MEMO

TO:	ActivitySim Consortium
FROM:	Joel Freedman and David Hensle, RSG
CC:	
DATE:	April 11, 2022
SUBJECT:	Disaggregate Accessibility Model Implementation Plan

1.1 Introduction

This memorandum describes a design for disaggregate accessibility calculations in ActivitySim. The goal of this task is to establish a process to generate a set of

accessibilities that are consistent with the actual ActivitySim destination and mode choice models, and which can be used by ActivitySim model components requiring destination choice logsums, and post-processing procedures such as logsum based benefit cost assessments. The accessibilities can be used in any ActivitySim model as a substitute for the current aggregate origin-based (e.g. destination choice) accessibilities.

This model implementation plan describes the following features:

- Type of accessibility calculations covered by this task
- Population segments (household and person variables) to include in the synthetic population and which combinations of purposes, departure and arrival times, and modes to calculate accessibilities over.
- User interface for the population\tour generator and level of configurability
- Other options to consider such as whether to provide the ability to output mode choice logsums at a TAZ and/or district level for use in land-use models and/or as a substitute for tour specific mode choice logsums in destination choice.

RSG will integrate the calculator with the MTC extended ActivitySim example and update one of the model components to use a disaggregate accessibility to ensure that the process works as anticipated. RSG is re-estimating the ActivitySim models as part of the SANDAG ABM3 deployment and will use the disaggregate accessibility calculator to refresh the accessibility coefficients used in ActivitySim which would then be made available to other implementations.

1.2 Overview of Accessibilities

Accessibilities used in activity-based travel models can be broadly categorized as follows:

- *Origin-destination accessibilities*. These accessibilities are calculated between an origin and a destination Transportation Analysis Zone (TAZ) or Micro-Analysis Zone (MAZ). These accessibilities can take the form of travel time, generalized cost, or a mode choice logsum. They are used in models where both the tour or trip origin and destination are known. For example, the mode choice logsum for work location choice can be used in auto ownership, because work location choice is run before auto ownership. Mode choice logsums are also used in destination choice models. In destination choice, even though the destination is unknown, the logsum is created for a subset of sampled destination zones since calculating a mode choice logsum for every tour or trip to every potential destination zone is typically computationally prohibitive. The mode choice probability equation is shown in Equation 1, where $P_{a*,i}$ is the probability of choosing mode *i* for decision-maker *a**, and *V* is the utility for mode *i*. Equation 2 shows the mode choice logsum ML_{a*} for decision-maker *a*,* calculated as the natural log of the denominator of the mode choice model.
- *Origin-based accessibilities*. These accessibilities are for origin TAZ or MAZs. They represent the accessibility of the origin zone to all potential destinations. In the current version of ActivitySim, these accessibilities are defined as the total employment or total retail employment accessible within a certain range of travel time for a certain mode; for example, total retail employment within 30 minutes of peak transit time. The logsum of the destination choice model is a more sophisticated treatment of origin-based accessibility due to its ability to capture accessibility for a certain tour or trip purpose, considering all modes of transportation and variables associated with mode choice such as cost and socio-economic variables. Equation 3 shows the destination choice logsum equation. S, is the size term of zone *j*, γ is the coefficient on mode choice logsum, and ML_{a*} is the mode choice logsum term for decision-maker a* (*i* and *j* zones are implied.

Equation 1: Mode Choice Equation

$$P_{a*,i} = \frac{e^{V_{a*,i}}}{\sum_{a \in A} e^{V_{a,i}}}$$

Equation 2: Mode Choice Logsum

$$ML_{a*} = ln(\sum_{a \in A} e^{V_{a,i}})$$

Equation 3: Destination Choice Equation

$$A_i = ln\left[\sum_{j=1}^{I} S_j \, x \, exp\left(\gamma \, ML_{a*}\right)\right]$$

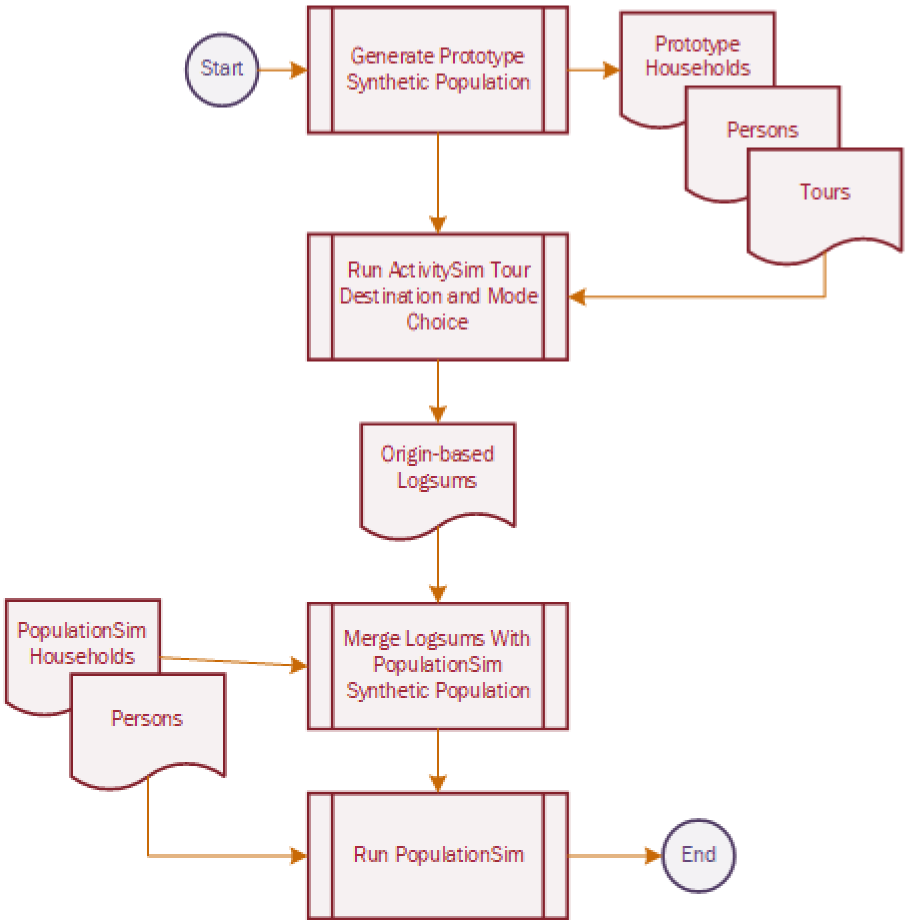
The current version of ActivitySim considers only simple, travel time related origin-based accessibilities. More recent versions of activity-based models, including CT-RAMP and DaySim, use destination-choice based logsums. However, these logsums are calculated using simplified mode and destination choice models that do not utilize the same model specifications as the actual mode and destination choice models used in the disaggregate activity-based model.

For example, the version of CT-RAMP developed for San Diego Association of Governments, and subsequently applied to Metropolitan Transportation Commission (Travel Model Two), Southern Oregon, and Southeast Florida Regional Planning Model calculates 50 different destination choice logsums in a pre-processor. The logsums are a combination of purpose, auto sufficiency and mode. Because they do not use the actual destination or mode choice models, the utility equations are maintained in separate set of Utility Expression Files and there are separate Java classes to create them. The utility expressions are simplified versions of the real models; they do not include all of the alternatives in the real models, or the actual utilities used in the actual models. The alternative-specific constants in the models are approximated. Changes (new modes such as ride-hailing, policies such as VMT taxes, or calibration of constant terms to new data) made to the actual tour or trip mode choice models do not affect the origin-based accessibilities. We wish to address these shortcomings in this task.

1.3 Exploration of SANDAG CT-RAMP Destination Choice Logsums

In order to create a set of disaggregate accessibilities using the ActivitySim destination and mode choice models, a 'fake' or 'prototypical' synthetic population must be created covering all market segments of interest, and a set of tours must be defined for each household and person covering all tour purposes of interest. ActivitySim destination and mode choice models will be run for the tours specified for prototypical population and written out to disk. The database of logsums will then be merged with the actual synthetic population for use in ActivitySim models. This procedure is shown graphically in Figure 1.

Figure 1: Proposed Procedure for Generating Origin-Based Logsums in ActivitySim



This section describes the household, person, and tour segmentation that will be used in the initial specification of the procedure. The next section describes how the user will be able to modify this specification.

To better understand how accessibilities are used in CT-RAMP, we explored the accessibilities used in the SANDAG model system. As noted above, these accessibilities are segmented by:

- Purpose, relating to the size term used in the destination choice model. The purposes are escort, shop, other maintenance, eating out, social/visiting, other discretionary, and at-work. There are also a set of size terms for total employment (which could be used as a proxy for work location choice), for total households, and a "total non-mandatory" segment which is an asserted set of coefficients for all non-mandatory purposes combined. Note that there are no explicit mandatory purposes represented, since work and school location choice are run early in the model system, so origin-destination based accessibilities are used for these purposes.
- Auto sufficiency, relating to the alternative specific constants used in the simplified mode choice model. The constants have the effect of weighting the

contribution of each mode utility (SOV, HOV, transit, and non-motorized) in the overall mode choice logsum for the auto sufficiency group. For example, in the 0-auto segment, transit and non-motorized modes have higher alternative-specific constants than for auto-owning households. There are three auto sufficiency segments, relating to 0 autos, autos < workers, and autos >= workers.

- Mode, relating to the mode choice logsum used as a cost variable. Destination choice logsums are calculated for each purpose and auto sufficiency combination listed above, for SOV modes versus HOV modes. This segmentation was added to support joint tour frequency models, which do not consider drive-alone mode utilities. Transit and non-motorized mode choice logsums are calculated separately for total employment and total household size terms.

We read the table of accessibilities for a base-year (2016) SANDAG model run into a Python Jupyter notebook and explored the correlation between the accessibilities within each auto sufficiency category, segmented by maintenance (escort, shop, other maintenance) and discretionary (eating out, social/visiting, other discretionary) purposes. The results of the analysis (shown in the Appendix) demonstrate a very strong correlation between accessibilities within each segment. The only destination choice logsum terms with a less than 0.9 correlation are in the 0 auto segment; for example between escort and shop, and between SOV logsums and HOV logsums. However, the correlations are still very high, between 0.7 and 0.8, and the correlations tend to be relatively high between all escort/shop and other maintenance, and also between eating out/social and other discretionary. We also note that within a specific purpose such as Shop or Other Discretionary, the correlation between the 0 auto segment and the autos< workers or autos>=workers segments tends to be lower than the correlations between auto owning segments. These correlations inform our recommendations for a more concise set of tour purposes as listed below.

1.4 Recommended Population Segments AND Tours

Table 1 shows all household variables in the synthetic population household file. Each variable name is shown as a row in the table, along with a brief description of the variable. There are three variable "types" indicated in the table, as follows:

- Constant: The variable will not change across households in the prototype population.
- Sequential: Variables such as IDs whose values range from 1 to n
- Controlled: Variables such as household income and auto ownership, whose values will be controlled by the user

We suggest varying household income and household vehicles in the synthetic population, with three levels of income and three levels of auto ownership (0 autos, autos<workers, autos>=workers) so there would be 3x3 or 9 total household segments in the synthetic population.

Table 2 shows the proposed person variables in the synthetic population. Each household would have two persons; a full-time working female age 35 and a non-working male age 55. The full-time worker would have a Mandatory activity pattern (1 work tour) and the non-working adult would have a Non-mandatory activity pattern (2 non-mandatory tours). The default values for tour start and end times and auto operating costs will be used in the mode choice logsum calculations. (This would correspond to the non-household vehicle option in the context of the vehicle allocation model.)

We propose to generate the following tours for destination choice logsums:

- Person 1: Work tour. Although the current SANDAG model does not use origin-based logsum for work, it does create an accessibility to all employment logsum, so this logsum would be a proxy for that. Also it would allow the user to calculate origin-based user benefits for work travel.
- Person 2: Other Maintenance tour (due to strong correlations of logsums between different types of maintenance activities with Other Maintenance purpose.
- Person 2: Other Discretionary tour (due to strong correlations of logsums between different types of discretionary activities with Other Discretionary purpose.

MEMO

Table 1: Household Variables

Variable Name	Description	Type	Value(s)
hhid	Unique Household ID	Sequential	1 to n where n is total number of households in file
household_serial_no	Household serial number	Sequential	1 to n where n is total number of households in file
taz	TAZ of household	Controlled	The set of TAZs to populate
maz	MAZ of household	Controlled	The set of MAZs to populate (if applicable)
hinccat1	Household income category:	Controlled	1 (<$30k), 2 ($30-100k), 4 ($100-$150k), corresponding to hinc values below
hinc	Household income	Controlled	$14k, $67k, $120k (using 10th, 50th, and 75th percentiles)
hworkers	Number of workers in household	Constant	1
veh	Number of vehicles in household	Controlled	0, 1, 2
persons	Number of persons in household	Constant	2
hht	Household/family type:	Constant	1 = Family household: married-couple (2035 mode)
bldgsz	Building size - Number of Units in Structure & Quality:	Constant	2 = One-family house detached (2035 mode)

Table : Person Variables

Column Name	Description	Unique value for Person 1	Unique value for Person 2	Other

hhid	Household ID			Set according to hhid in household file
perid	Person ID			Sequential 1 through n where n is total number of persons in file
pnum	Person Number	1	2	
age	Age of person	35	55	
s e x	Gender of person	2 (female)	1 (male)	
military	Military status			4 (No)
pemploy	Employment status	1 (employed full-time)	3 (unemployed)	
pstudent	Student status			3 (not attending school)
ptype	Person type	1 (full-time worker)	4 (non-working adult)	
educ	Educational Attainment:			13 (Bachelors)
grade	Grade school attending			0 (not attending)
weeks	Weeks worked	1 (50 to 52 weeks)	0	
hours	Hours worked	35	0	
timeFactorWork	Work travel time factor			1 (mean)
timeFactorNonWork	Nonwork travel time factor			1 (mean)
DAP	Daily Activity Pattern	M (Mandatory)	N (Non-Mandatory)	

MEMO

1.5 Implementation

We propose making the implementation configurable so that the user does not need to code in Python in order to change the characteristics or segments of the prototype population or the tours for which to calculate accessibilities.

Inputs include the following:

1) household_accessibility_spec.csv: A csv file listing each household variable, the type of variable, and the attribute levels, similar to Table 1, as shown below. Note the use of semi-colon to differentiate between attribute levels. Also note that certain variables (hhid, household_serial_no, taz, and maz) are special variables where no attributes are specified. The code will need to populate these variables internally.

   Example household_accessibility_spec.csv file:

```
variable,type,values
hhid,sequential,
household_serial_no,sequential,
taz,controlled,
maz,controlled,
hinccat1,controlled,1;3;4
hinc,controlled,14000;67000;120000
hworkers,constant,1
veh,controlled,0;1;2
persons,constant,2
hht,constant,1
bldgsz,constant,2
```

2) person_accessibility_spec.csv: A csv file listing each person variable for each person indicated by `persons` in the household_accessibility_spec.csv file and the attribute levels, similar to . For example:

   Example person_accessibility_spec.csv file:

```
Variable,person_1,person_2
hhid,,
perid,,
pnum,1,2,
age,35,55
sex,2,1
military,4,4,
pemploy,1,3
pstudent,3,3
```

```
ptype,1,4
educ,13,13
grade,0,0
weeks,1,0
hours,35,0
timeFactorWork,1,1
timeFactorNonWork,1,1
DAP,M,N
```

3) tour_accessibility_spec.csv: A file listing the tours to be created and the person to which each tour belongs.

   Example tour_accessibility_spec.csv file:

```
person,tour_num,purpose
1,1,work
2,1,shopping
2,1,othdiscr
```

4) accessibility_config.yaml: A file with settings for the accessibility calculator such as the names of the above input files and the actual synthetic population files to merge the accessibilities with after creating them. The yaml file will have the following additional settings:

   a. Optional: The input zonal data file and a field in the input zone data file indicating which zones to generate prototype synthetic population for. If this field is not provided, the code will assume that it should generate households for all input zones in the zone data file. If it is provided, it would only generate households for those zones with a 1 in the input field. This can be used in the case of 2-zone or 3-zone implementations where generating prototype synthetic population in all zones might result in inordinately long runtimes. In such cases the user can use spatial analysis such as k-means clustering to first identify which micro-analysis zones to include for creation of the synthetic population. If this option is used, the user must also specify a file listing, for each zone with a 1 in the input zone file, which zones to expand the logsum to. This file should cover all zones in the zone data so that the merging process can be completed (see below).

   b. A setting indicating the total number of destinations to sample in the destination choice model. This will override the setting in the tour destination choice yaml files. It provides the user the ability to reduce Monte Carlo simulation variance by increasing the sample rate for destinations used for accessibility calculations.

The code will automatically create the prototype synthetic household and person file, and choosers table for destination choice. It will then run the destination choice models for the tours listed in the tour table and write out the logsums for each tour to an output file.

After the logsums are created, they must be merged with the actual synthetic population so that they can be used in ActivitySim models as explanatory variables. In the case that the logsums are created for the full set of zones, the merge process should be straightforward. Households are matched based on their zone, auto sufficiency and income level. The logsums from all tours generated for the prototype population would then be merged with the household file and saved as additional fields.

In the case that logsums are created for only a subset of zones, then the logsums must first be expanded or copied to the other zones for which logsums were not created before they are merged with the synthetic household file. In this case, an additional file listing proximate zones for each zone must be provided in the accessibility_config.yaml file. Presumably this file would be created based on a closest neighbor calculation.

MEMO

APPENDIX A. ACCESSIBILITY CORRELATIONS

TABLE : 0 AUTO MAINTENANCE PURPOSE ACCESSIBILITY CORRELATIONS

Purpose_Mode_Autos	ESCORT_HOV_0	SHOP_HOV_0	MAINT_HOV_0	SHOP_SOV_0	MAINT_SOV_0
ESCORT_HOV_0	1	0.998791	0.998357	0.795862	0.82752
SHOP_HOV_0	0.998791	1	0.997578	0.79236	0.82173
MAINT_HOV_0	0.998357	0.997578	1	0.789151	0.821699
SHOP_SOV_0	0.795862	0.79236	0.789151	1	0.977054
MAINT_SOV_0	0.82752	0.82173	0.821699	0.977054	1

TABLE : AUTOS<WORKERS MAINTENANCE PURPOSE ACCESSIBILITY CORRELATIONS

Purpose_Mode_Autos	ESCORT_HOV_1	SHOP_HOV_1	MAINT_HOV_1	SHOP_SOV_1	MAINT_SOV_1
ESCORT_HOV_1	1	0.998856	0.998381	0.998882	0.998383
SHOP_HOV_1	0.998856	1	0.997646	0.99999	0.997615
MAINT_HOV_1	0.998381	0.997646	1	0.997653	0.999992
SHOP_SOV_1	0.998882	0.99999	0.997653	1	0.997638
MAINT_SOV_1	0.998383	0.997615	0.999992	0.997638	1

TABLE : AUTOS>=WORKERS MAINTENANCE PURPOSE ACCESSIBILITY CORRELATIONS

Purpose_Mode_Autos	SHOP_HOV_2	MAINT_HOV_2	SHOP_SOV_2	MAINT_SOV_2	
ESCORT_HOV_2	1	0.998816	0.998369	0.998231	0.997664
SHOP_HOV_2	0.998816	1	0.997612	0.999722	0.997255
MAINT_HOV_2	0.998369	0.997612	1	0.997513	0.999747
SHOP_SOV_2	0.998231	0.999722	0.997513	1	0.997683
MAINT_SOV_2	0.997664	0.997255	0.999747	0.997683	1

Table : 0 AUto Discretionary Purpose Accessibility Correlations

Purpose_Mode_Autos	EAT_HOV_0	VISIT_HOV_0	DISCR_HOV_0	EAT_SOV_0	VISIT_SOV_0	DISCR_SOV_0
EAT_HOV_0	1	0.999893	0.940089	0.804684	0.80566	0.67907
VISIT_HOV_0	0.999893	1	0.938152	0.801872	0.803067	0.67414
DISCR_HOV_0	0.940089	0.938152	1	0.842484	0.842726	0.770112
EAT_SOV_0	0.804684	0.801872	0.842484	1	0.999546	0.928212
VISIT_SOV_0	0.80566	0.803067	0.842726	0.999546	1	0.924784
DISCR_SOV_0	0.67907	0.67414	0.770112	0.928212	0.924784	1

Table : Autos<Workers Discretionary purpose Accessibility Correlations

Purpose_Mode_Autos	EAT_HOV_1	VISIT_HOV_1	DISCR_HOV_1	EAT_SOV_1	VISIT_SOV_1	DISCR_SOV_1
EAT_HOV_1	1	0.999903	0.939738	0.99999	0.999903	0.939586
VISIT_HOV_1	0.999903	1	0.937894	0.999882	0.99999	0.937724
DISCR_HOV_1	0.939738	0.937894	1	0.939804	0.937964	0.999977
EAT_SOV_1	0.99999	0.999882	0.939804	1	0.999902	0.939682
VISIT_SOV_1	0.999903	0.99999	0.937964	0.999902	1	0.937824
DISCR_SOV_1	0.939586	0.937724	0.999977	0.939682	0.937824	1

Table : Autos>=Workers Discretionary purpose Accessibility Correlations

Purpose_Mode_Autos	EAT_HOV_2	VISIT_HOV_2	DISCR_HOV_2	EAT_SOV_2	VISIT_SOV_2	DISCR_SOV_2
EAT_HOV_2	1	0.999896	0.940091	0.999775	0.999587	0.940736
VISIT_HOV_2	0.999896	1	0.938172	0.999805	0.999809	0.939
DISCR_HOV_2	0.940091	0.938172	1	0.93792	0.936025	0.999528
EAT_SOV_2	0.999775	0.999805	0.93792	1	0.99991	0.939203
VISIT_SOV_2	0.999587	0.999809	0.936025	0.99991	1	0.937439
DISCR_SOV_2	0.940736	0.939	0.999528	0.939203	0.937439	1

TABLE : SHOP PURPOSE ACCESSIBILITY CORRELATIONS, ALL AUTO SUFFICIENCY GROUPS

	SHOP_HOV_0	SHOP_HOV_1	SHOP_HOV_2	SHOP_SOV_0	SHOP_SOV_1	SHOP_SOV_2
SHOP_HOV_0	1	0.99993	0.999998	0.79236	0.999937	0.99969
SHOP_HOV_1	0.99993	1	0.999946	0.787851	0.99999	0.99991
SHOP_HOV_2	0.999998	0.999946	1	0.791654	0.999951	0.999722
SHOP_SOV_0	0.79236	0.787851	0.791654	1	0.788362	0.78313
SHOP_SOV_1	0.999937	0.99999	0.999951	0.788362	1	0.9999
SHOP_SOV_2	0.99969	0.99991	0.999722	0.78313	0.9999	1

TABLE : OTHER DISCRETIONARY PURPOSE ACCESSIBILITY CORRELATIONS, ALL AUTO SUFFICIENCY GROUPS

	DISCR_HOV_0	DISCR_HOV_1	DISCR_HOV_2	DISCR_SOV_0	DISCR_SOV_1	DISCR_SOV_2
DISCR_HOV_0	1	0.999873	0.999996	0.770112	0.99988	0.999456
DISCR_HOV_1	0.999873	1	0.999908	0.762495	0.999977	0.999842
DISCR_HOV_2	0.999996	0.999908	1	0.768826	0.999911	0.999528
DISCR_SOV_0	0.770112	0.762495	0.768826	1	0.763439	0.754723
DISCR_SOV_1	0.99988	0.999977	0.999911	0.763439	1	0.999831
DISCR_SOV_2	0.999456	0.999842	0.999528	0.754723	0.999831	1