1) in the Base to at least 10 trips in the build scenario
    - Number of cells in the matrix that change from at least 10 trips in the Base to less than 1 trip (0 to 1) in the build scenario
    - The count, RMSE and %RMSE for cells where either the Base or the build scenario has a minimum of 10 trips
    - The count, RMSE and %RMSE for cells where either the Base or the build scenario has a minimum of 50 trips
- Segment summaries by zone pairs/periods that are affected vs. not affected. Affected zones are defined as:
    - Employment scenario: a trip whose origin or destination lies in an employment-scenario zone. Transit scenario: a trip whose origin and destination are both in transit-affected zones.
    - Zones affected can be found in the input folder in the GitHub ActivitySim/Asim_eet_viz repository

Scenario Results

For each scenario, we summarize each model, noting first what our expectations are for the model based on our understand of what changed in the build scenario compared to the base scenario, and our knowledge of the variables and coefficients in the model. Note that in some cases, we may not have a clear expectation of the change. We then describe what we observe in the visualizer and aggregate summary spreadsheet described above. We do not replicate the tables provided in these tools in the discussion below due to time and budget constraints; readers are encouraged to follow along with the discussion by browsing the results directly in the aforementioned tools.

Transit Scenario Results

Long-Term Models: Auto ownership

Expected results: We expect households to reduce auto ownership in the affected area as a result of improved AM and MD transit utilities. We do not expect households to increase auto ownership, and we do not expect auto ownership to change outside of the affected area.
Findings: We see more households reducing auto ownership with MC simulation (74) than with EET simulation (48). We see roughly the same number of households

increasing auto ownership (9 in EET vs. 8 in MC). These are very small percentages of total households in the affected area (172,540). The change in auto ownership with MC simulation is always +/- 1 car, whereas in EET there are changes between nonadjacent car choices. The use of MC simulation with fixed random numbers implies an order which in the case of auto ownership may be desirable but restricts the outcomes to switch to adjacent alternatives. Neither EET nor MC had any changes in auto ownership for households outside the affected area

Long-Term Models: Mandatory Location Choice

Expected results: We expect some workers\students to change work\school location to zones within the affected area as a result of improved utilities. We expect minimal changes to work\school location for persons outside of the affected area - they should only change to the simulation-based constraint mechanism used to ensure that the number of workers or students who choose each zone is proportional to the employment or enrollment in each zone.

Findings: In the affected area, 0.6% of work locations change in EET, vs. 3% in MC In the unaffected area, no work locations change in either EET or MC. In the affected area, 3.2% of school locations change in EET, vs. 3.1% in MC. In the unaffected area, 6.5% of school locations change in EET, vs 7.6% in MC. Overall the changes in mandatory location choice suggest more stability in the EET model with respect to changes in the affected area for work, but relatively similar changes for school location choices across both MC and EET for school. There are more workers who live in the affected area selecting to work in the affected area due to the change in transit accessibility, but the change is more pronounced for MC simulation. There is more switching in MC simulation because small changes in utility can lead to big changes in outcomes, especially for models with a lot of alternatives. So, these findings are expected,

Note that the constraint mechanism may not change the chosen workplace or school location if the changes are less than the threshold. I think it makes some sense that school is changing more, because enrollment is “lumpier” than employment – in other words, it is more concentrated in a limited number of zones. So, changes in utility can have a bigger effect on any one zone, making it more likely that the constraint mechanism needs to compensate.

Long-Term Models: Transit Subsidy

Expected results: We expect an increase in workers\students to have subsidized transit due to the improved household and employment transit accessibilities in the affected area. We do not expect changes to transit subsidies outside the affected area.

Findings: Transit subsidies increase for 114 persons in the affected area for MC and 48 persons in EET. Transit subsidies decrease for 73 persons in MC and only 8 persons in EET. There are slight increases in transit subsidies in both MC (3 persons) and EET (5 persons) in the unaffected area.

Long-Term Models: Transit Pass Ownership

Expected results: We expect an increase in persons to own transit passes due to the improved household transit accessibilities in the affected area. We do not expect changes to transit pass ownership outside the affected area.

Findings: Transit pass ownership in the affected area increases for 208 persons and decreases for 110 persons in MC, versus 98 and 17 persons respectively for EET. The ratio of transit pass changes in EET is more logical than for MC. There is no change to transit pass ownership in the unaffected area for MC and only 1 person increases their transit pass ownership in the unaffected area for EET.

Daily Models: Coordinated Daily Activity Patterns

Expected results: We expect more active patterns for persons in the affected area due to improvements in accessibilities. We do not expect changes in activity patterns for persons outside the affected area.

Findings: 130 persons switched from a home pattern to either a mandatory or non-mandatory pattern in MC versus only 9 persons switching in EET. We observe 241 persons becoming less active (switching from mandatory or non-mandatory to home) in MC versus only 12 persons becoming less active in EET. Essentially there is more simulation noise in CDAP in MC and very little elasticity to accessibility in CDAP in EET. There is no change in CDAP outside the affected area for either MC or EET.

Daily Models: Non-Mandatory Tour Frequency

Expected results: We expect more non-mandatory tours to be generated in the affected area due to improvements to accessibility. We expect minimal changes to non-mandatory tour frequency in the unaffected area; such changes would only be due to indirect effects of changes in mandatory tour locations due to simulation-based constraints.

Findings: There are 923 persons with more non-mandatory tours in the affected area in and 923 persons with fewer non-mandatory tours in the affected area in MC. There are smaller changes in EET 230 persons with more non-mandatory tours and 323 persons with fewer non-mandatory tours. Overall the changes are more muted in EET but persons with fewer non-mandatory tours are unexpected and should be evaluated further. We also observe changes in non-mandatory tours outside the affected area in both MC and EET, though the EET simulation has roughly 50% less change than MC. Again, it would be helpful to trace utility calculations for some choosers to better understand why the choice outcomes are changing outside the affected area.

Tour Models: Non-Mandatory Tour Destination Choice

Expected results: We expect non-mandatory tour destinations for tours in the affected area to change more than tours outside the affected area. We also expect tour lengths to decrease, but tour length is not currently summarized in the visualizer. This would be a useful future addition to the tool.

Findings: 2899 non-mandatory tours (1.7%) change their primary destination in the affected area in MC versus 1411 non-mandatory tours (0.8%) in EET. Changes in non-mandatory tour destination in the unaffected area are roughly the same in MC and EET: 35817 tours (1.5%) change their primary destination in MC versus 36210 tours (1.5%) in EET. This is a surprising result given that the other models show smaller changes in the unaffected area for EET and should be further investigated.

Tour Models: Mandatory Tour Scheduling

Expected results: We expect mandatory tours to shift their departure time to A.M. peak and Midday skim and assignment periods due to the improved accessibility in these periods compared to the base alternative in the affected area. There may be departure shifts to early AM as well to take advantage of the better transit time in the AM and especially MD periods for the return tour leg. There should not be departure time shifts to later periods. We expect arrival periods to shift to AM or MD. There may be shifts to arrival time periods as well, given that tours that left in the EA Period in the base and shifted to AM or MD in the build may stay longer at work, and tours that depart earlier in the build may leave work earlier. We expect minimal shifts to time of day choice for mandatory tours outside the affected area.

Findings: 69 mandatory tours shifted their departure time to the AM period, while 24 less tours chose to depart in the MD period in MC. In EET, 35 tours shifted their departure time to the AM period, and 18 fewer tours chose to depart in the MD period. The AM shifts are expected; the midday shifts suggest that improving accessibility in the AM period caused a shift from the MD period even though its accessibility also improved. The tabulation provides evidence of this; the greatest changes from the MD period are shifts to the AM period. Departures in the EA period also increased in both MC and EET, with slightly more tours shifting to the EA period in EET (9) than in MC (2); these shifts are also expected to take advantage of returning in the AM or MD periods.

Mandatory tour departure time shifts in the unaffected area are about 2.5x higher for MC simulation than for EET, and arrival time shifts are about 8x higher in MC simulation than for EET, though as a percentage of total mandatory tours in the unaffected area the shifts are quite small in either case.

Tour Models: Non-Mandatory Tour Scheduling

Expected results: We expect similar results as mandatory tours.

Findings: There are similar findings as described above for mandatory tours though changes are close between MC and EET runs. In the affected area, 25 non-mandatory tours shift their departure period to the AM period and 18 fewer tours depart in the MD period versus 37 and 19 less tours respectively for EET. Note that EET has no mandatory or non-mandatory tours shifting to EA, PM, or EV periods, while MC has 14 more non-mandatory departures in the PM period, which is unexpected.

Changes in departure and arrival time for non-mandatory tours in the unaffected area are generally smaller for EET than MC. For example, there are 331 less tours departing in the MD period in MC compared to 12 more tours in EET. Note that there are over twice as many non-mandatory tours in MC (roughly 48k) that cannot be matched between the base and build alternative as there are in EET (roughly 19k). Both are less than 1% of total non-mandatory tours in the non-affected area though.

Tour Models: Mode Choice

Expected results: We expect shifts to transit from other modes for tours in the affected area. We do not expect shifts to other modes, and we expect minimal tour mode choice changes outside the affected area.

Findings: 478 mandatory tours and 751 non-mandatory tours shifted from non-transit to transit in the affected area in MC, while 435 mandatory tours and 726 non-mandatory tours shifted to transit in the affected area in EET - very similar shifts. However in MC simulation, more tours shifted from adjacent modes than in EET. 672 total tours shifted from non-motorized to transit in MC, compared to 480 tours in EET. Most of the difference is accounted for in shifts from auto to transit.

Shifts to other modes in the affected area were much lower in EET compared to MC; 491 tours shifted to a mode other than transit in MC, versus only 16 tours in MC.

In the unaffected area, tour mode shifts were more pronounced for MC than EET. 632 tours shifted to transit and 776 tours shifted to another mode in MC. In EET, 348 tours shifted to transit and 252 tours shifted to something else.

Tour mode shifts for at-work subtours were fairly modest for both MC and EET in the region and are not included in the description of tour mode changes above.

Tour Models: Stop Frequency

Expected results: We have no a priori expectations regarding stop frequency models for tours in the affected area. On the one hand, we expect increases in stop frequency due to increased accessibilities; on the other hand, we might expect decreased stop frequency in the affected area due to the shift to transit. In the unaffected area, we expect minimal change in stop frequency.

Findings: Slightly more tours increased the number of stops (5004) as decreased stops (4386) in the affected area in MC. Changes in stop frequency were much smaller in EET; 1969 tours increased stops while 1373 tours decreased stops. Changes in stop frequency in the unaffected area was twice as high in MC as it was in EET; 76430 tours changed stop frequency in the unaffected area in MC versus 38749 tours in EET.

Trip Models: Trip Destination Choice

Expected results: We expect stop destinations for tours in the affected area to change more than stop destinations for tours outside the affected area. We also expect trip lengths to decrease, but trip length is not currently summarized in the visualizer. This would be a useful future addition to the tool.

Findings: Over twice as many (5221) trips changed their destination in MC in the affected area versus EET (2367). More trips also changed their destination in the unaffected area in MC (36899) versus EET (29727).

Trip Models: Trip Mode Choice

Expected results: We expect trips to switch to transit and walking in the affected area as a result of increased transit accessibilities and tour mode shifts to transit in the affected area. We expect minimal trip mode choice changes outside the affected area.

Findings: 4657 trips switched to transit in the affected area in MC versus 3176 trips in EET. Of these, 2351 (50%) switched from non-motorized modes in MC versus 1478 (47%) trips in EET. 1535 trips switched to a non-transit mode in MC versus only 155 trips in EET. The transit mode share increased by 14% in both scenarios. Note that the mode changes in both scenarios are less than 1%.

In the unaffected area, 2674 trips switched to transit and 2228 trips switched to a non-transit mode in MC, versus 1118 and 762 trips respectively in EET.

Aggregate Results

The base versus transit summary shows similar results between MC and EET in terms of total trips by mode and time period in total, with slightly less transit trips in EET. The 'Zones affected transit' page breaks these summaries out for affected zones; we also see similar aggregate results between MC and EET. However, we notice an increase in shared ride trips in the early AM period in MC; this illogical result is not shown for EET. Overall, transit trips increase by 29.7% in the AM peak period and 24.7% in the midday period in the affected area in MC. Similar results are shown for EET; 27.8% and 25.1% respectively. Note that the increase in total transit trips does not track exactly with the above discussion of disaggregate results. This is due to two differences between the analysis:

1) The disaggregate summaries above count joint trips as one record per joint trip, while the aggregate summaries count person trips.

2) Because the disaggregate summaries tabulate base versus build results for each chooser, a consistent method must be used to define how to classify whether each chooser is in the affected area or not across both scenarios. This is not the case in the aggregate summaries, where a person trip could be counted as affected in one scenario and not affected in another scenario depending on the trip origin and destination TAZ.

Subsequent summaries of the disaggregate data (not shown) indicate that the treatment of joint travel is the most significant cause of the differences.

The base versus build root mean square errors are similar between MC and EET trip tables, but when normalized to the mean, the percent root mean square error for EET is lower than for MC.

Employment Scenario

The primary expectation for the employment scenario is that the increase in employment will cause more tours to change their primary destination to the affected area, and that more trip origins and destinations will shift to the affected area. We expect fewer changes to tours further from the affected area, though we have not implemented such summaries in the results. Therefore unlike the summary of the transit scenario above, we keep the disaggregate summary fairly broad, but add spatial analysis of the effects of the scenario on destination choice outcomes at the tour level.

Long Term Models: Auto Ownership

There is much greater change in auto ownership in MC than EET. Across the entire region, over 10k households changed their auto ownership in MC versus only slightly over 2k households in EET.

Long-Term Models: Mandatory Location Choice

Across the entire region, in MC simulation, 85.2% of workers changed their workplace and 22.9% of students changed their school location, compared to 56.9% and 15.1% respectively in EET. This demonstrates much more variation in destination choice using MC compared to EET for land-use scenarios.

Long-Term Models: Transit Subsidy and Pass Ownership

In MC simulation, 27685 persons changed their transit subsidy choice and 40990 persons changed their transit pass ownership choice, while in EET 18329 persons changed their transit subsidy choice and 25818 persons changed their transit pass ownership choice. In both cases, the shares of persons who change their alternative is less than 1%, but EET is changing less than MC in both cases.

Daily Models: Coordinated Daily Activity Patterns and Tour Frequency

In MC simulation, 162840 persons changed their daily activity pattern, while in EET, only 14371 persons changed their daily activity pattern. The change in daily activity patterns contributes to significant changes in tour frequency in MC compared to EET. 64711 workers change their mandatory tour frequency in MC compared to only 17,092 workers changing their mandatory tour frequency in EET. Nearly 323k persons change their non-mandatory tour frequency in MC (almost 10%) versus slightly more than 50k persons (less than 2%) in EET.

Tour Models: Non-Mandatory Tour Destinations

In MC simulation 37.1% of non-mandatory tours change their primary destination, compared to 19.5% in EET. There are also 12% of non-mandatory tours in the base scenario that could not be matched with a build scenario tour in MC, compared 1.8% base tours that could not be matched to build tours in EET. Very similar shares are shown for build tours that could not be matched to base tours in MC and EET.

Tour Models: Tour Scheduling

Nearly 400k mandatory tours (26%) changed their departure period and over 670k (45%) mandatory tours changed their arrival period in MC, versus 85k (6%) and 82k (6%) mandatory tours respectively in EET. Over 614k (21%) non-mandatory tours changed their departure period and over 956k (32%) tours changed their arrival period in MC versus 160k (5%) and 157k (5%) tours respectively in EET.

Tour Models: Mode Choice

In MC simulation, 448k tours (9%) changed their tour mode compared to 201k (4%) tours in EET.

Tour Models: Intermediate Stop Frequency

Nearly 540k tours (10%) changed their stop frequency in MC versus 114k tours (2%) in EET.

Trip Models: Trip Destination Choice

65.1% of trips changed their destination in MC versus 48% of trips in EET. However, there were many more trips that could not be matched (no common chooser ID) between base and build alternatives in MC simulation (1.6M, or 12% of total) than in

EET simulation (296k, or 2%) due to changes in stop frequency. Thus there is significantly more variation in trip destination choice in MC than in EET, and significantly more changes to overall tour structure in MC than in EET for this scenario test.

Trip Models: Trip Mode Choice

Approximately 180k (a bit more than 1%) trips changed their trip mode in MC versus 73k (less than 1%) of trips in EET.

Spatial Analysis

Figure 6 and Figure 7 show how tour destinations change in MC simulation and in EET simulation respectively. Though the maps are similar, it does appear that there are more increases in tour destinations outside the affected area in MC simulation than in EET. Figure 8 and Figure 9 show the number of tours by tour origin TAZ whose destination changed in MC and EET respectively. These plots demonstrate that far fewer tours change their destination in EET than in MC, and those tours that do change their primary destination tend to be more spatially concentrated around the affected area.

Aggregate Analysis

The base vs. employment scenario comparisons show that total auto trips increase in both MC and EET by similar percentages. Transit, however, diverges: EET shows slight decreases across most time periods except that EA shows a slight increase, whereas MC's transit trips decrease in PM and EA but increase in AM, MD and EV. Within the employment-affected zones, trip grows for motorized modes in both EET and MC by similar percentages, yet SR2/3 in EA surges far more in EET than in MC. In terms of transit trips, EET generally carries more transit trips than MC, except in the AM.

Table 3 shows Root Mean Square Error (RMSE) and percent difference in RMSE for aggregate trip tables by time period and mode for MC versus EET. As shown in the table, there are significant reductions in RMSE for all modes and time periods with EET as compared to MC.

Table 3: Root Mean Square Error for Aggregate Trip Tables by Time Period and Mode, Monte Carlo Simulation versus Simulation with Explicit Error Terms

TIME PERIOD	AUTO			TRANSIT			WALK\BIKE		
	MC	EET	PCT DIFF	MC	EET	PCT DIFF	MC	EET	PCT DIFF
EA	0.10	0.08	-22.1%	0.02	0.01	-50.0%	0.04	0.02	-50.0%
AM	0.33	0.26	-21.0%	0.05	0.03	-40.0%	0.15	0.11	-26.7%
MD	0.88	0.82	-7.2%	0.07	0.05	-28.6%	0.56	0.52	-7.1%
PM	0.42	0.35	-17.9%	0.06	0.04	-33.3%	0.20	0.14	-30.0%
EV	0.28	0.21	-23.6%	0.04	0.03	-25.0%	0.15	0.09	-40.0%

Figure 6: Change in Tours by Tour Destination TAZ, Monte Carlo Simulation

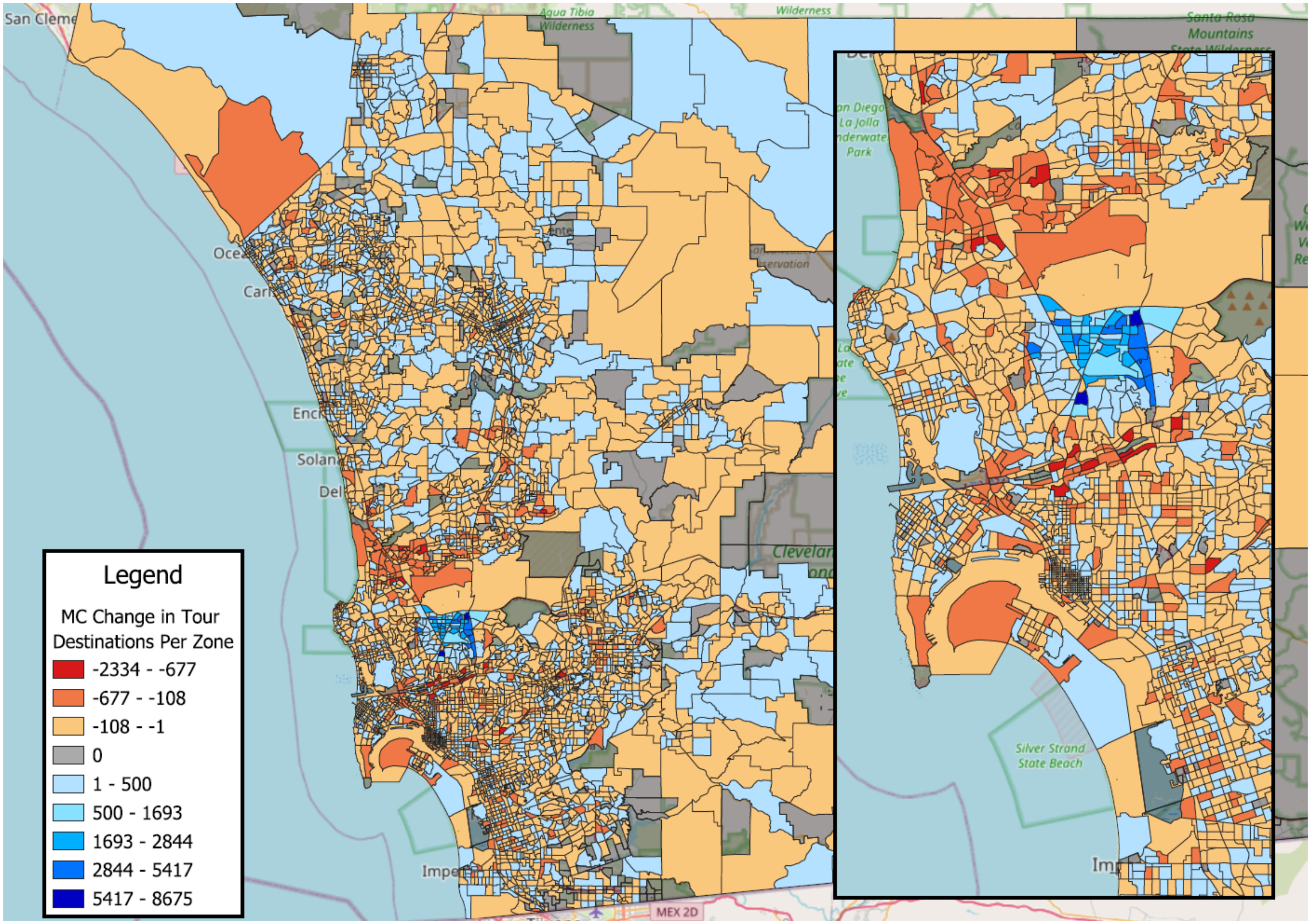


Figure 7: Change in Tours by Tour Destination TAZ, Simulation with Explicit Error Terms

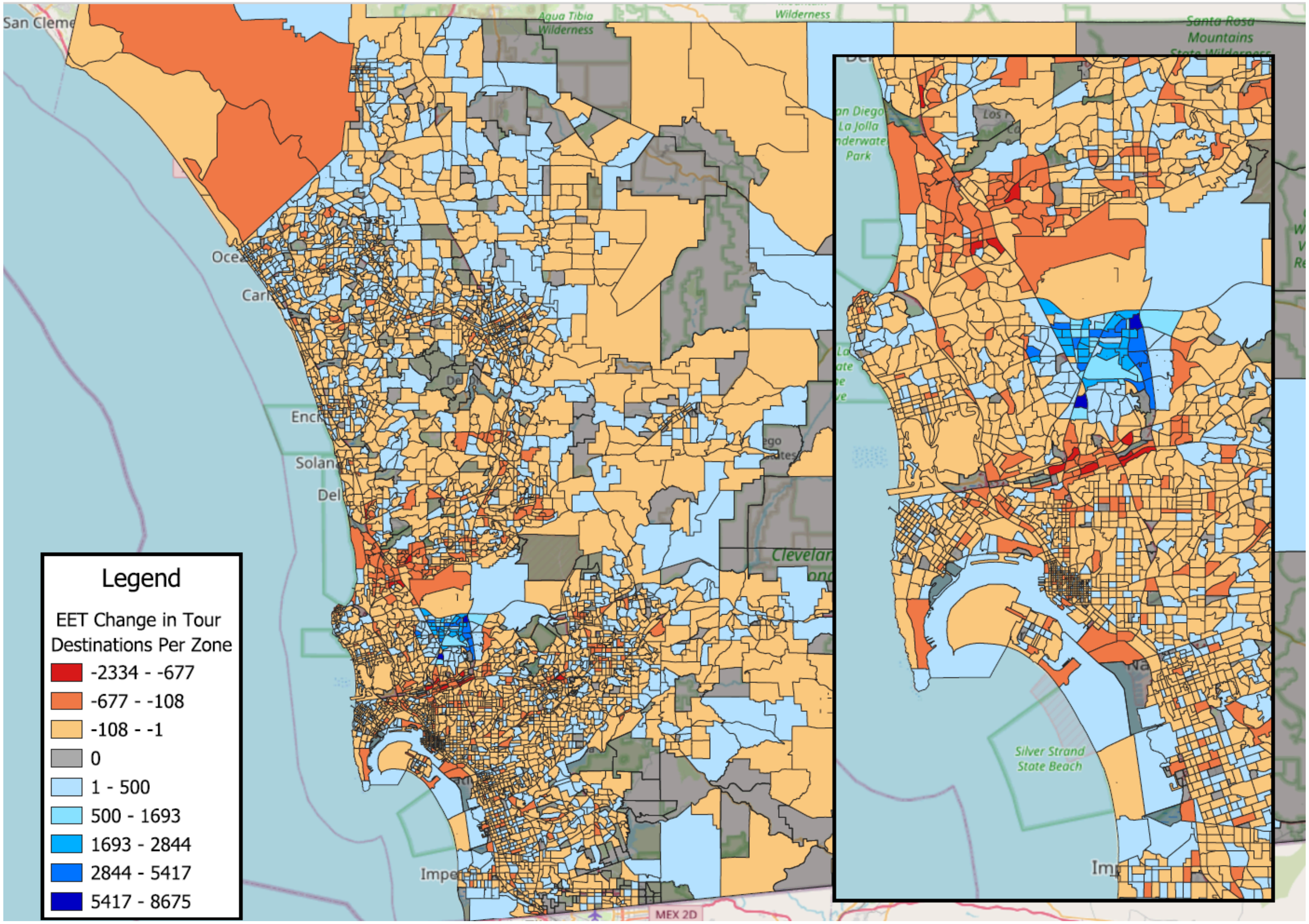


Figure 7: Change in Tours that Changed Tour Destination by Tour Origin TAZ, Monte Carlo Simulation

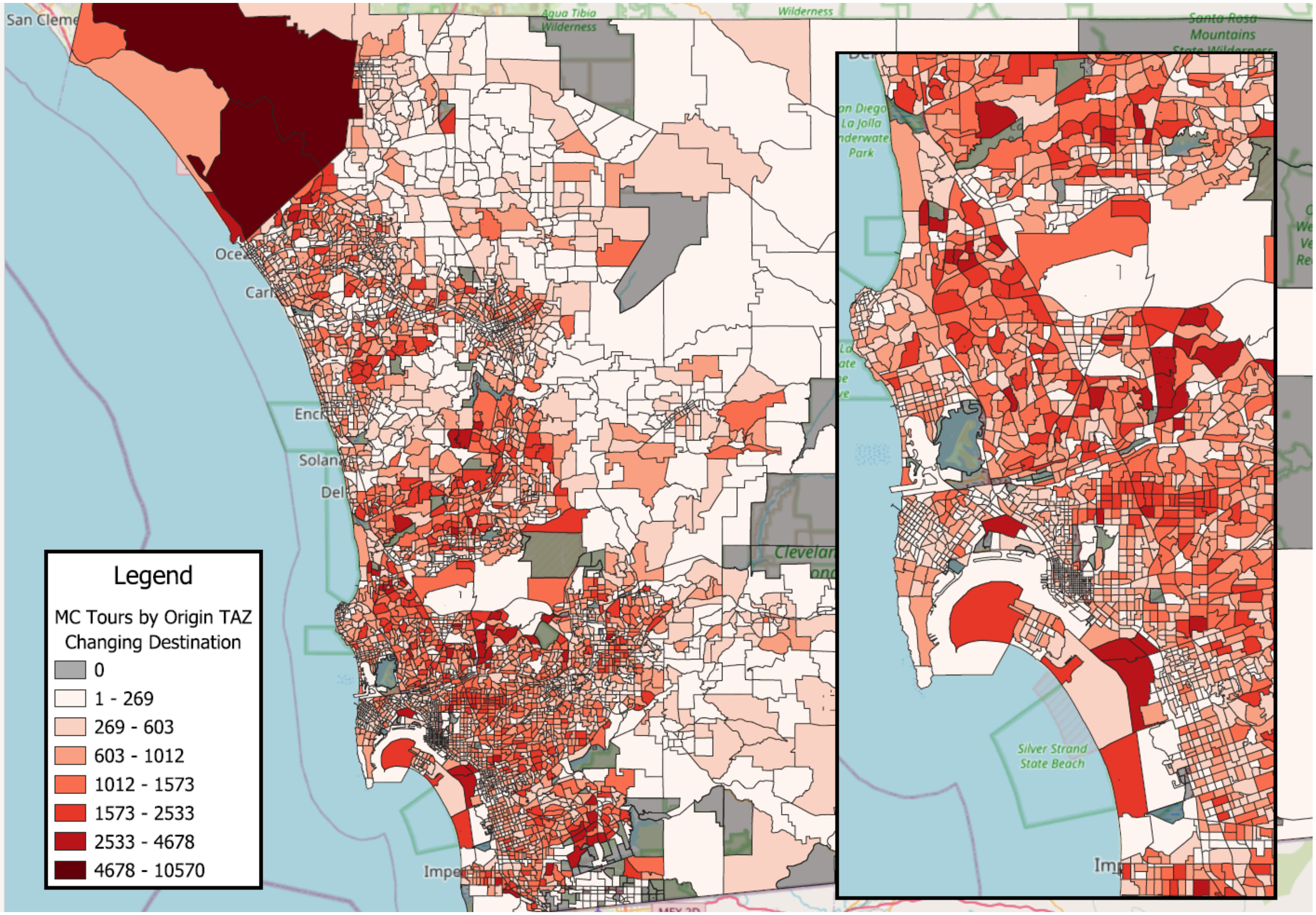
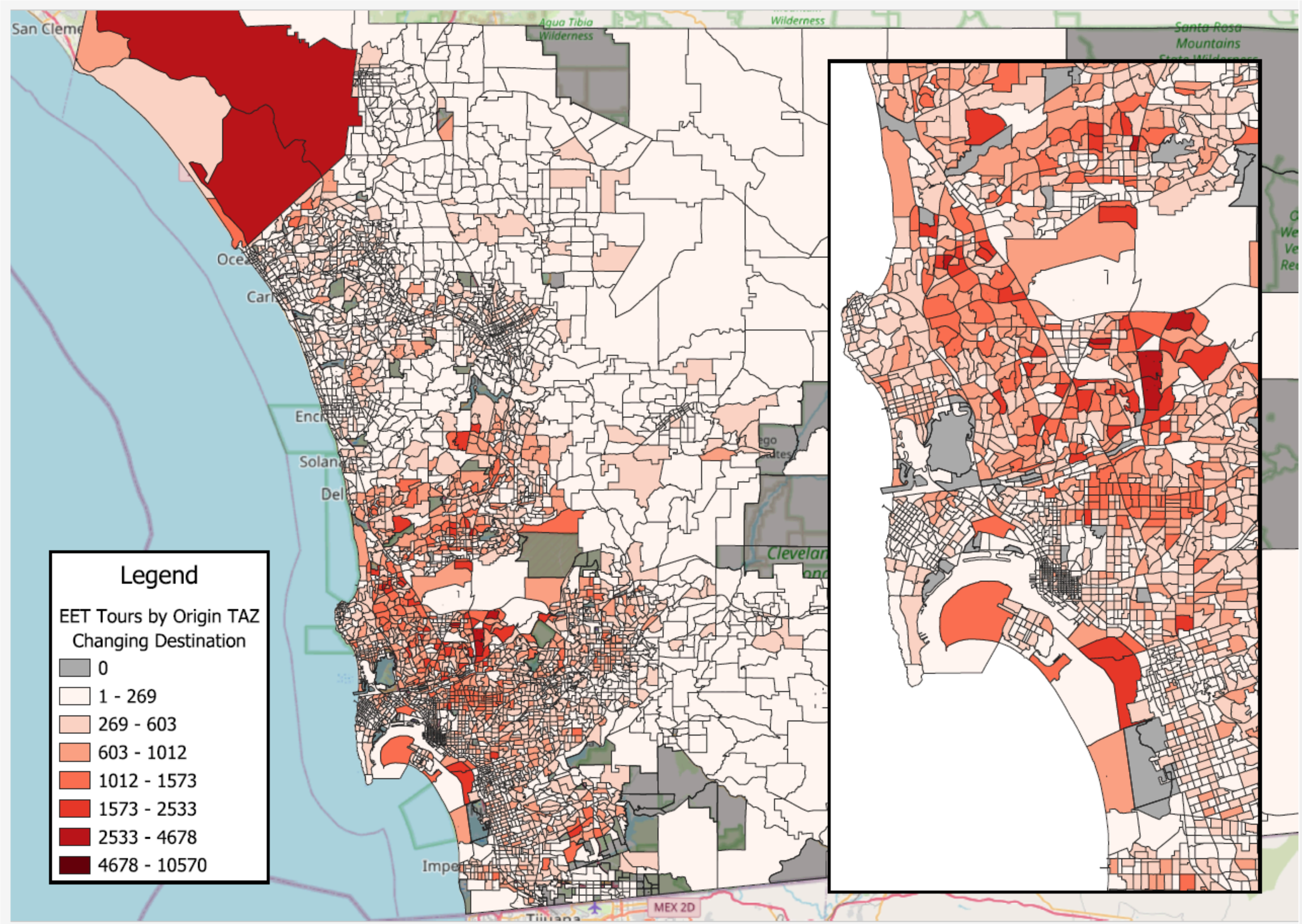


Figure 8: Change in Tours that Changed Tour Destination by Tour Origin TAZ, Simulation with Explicit Error Terms


Legend
EET Tours by Origin TAZ Changing Destination
0
1 - 269
269 - 603
603 - 1012
1012 - 1573
1573 - 2533
2533 - 4678
4678 - 10570
Agua Tibia Wilderness
Santa Rosa Mountains
MEX 2D

Sameness Assessment

The motivation for the EET approach is that model owners want changes in choices across scenarios to be a function of the systematic utility component, rather than the random utility component. To operationalize EET, we must therefore identify a choice in the build scenario that is the same as a choice in the base scenario (i.e., we cannot use the same error term across two choices unless we identify which two choices are the same). In practice, analysts would rarely, if ever, compare individual choices. But they may want to aggregate over a small number of choices or they may want to know how they should think of choices across two scenarios as the same or not (i.e., should they be compared). It is therefore important to know and think through the best way to join choices across scenarios. For the balance of this section, we use the word "sameness" to define how choices across scenarios are paired up.

To illustrate the challenges and opportunities for various sameness definitions, this analysis uses a single ActivitySim component: tour mode choice. Identical issues exist for any ActivitySim model component that is run more than once for a given household or person. (For model components, such as automobile ownership, that are only run once per simulation, the definition of sameness is clear and unambiguous: there is only one automobile ownership choice per household in the base and build scenarios, so pairing them is trivial.) The analysis uses the Transit Scenario described previously.

The definition of sameness used in the EET prototype code is designed for Monte Carlo simulation. Error terms for Monte Carlo simulation have the following two requirements:

1. Uniqueness across choices that were assumed to be independent in model estimation. For example, if a single person engages in two shopping tours, it is common to treat these two tours as independent observations in model estimation. We should therefore give these two shopping tours independent error terms in the simulation, so that the estimation and application are consistent.
2. Generation of seeds from artifacts of the subject choice, which facilitates reproducibility. It's easy to satisfy requirement (1) Uniqueness, as you can just keep drawing random numbers. But to satisfy this requirement, we craft random number seeds from household_id, person_id, model_id, etc. -- artifacts of the subject choice (i.e., attributes of the subject choice that can be re-used when crafting a random number seed, which allows the same random number seed to be used when re-running a scenario). This facilitates reproducibility for a single scenario. ActivitySim, as documented in the previous memorandum, uses counting indices, e.g., the first shopping tour gets an index of one, the second

shopping tour gets an index of two, so that requirement (1) Uniqueness continues to be satisfied.

With the explicit error term approach, we must satisfy a third requirement when creating random number seeds, which is (3) Sameness. To compare choices across scenarios, we must identify choices or choice contexts that are the same.

In tour mode choice, the Monte Carlo method of creating random seeds creates a unique seed for each tour, specific to each household, person, and tour purpose. If, for example, Person X engages in two shopping tours, the first shopping tour is given an index of 1 and the second shopping tour is given an index of zero. The random seed is then a random integer based on household_id, person_id, purpose, and tour_type_num (this is the index mentioned above, e.g., the first shopping tour gets an index of one). For the prototype EET code, this approach was retained. Depending on an analyst's view of sameness, this approach is potentially problematic. Consider the following example:

Base Scenario:

Household ID: A

Person ID: X

Shopping Tour 1: Departs at 9 am for TAZ 1, returns at 10 am

Shopping Tour 2: Departs at 7 pm for TAZ 18, returns at 8 pm

Build Scenario:

Household ID: A

Person ID: X

Shopping Tour 1: Departs at 7 pm for TAZ 18, returns at 8 pm

In this example, a change in the build scenario motivated the elimination of this person's first shopping tour in the base scenario. Using the prototype EET code, the same error term for shopping tour 1 in the base scenario is used for shopping tour 1 in the build scenario. To most analysts, this is an error: shopping tour 2 in the base scenario is very similar to shopping tour 1 in the build scenario. Said another way, if an analyst were asked, which of the two shopping tours in the base scenario is the same (i.e., should be compared directly to) as the one shopping tour in the build scenario, most analysts would select base shopping tour 2.

Analysis

The bad news is that, for any model component that runs more than once in the simulation, there is no perfect definition of sameness: every definition is subjective and the pros and cons of any definition can be debated/discussed. The good news is that there are alternatives to the method used in the prototype code, which would involve adding more conditions to check for sameness. For example, one could define sameness for tour mode choice as two tours that have the same household_id, person_id, purpose, depart_time, arrive_time, and primary_destination_taz.

The Jupyter notebook tour-mode-choice-sameness-assessment.ipynb included in the asim_eet_viz repository does the following:

1. Reads in the base and build tour mode choice output files (final_tours.csv) for the transit scenario tested as part this assessment.
2. The error terms for the EET method are based on the tour_id column. We therefore know, across the base and build scenarios, which choices are identified by the prototype code as the same.
3. Five distinct definitions of sameness are explored. Each of the five is described below.
4. For each of the five definitions, the performance of the prototype code is assessed.
5. The results are summarized in a large CSV file that is rendered with the sameness-investigations.twb Tableau workbook.
6. The Jupyter notebook also highlights examples that are in line with the two shopping tour/one shopping tour example that motivated this examination, i.e., this is both a theoretical and practical problem.

The five definitions explored in the notebook are as follows:

1. "Basics", which defines sameness as tours with the same household_id, person_id, and purpose.
2. "Basics + tour_type_num", which is the same as (1), but with tour_type_num, which is a purpose-specific unique index for each tour. This is the definition used in the prototype code.
3. "Basics + destination", which is the same as (1), but with the tour destination_taz added.
4. "Basics + start_time", which is the same as (1), but with the tour start_time added.

5. "Basics + start_time + duration + origin + destination", which is the same as (1), but with start_time, duration, origin, and destination added.

Mechanically, the analysis assesses the performance of the prototype code against these definitions as follows:

1. Create a definition.
2. Join the tour mode choice outcomes using this definition for the base and build scenarios.
3. Using the tour_id variable, determine if the joined tours received the same error term in the prototype code.

Once joined, each base scenario tour mode choice decision is given one of four outcomes, as follows:

1. "Success: Correct match to build" - If, per the proposed sameness definition, the base and build joined tours received the same error term using the prototype code.
2. "Success: Nothing comparable in build" - If, for example, Person A engaged in one shopping tour in the base scenario and no shopping tours in the build scenario, there is nothing comparable in the build scenario to compare to the subject tour in the base scenario.
3. "Failure: Incorrect match in build" - If two tours are joined using the subject sameness definition, but these tours have different error terms.
4. "Failure: Error term not generated in build" - If, for example, Person A engaged in two shopping tours in the base scenario and one shopping tour in the build scenario, and the second shopping tour in the base scenario was deemed the same, per the subject sameness definition, as the first shopping tour in the build scenario. The error term we want in this example does not exist in the build scenario.

Key Results

The "Basics + tour_type_num" sameness definition, which is definition number 2, should have performed perfectly, i.e., all the choices should be labeled "Success": it is the definition used in the prototype code. And it nearly does, as shown in the table below: it succeeds 99.9 percent of the time. The failures are due to either (a) a coding error or, more likely, (b) a misunderstanding on the part of the analyst as to the meta data needed to define sameness. The minor differences are not relevant to the key findings of the analysis.

Definition of Sameness	Success: Correct match to build	Success: Nothing comparable in build	Failure: Incorrect match to build	Failure: Error term not generated in build	Total
Basics (hh_id, person_id, purpose)	99.77%	0.09%	0.00%	0.14%	100.00%
Basics + tour_type_num	99.73%	0.16%	0.05%	0.07%	100.00%
Basics + destination	98.13%	0.19%	1.65%	0.04%	100.00%
Basics + start_time	98.92%	0.20%	0.86%	0.02%	100.00%
Basics + start_time + duration + origin + destination	97.62%	0.22%	2.16%	0.01%	100.00%

The other sameness definitions explored in the analysis are oriented towards behavioral definitions of sameness. An attractive definition for tour mode choice is "Basics + start_time + duration + origin + destination". Here, if an analyst assumes this definition of sameness across a base and build scenario, the prototype code succeeds about 97.8 percent of the time in the transit example. It will fail more than 2 percent of the time, which is about 97,000 tours in the SANDAG example.

Practical Implications

If an analyst assumes a sameness definition of "Basics + start_time + duration + origin + destination" and the ActivitySim EET implementation fails about 2 percent of the time, does this matter? The short answer is probably not – it would only matter if the types of comparisons shown on page 26 were important. But because the EET method requires making pairwise comparisons across scenarios, it seems wise to take the short amount of time required to come up with a thoughtful definition of sameness for each model component.

For tour mode choice, seeding the error term generator with a value that considers household_id, person_id, purpose, start_time, duration, origin, and destination is likely a better way forward than using the prototype code's current approach. Said another way: analysts using ActivitySim are more likely to consider tours with the same purpose, start time, duration, origin, and destination to be the same than they would tours that have the same, for each person, sequential index by purpose.

More or less strict: Independence Violations

As noted previously, there is no perfect definition of sameness for any model component run more than once for a household or person in the simulation. All definitions are subjective and the pros/cons of all definitions can be debated. However, less strict definitions are more likely to result in violations of the independence of choices. For example, two shopping tours made by the same person in the same day are very likely to be treated as two, independent observations in the tour mode choice model estimation. The math behind a logit model, which leads to its elegant closed form solution, requires that the choices are independent of each other, i.e., the error terms are not correlated across choices. If the selected sameness method, e.g., the "Basics" shown above, uses the same error term in application for two shopping tours made by the same person, the application will have violated the assumptions used in estimation: it will use the same error term across observations, which leads to correlations across choices that should be treated as independent. This is why the current method used in ActivitySim adds the number of the tour to the random number seed.

To estimate the quantity of potential independence violations, we summarize the number of tour mode choice outcomes that, when viewed through each sameness definition, appear more than once in the base SANDAG dataset. The results are shown in the table below with the frequency of indepedence violations shown in the columns (e.g., because some persons make 9 tours of the same purpose, the independence count goes up to 9 for the "Basics" definition).

	IIA Count										
Definition of Sameness	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**
Basics (hh_id, person_id, purpose)	71.34%	18.69%	4.24%	3.53%	1.36%	0.65%	0.12%	0.05%	0.01%	0.00%	100.00%
Basics + tour_type_num	99.99%	0.01%									100.00%
Basics + destination	86.86%	12.61%	0.31%	0.20%	0.01%	0.01%					100.00%
Basics + start_time	98.61%	1.36%	0.03%	0.00%							100.00%
Basics + start_time + duration + origin + destination	99.98%	0.02%									100.00%

Using a less strict definition of sameness, e.g., "Basics", "Basics + start_time", results in more potential independence violations than the more strict definitions. (As before, the "Basics + tour_type_num" definition should perform perfectly on this measure and fails, likely because joint tours and individual tours were not considered as different for the purposes of calculation of this measure, though the chooser ID in ActivitySim does vary by joint versus individual tours of the same purpose.

More or less strict: Response variability

Theoretically, independence violations are unattractive. But practically, many model owners may not be bothered by them, in particular because there is an upside to accepting more independence violations, which is a reduction in what we will call model response variability.
To illustrate this point, consider an extreme example: every tour mode choice decision in the simulation gets the same error terms by mode. Meaning, the random utility component for transit is, say, -0.50, for every tour mode choice decision, and the random utility component for drive alone is, say, 0.25, for every tour mode choice decision, and on and on. In this case, if transit level-of-service is improved, the response of the model would be sharp. Meaning, for any two travelers with similar systematic utilities across modes, the elasticities with respect to the transit level-of-service change would be similar. The heterogeneity of the response to the level-of-service change would be diminished.

This extreme example is only presented to demonstrate the point. Using the same error term for every tour mode choice decision is a bad idea, for many reasons, including that it would over- or under-state the model's response to the change, depending on the random number seed.

Practically, however, analysts may prefer less model response variability in exchange for some independence violations. Assessing the practical pros and cons of this would be an interesting and useful study, but is beyond the scope of the present work.

A rough-and-ready way to assess the likely model response variability is by counting the unique error terms in the tour mode choice outcomes. This is the opposite side of the coin of independence violations. These results are shown in the table below. The stricter the definition of sameness, the more unique error terms, meaning the more variability in the component response. (As before, the "Basics + tour_type_num" should be perfect).

Definition of Sameness	**Share of Error Terms that are Unique**
Basics (hh_id, person_id, purpose)	83.34%
Basics + destination	92.97%
Basics + start_time	99.28%

Basics + start_time + duration + origin + destination	99.99%
Basics + tour_type_num	99.99%

Optimal

So what choice is optimal? There is no perfect definition of sameness, and any definition has its pros and cons. We should use a definition of sameness that the ActivitySim consortium feels comfortable with. It is clear that using a definition that does not include tour number (e.g. the “Basics” alternative) is not advised due to violations of independence. Using more strict definitions of sameness than is used in the current code would make comparing base versus build comparisons more concise, but would likely introduce complexity to the code - requiring reading and merging choice outcomes from a base run to a build run in order to determine when to vary random number seeds and when to keep them constant. It may also introduce more simulation variance into the responses; as noted above, some of the model responses to the transit scenario are relatively small; letting error terms vary for more choices may reduce the signal-to-noise ratio for similar runs.

Conclusions and Next Steps

This report conclusively demonstrates the following:

1. Monte Carlo Simulation suffers from problems with ordering alternatives as noted in the task design document which manifest when comparing base to build model outcomes across individual decision-makers[2]. This is demonstrated in the auto ownership and mode choice results from the transit scenario, where changes in adjacent alternatives are more significant in MC than in EET. It is the result of fixing random number seeds by chooser ID, which reduces Monte Carlo simulation variance, but introduces adjacency errors in base versus build outcomes.

2. Monte Carlo Simulation demonstrates more simulation variance than simulation with explicit error terms across a range of disaggregate choice outcomes and with respect to aggregate results. This is very evident in the employment scenario where there are much more significant changes across all choice models due to the increase in employment in Kearney Mesa.

[2] Importantly, this problem with Monte Carlo simulation only occurs if the same random seed is given to the base and build scenarios. It need not be. Using a different seed, however, would result in more simulation variability across the scenarios. Additional analysis would be useful in quantifying this effect for the two example scenarios.

3. The ‘uncomparable choosers’ issue - that is, tours and trips that cannot be matched between base and build alternatives by chooser ID due to a change in the number of tours by purpose or number of stops on each tour - is limited in the transit scenario but much more significant for the employment scenario due to the changes in day pattern, tour frequency, and stop frequency models. The issue is much more significant for MC simulation than for EET.

4. Both MC and EET methods show unexpected variation in results outside the affected area for the transit scenario, though MC shows more variation than EET. Further investigation would be helpful to understand why some of these changes are occuring. The authors posit that one reason may be the way in which the disaggregate accessibilities are calculated, where a sampling mechanism is used to select MAZs to generate residence locations for prototype households. Because the sampling mechanism does not guarantee that all market segments are covered in each TAZ, it is possible that some households may be merged with disaggregate accessibilities in a TAZ other than the one in which they reside. This can be modified and tested fairly easily. The other reason for changes outside the affected area is likely due to the simulation-based constraint mechanism. More carefully controlling for simulation variance in the algorithm may be useful.

Next Steps

The team recommends moving forward with production-ready software that supports explicit error terms, due to the reduction of simulation variance and improved ability to compare base versus build model outcomes compared to Monte Carlo simulation as documented above.

There are several logical next steps indicated by the above findings.

1) More detailed analysis of changes outside the affected area in the transit scenario. This analysis should initially focus on two key model components, possibly uncovering other issues, and potentially leading to software development activities:
    a. Disaggregate accessibilities: Explore how disaggregate accessibilities are affected by the build scenario and how those accessibilities are merged with households.
    b. Simulation-based constraint mechanism: Explore how the constraint mechanism iterates towards a solution and how the random number sequences used in the mechanism may be better controlled to reduce

simulation variance. There are ways to take advantage of explicit error terms in the constraint mechanism that may reduce simulation variance, reduce the number of decision-makers who are made worse off by the introduction of constraints, and potentially even reduce runtime.

2) We suggest several avenues to improve software performance. Note that these features may improve the performance of both MC and EET simulation methods.
    a. The current method used to draw random numbers could be replaced with a more modern method. The current method (Mersenne Twister) is based on a Java implementation that is around 20 years old. Newer methods carry much less overhead.
    b. The method used for sampling of alternatives could be replaced with a simpler method. The current method is referred to as 'importance-based sampling', and it is used to select a sample of alternatives that generally follows the utility distribution of the full choice model. It uses a simple destination choice model, where distance is a substitute for a mode choice logsum, to generate this sample. Then mode choice logsums are calculated for each destination in the sample and the full destination choice model is run for this much smaller set of alternatives. Simple random sampling and stratified random sampling are alternative methods. In simple random sampling, the choice set is drawn randomly from all alternatives (considering only alternatives with a positive size term). This could result in some choosers with a set of alternatives whose utilities are all very small. Stratified random sampling ensures that this is unlikely by using a districting system to control the sample. Either method would be much faster than the existing approach.
    c. The user could use MC simulation for sampling alternatives and EET for the full choice model, in order to reduce the runtime associated with sampling of alternatives for EET. However, this may introduce more simulation variance into the results.

3) Test the explicit error term code against the Monte Carlo simulation code for a real-world economic appraisal, as a measure of effectiveness of the code and the software enhancements recommended above.

Appendix

Current Method of Random Number Seeding in ActivitySim

ActivitySim currently relies on the Mersenne Twister pseudo-random number generator (RNG) for Monte Carlo simulation, developed in 1997 by Makoto Matsumoto and Takuji Nishimura. This RNG is popular because of its long period length as well as other statistical properties. However, its use in ActivitySim is somewhat unique in that most RNGs are used to continually draw random numbers once the algorithm is instantiated with a 'seed' value. In ActivitySim, the RNG is continually re-seeded with a unique value for each model that is a function of the decision-maker for whom the choice is being made. This is done to ensure that random number draws are independent of the order in which households, persons, tours, and trips are processed so that outcomes can be replicated across machines and\or with different parallel processing settings given the same inputs. It is worth noting that the properties of the random numbers generated by the algorithm in ActivitySim have not been evaluated as to their statistical properties such as ensuring that random numbers are uniformly distributed and uncorrelated despite re-seeding repetitively.

The seed values used to generate random number sequences are created by adding one or more IDs associated with the decision-maker to a hash value based on the name of the model being applied, with some exceptions as noted below. The IDs used in the RNG are shown in Table 6 (model hash is not shown). The RNG seeds follow the type of decision-making unit that the model is being applied to:

- Household level decisions are seeded with the household ID, which uniquely identifies each household in the synthetic population. An exception is the vehicle type choice model, which is seeded with the vehicle ID. Vehicle ID is generated by household ID such that it uniquely identifies vehicles owned by the household (changes in number of vehicles owned by one household does not affect the vehicle IDs of other households).
- Person level decisions are seeded with the person ID, which uniquely identifies each person in the synthetic population.
- Tour level decisions are seeded with the tour ID, which uniquely identifies each tour. Tour ID is generated based on person id, tour purpose, and tour number such that tour ids are unique and independent of other tours in a person's day plan (i.e., the ID of the first shop tour will be the same for each person, independent of other tours). Exceptions include:
    - Tour destination choice, which consists of two models; a sampling of alternatives model in which a subset of zones are first selected according to a simple destination choice model with size terms and distance terms, and a full model which includes other terms including mode choice logsums. The sampling of alternatives model is run repetitively for each zone sampled, so the RNG is re-seeded with tour ID + model hash and random numbers are drawn consecutively for each zone sampled.

- The vehicle allocation model which predicts which vehicle would be used for each of the auto modes (drive-alone, shared 2, and shared 3+) on a tour. In this case the RNG is seeded with tour ID + model hash and random numbers are drawn consecutively once seeded for each mode.

Trip level decisions are seeded with trip ID, which are based on tour id and stop count such that trip ids are unique. One exception is trip destination choice, which uses the same two-stage modeling approach as is used in tour destination choice. The trip mode choice sampling of alternatives model is re-seeded with trip ID + alternative number + model hash for each zone sampled.

Table 6: Random number seeds used in ActivitySim

Model	Household ID	Person ID	Tour ID	Trip ID
auto_ownership_simulate	✓			
work_from_home		✓		
school_location	See text			
workplace_location	See text			
transit_pass_subsidy		✓		
transit_pass_ownership		✓		
vehicle_type_choice	See text			
free_parking		✓		
telecommute_frequency		✓		
transponder_ownership	✓			
cdap_simulate	✓			
mandatory_tour_frequency		✓		
mandatory_tour_scheduling			✓	
school_escorting	✓			
joint_tour_frequency_composition	✓			
joint_tour_participation			✓	
joint_tour_destination	See text			
joint_tour_scheduling			✓	

Model	Household ID	Person ID	Tour ID	Trip ID
non_mandatory_tour_frequency		✓		
non_mandatory_tour_destination	See text			
non_mandatory_tour_scheduling			✓	
vehicle_allocation			✓	
tour_mode_choice_simulate			✓	
atwork_subtour_frequency			✓	
atwork_subtour_destination	See text			
atwork_subtour_scheduling			✓	
atwork_subtour_mode_choice			✓	
stop_frequency			✓	
trip_purpose			✓	
trip_destination	See text			
trip_purpose_and_destination				✓
trip_scheduling				✓
trip_mode_choice				✓
parking_location				✓

The initial code for simulation with Explicit Error Terms uses the same random number seeds as are shown above, but instead of drawing just one random number from the RNG, the software draws random numbers for each alternative in the model in the order in which they appear in the utility specification. For example, in an auto ownership model with 4 alternatives, household ID is used to seed the random number generator, and repeated draws are made for each auto ownership alternative. Note that these draws are repeated without respect to availability of alternative. For example, random numbers are drawn for each zone in the sampling of alternatives destination choice model regardless of whether that zone has a positive size term.