5 Representative Population Data

The SACSIM model requires a detailed population file with representation of key demographics, such as household size, income and age of the population. SACOG uses the following process to generate the representative population file. The key software used in the process is known as “PopGen”, developed by the School of Sustainable Engineering and the Built Environment at Arizona State University along with additional research institutions and metropolitan planning organizations. The current PopGen version is maintained and hosted by the Mobility Analysis Research Group. The program generates synthetic populations whereby both household-level and person-level characteristics of interest can be matched. Moreover, PopGen facilitates the use of Census data. This chapter describes the various aspects involved in the process of generating the representative population for transportation demand forecasting.

5.1 Land Use Scenarios

Parcel-based dwelling units maintained in the land use scenario model is the base for generating representative population. Land use scenarios include “yield” estimates of a range of land use variables at parcel level. Two variables are used directly in SACSIM: dwelling units and jobs by sector. The yield estimates are based on the “place type” (generalized land use) of the parcel, the parcel area, and a number of physical, environmental, or policy constraints. Yield estimates are calibrated to match small area inventories of dwellings and jobs for the base year. Yield estimates are made for future year growth based on the future year place type and development status of each parcel in the future year scenario, and constraints expected to be in place at each parcel in the future year.

5.2 Household/Population Demographics

Travel demand forecasting models like SACSIM have always relied heavily on representations of key demographic characteristics in the population input file used for modeling. The most common demographic variables included in travel demand models are:

- Household size—the number of persons in the household
- Number of workers—the number of working adults in the household
- Household income—usually classified by three or more income categories.

Increasingly, ***age*** has become a variable of interest for travel demand modeling. In part, the increase in interest is related to the so-called "graying" of the population in the United States and elsewhere. Most of older/state-of-practice travel demand models do not account for age in the population demographics. SACSIM is the first of a new generation of "activity-based" travel demand models which allow for more realistic and thorough accounting of demographic variables like age.

The following variables and categories, at census tract level, are used for the household controls in the current process:

- Household size (4 categories: 1, 2, 3, and 4-or-more persons per household)
- Workers per household (4 categories: 0, 1, 2 and 3-or-more workers per household)
- Household Income in 2016 dollars (5 categories: less-than-$20,000; $20,000 to $39,999; $40,000 to $59,999; $60,000 to $99,999; and $100,000-or-more)
- Age of householder, or head-of-household (3 categories: <35 years; 35 to 64 years; and 65-or-more years)
- Ethnicity: (4 categories: 1. White Non-Hispanic, 2. Black Non-Hispanic, 3. Hispanic, and 4. Asian, Pacific Islander, or Other; including people who identify with two ethnic groups and Native Americans and Pacific Islanders. Note SACOG only controlled ethnicity for base year demographic population using 5 year ACS data, but did not use ethnicity as a control for any future year forecasting.

In addition to these household-level variables, the number of university student "households" clustered near to colleges or universities are controlled. As well as the number of senior "households" at senior living facilities. Dorm students and senior living persons are accounted for as Group Quarter population throughout the development of the representative population process. Senior households are still included in general household and population summaries. Person-level controls on age are also included. There are four categories of age controls: 14 years and younger, 15 to 34 years, 35 to 64 years and 65 years and older.

5.3 Base Year household/population demographics

The base year for SACSIM19 is 2016. The ACS data, tract level 5-year sample 2012-2016, provides a comprehensive demographic portrait of the region's population. Based on these ACS tract level totals and ACS Public Use Microdata Sample (PUMS) data, SACOG uses a software tool PopGEN to develop a representative, or synthetic, population to use to model travel within the region using SACSIM.

Table 5-1 through Table 5-5 provide demographic distributions of the five main control variables at county-level in 2016 representative population file. Although reported here at county level, these control variables were established at census tract level.

Table 5-1 Household Size Distribution

SACSIM19 Representative Sample Data

County	1 person	2 persons	3 persons	4+ persons
El Dorado*	21%	40%	17%	22%
Placer*	24%	36%	15%	24%
Sacramento	27%	31%	16%	27%
Sutter	22%	30%	17%	31%
Yolo	24%	32%	18%	27%
Yuba	20%	32%	17%	31%
Total	25%	33%	16%	26%

Source: SACOG 2020.
Based on American Community Survey 2012 5-year Sample data, excludes Tahoe Basin.

Table 5-2 Workers per Household Distribution

SACSIM19 Representative Sample Data

County	0 worker	1 workers	2 workers	3+ persons
El Dorado*	32%	37%	28%	4%
Placer*	30%	35%	29%	6%
Sacramento	27%	40%	27%	6%
Sutter	29%	38%	26%	7%
Yolo	26%	36%	31%	7%
Yuba	32%	40%	24%	5%
Total	28%	38%	28%	6%

Source: SACOG 2020.
Based on American Community Survey 2012 5-year Sample data, excludes Tahoe Basin.

Table 5-3 Household Income Distribution

SACSIM19 Representative Sample Data

County	< $20K	$20K - $40K	$40K - $60K	$60K - $100K	$100K+
El Dorado*	11%	16%	13%	22%	39%
Placer*	11%	14%	13%	23%	39%
Sacramento	17%	19%	16%	23%	26%
Sutter	18%	20%	16%	23%	22%
Yolo	19%	17%	15%	20%	28%
Yuba	18%	24%	18%	23%	16%
Total	16%	18%	15%	23%	28%

Source: SACOG 2020.
Based on American Community Survey 2012 5-year Sample data, excludes Tahoe Basin.

Table 5-4 Year 2020 Age of Householder Distribution

SACSIM19 Representative Sample Data

County	< 35 Years	35 - 64 years	>= 65 Years
El Dorado*	9%	60%	31%
Placer*	14%	57%	29%
Sacramento	21%	57%	21%
Sutter	20%	56%	24%
Yolo	28%	52%	20%
Yuba	27%	53%	20%
Total	20%	57%	23%

Source: SACOG 2020.
Based on American Community Survey 2012 5-year Sample data, excludes Tahoe Basin.

Table 5-5 Year 2016 Household Population Age Distribution

County	<= 14 years	15 - 34 years	35 - 64 years	>= 65 years	Population
El Dorado*	18%	21%	43%	19%	147,202
Placer*	20%	23%	39%	18%	363,896
Sacramento	21%	29%	37%	13%	1,476,573
Sutter	22%	28%	36%	14%	96,392
Yolo	18%	38%	33%	11%	214,784
Yuba	24%	31%	34%	11%	77,464
Total	20%	28%	37%	14%	2,376,311

Source: SACOG 2020.
Based on American Community Survey 2012 5-year Sample data, excludes Tahoe Basin.

5.4 Future Year household/population demographics

SACOG has relied on the Center for Continuing Study of the California Economy (CCSCE) to prepare the long range, region-level projections of population, housing and jobs. These projections were adopted by the SACOG Board of Directors, for use in development and analysis of the MTP/SCS scenarios. The Board-adopted region-level projections serve as control totals and guidance for development of spatially detailed, small area projections.

When establishing marginal controls for future years, a new process is implemented. For the residential areas already developed in 2016 base year, average demographics from 2016 representative population by generalized place type (see Table 5-11 for descriptions), community type and census tract are applied. For the future developing residential areas, regional average demographics by community type and generalized place type from base year 2016 (see Table 5-12 through Table 5-14) is used. Then, the regional demographic projections from CCSCE and expected demographic changes are factored into the base year controls, to ensure that expected/predicted changes in total population, household size, workers, household income, and age are reflected in the future year population files. Table 5-6 through Table 5-10 show the regional demographic 2016 base year and 2040 projections for the control variables used to generate future year representative population. Finally, Table 5-15 through Table 5-18 show the final MTP/SCS future year preferred land use scenario (2040) demographic tables by land use type.

Table 5-6 Age of Householder Distribution of Year

Year	< 35 Years	35 - 64 years	>= 65 Years
2020	21%	56%	24%
2036	17%	50%	33%

Source: SACOG 2020.

Table 5-7 Population Age Distribution of Year 2016/2040

Year	<= 14 years	15 - 34 years	35 - 64 years	>= 65 years
2016	20%	28%	38%	14%
2040	16%	25%	37%	22%

Source: SACOG 2020

Table 5-8 Regional Household Size Distribution of Year 2016/2040

Year	1 person	2 persons	3 persons	4+ persons
2016	25%	33%	16%	26%
2040	27%	32%	16%	25%

Source: SACOG 2020.

Table 5-9 Workers per Household Distribution of Year 2016/2040

Year	0 worker	1 workers	2 workers	3+ workers
2016	28%	38%	28%	6%
2040	30%	37%	27%	5%

Source: SACOG 2020.

Table 5-10 Household Income Distribution of Year 2016/2040

Year	< $20K	$20K - $40K	$40K - $60K	$60K - $100K	>= $100K
2016	16%	18%	15%	23%	28%
2040	16%	18%	16%	22%	28%

Source: SACOG 2020. Income breaks are in year 2016 dollars.

Table 5-11 Description of Generalized Place Types

Place Type #	Description	Examples of Place Types from Land use Model
1	High density	High density attached residential/Very high density attached residential
2	Medium-High density	Medium-High density attached/detached residential
3	Medium density	Medium density attached/detached residential, mobile home
4	Low density	Low density detached residential
5	Very low density	Farm home/Rural residential/Very low density detached residential
6	Mixed/Urban	Residential/Retail Mixed Use Low/High, Urban Attached Residential

Source: SACOG 2020.

Table 5-12 Typical Household Demographics by Community Type and Place Type: Household Size & Workers

Community Type	Place Type	Household Size				Worker per Household			
		HH size 1	HH size 2	HH size 3	HH size 4+	0 worker HH	1 worker HH	2 worker HH	3+ worker HH
Center / Corridor	High Density	45%	18%	11%	26%	51%	32%	11%	6%
	Medium High Density	33%	29%	20%	19%	44%	37%	15%	4%
	Medium Density	36%	37%	12%	15%	45%	35%	17%	3%
	Low Desnity	24%	38%	14%	24%	30%	37%	27%	6%
	Very Low Density	21%	41%	16%	22%	29%	37%	29%	5%
	Mixed Use / Urban	33%	7%	20%	40%	47%	33%	20%	0%
Developing / Established	High Density	56%	24%	9%	12%	41%	41%	15%	3%
	Medium High Density	35%	33%	14%	19%	37%	41%	18%	4%
	Medium Density	36%	33%	12%	18%	41%	33%	21%	4%
	Low Desnity	23%	32%	17%	27%	21%	41%	31%	6%
	Very Low Density	18%	30%	21%	32%	21%	43%	30%	6%
	Mixed Use / Urban	66%	23%	6%	6%	38%	46%	14%	2%
Agricultural / Rural Residential	High Density	47%	24%	11%	18%	48%	34%	14%	4%
	Medium High Density	32%	32%	14%	21%	39%	39%	18%	5%
	Medium Density	30%	32%	13%	25%	36%	34%	24%	6%
	Low Desnity	20%	32%	18%	30%	22%	38%	33%	7%
	Very Low Density	17%	36%	19%	28%	21%	39%	35%	6%
	Mixed Use / Urban	48%	23%	9%	20%	56%	29%	12%	4%

Source: SACOG 2020.

Table 5-13 Typical Household Demographics by Community Type and Place Type: Household Income & Age

Community Type	Place Type	Household Income					Householder's Age		
		< 15K	15K-30K	30K-50K	50K-75K	>=75K	Age < 35	Age 35-64	Age >= 65
Center / Corridor	High Density	20%	21%	12%	23%	25%	5%	18%	77%
	Medium High Density	22%	21%	16%	23%	19%	6%	31%	62%
	Medium Density	22%	20%	16%	23%	20%	9%	47%	44%
	Low Desnity	14%	20%	19%	26%	21%	14%	60%	26%
	Very Low Density	11%	16%	13%	24%	36%	12%	64%	24%
	Mixed Use / Urban	13%	13%	27%	13%	33%	7%	0%	93%
Developing / Established	High Density	31%	22%	16%	16%	15%	30%	38%	32%
	Medium High Density	26%	24%	17%	18%	15%	24%	44%	33%
	Medium Density	23%	24%	16%	19%	18%	23%	45%	32%
	Low Desnity	18%	20%	17%	23%	23%	30%	60%	10%
	Very Low Density	18%	23%	16%	21%	23%	29%	64%	7%
	Mixed Use / Urban	31%	20%	14%	15%	21%	36%	40%	25%
Agricultural / Rural Residential	High Density	28%	22%	17%	18%	15%	18%	35%	48%
	Medium High Density	21%	22%	17%	20%	20%	17%	42%	41%
	Medium Density	18%	19%	16%	23%	24%	19%	51%	31%
	Low Desnity	12%	16%	15%	25%	32%	21%	64%	15%
	Very Low Density	9%	13%	13%	22%	42%	17%	69%	14%
	Mixed Use / Urban	27%	26%	16%	18%	14%	12%	30%	57%

Source: SACOG 2020.

Table 5-14 Typical Person Age & Ethnicity Group Demographics by Community Type and Place Type

Community Type	Place Type	Person's Age				Person's General Ethnicity Group			
		<= 14 year	15 - 34 yea	35 - 64 yea	>= 65 year	White, Non-Hispanic	Black, Non-Hispanic	Hispanic	Asian and Other
Center / Corridor	High Density	12%	16%	34%	38%	55%	41%	0%	5%
	Medium High Density	11%	21%	31%	36%	65%	23%	1%	11%
	Medium Density	12%	20%	38%	29%	71%	19%	1%	8%
	Low Desnity	18%	25%	42%	16%	65%	20%	2%	13%
	Very Low Density	19%	22%	43%	16%	76%	13%	1%	10%
	Mixed Use / Urban	14%	14%	31%	41%	65%	35%	0%	0%
Developing / Established	High Density	13%	34%	32%	22%	50%	26%	7%	17%
	Medium High Density	16%	31%	33%	20%	42%	29%	8%	21%
	Medium Density	18%	29%	33%	20%	48%	26%	7%	19%
	Low Desnity	22%	34%	37%	7%	42%	27%	8%	23%
	Very Low Density	27%	31%	37%	5%	44%	28%	6%	22%
	Mixed Use / Urban	8%	38%	35%	19%	56%	21%	7%	17%
Agricultural / Rural Residential	High Density	15%	28%	29%	29%	43%	30%	7%	20%
	Medium High Density	16%	27%	32%	25%	49%	25%	6%	20%
	Medium Density	19%	28%	35%	18%	50%	23%	6%	22%
	Low Desnity	23%	29%	40%	9%	51%	21%	5%	23%
	Very Low Density	23%	25%	43%	9%	66%	15%	3%	17%
	Mixed Use / Urban	16%	25%	25%	34%	49%	37%	2%	11%

Source: SACOG 2020.

Table 5-15 SACSIM19 forecasted Demographic Distributions (2040): Household Size

Landuse Type	HH size 1	HH size 2	HH size 3	HH size 4+
Agriculture	17%	38%	17%	28%
College/University	100%	0%	0%	0%
High Density Residential	45%	24%	12%	19%
Low Density Residential	19%	33%	17%	30%
Medium Density Residential	26%	32%	15%	27%
Medium-High Density Residential	31%	34%	15%	20%
Mixed Use	49%	23%	11%	17%
Rural Residential	18%	40%	17%	25%
Very High Density Residential	55%	24%	9%	12%
Very Low Density Residential	16%	37%	18%	28%
Regional	27%	32%	16%	25%

Source: SACOG 2020.

Table 5-16 SACSIM19 forecasted Demographic Distributions (2040): Household Workers

Landuse Type	0 worker HH	1 worker HH	2 worker HH	3+ worker HH
Agriculture	20%	44%	32%	4%
College/University	84%	16%	0%	0%
High Density Residential	47%	35%	15%	3%
Low Density Residential	22%	37%	35%	6%
Medium Density Residential	29%	38%	28%	6%
Medium-High Density Residential	39%	39%	18%	4%
Mixed Use	46%	37%	14%	3%
Rural Residential	27%	37%	32%	5%
Very High Density Residential	43%	41%	14%	2%
Very Low Density Residential	21%	37%	37%	5%
Regional	30%	37%	27%	5%

Source: SACOG 2020.

Table 5-17 SACSIM19 forecasted Demographic Distributions (2040): Household Income

Landuse Type	HH Income < 15K	HH Income 15K-30K	HH Income 30K-50K	HH Income 50K-75K	HH Income >=75K
Agriculture	13%	18%	22%	20%	27%
College/University	100%	0%	0%	0%	0%
High Density Residential	28%	22%	20%	14%	16%
Low Density Residential	13%	17%	20%	20%	29%
Medium Density Residential	18%	20%	21%	18%	23%
Medium-High Density Residential	23%	23%	22%	16%	16%
Mixed Use	27%	21%	20%	14%	18%
Rural Residential	11%	16%	19%	20%	34%
Very High Density Residential	32%	22%	18%	13%	16%
Very Low Density Residential	10%	14%	18%	20%	39%
Regional	18%	19%	20%	18%	25%

Source: SACOG 2020. Income breaks are in year 2000 dollars.

Table 5-18 SACSIM19 forecasted Demographic Distributions (2040): Householder Age

Landuse Type	HH Holder Age < 35	HH Holder Age 35-64	HH Holder Age >= 65
Agriculture	16%	68%	16%
College/University	95%	3%	3%
High Density Residential	14%	28%	58%
Low Density Residential	19%	61%	20%
Medium Density Residential	18%	51%	30%
Medium-High Density Residential	13%	34%	53%
Mixed Use	14%	30%	56%
Rural Residential	11%	60%	29%
Very High Density Residential	22%	33%	44%
Very Low Density Residential	14%	65%	21%
Regional	17%	50%	33%

Source: SACOG 2020.

5.4.1 Representative Population Generation Process

A representative population file includes one data record per person in the base year or future year to be analyzed. This section describes the steps SACOG uses to generate a representative population.

5.4.1.1 Prepare Input files for PopGen

As described above, SACOG uses an open source synthetic population software called, PopGen, which requires three sets of input files:

Marginal control files for the control variables, micro-samples collected by the Census Bureau, and geographic corresponding file describing relationships among levels of geographic areas. To prepare household marginal control file, dwelling units from land use model are aggregated to census tracts and are converted to households by applying adjusted occupancy rates based on ACS 5-year sample data at census tract level. Then the households are split by predefined categories of control variable using ACS 5-year demographic profiles for the base year or adjusted base year demographics with factors of expected/projected future demographics for the future year. The population control totals for each tract are estimated using the number of households by household size controls. Then the age demographic profile from ACS 5-year sample from base year is applied to create population controls by age for the base year, while adjusted age demographics by regional projected factors is applied to create future year population controls. The population file also includes persons ethnic categories based on ACS 5-year sample, however, this ethnic category information is not used in the at all by SACSIM19 therefore does not have an effect in any forecasting. The Group Quarter marginal control file is created with two categories,

1. Representing dorm students in UC Davis and California State University Sacramento,
2. Representing seniors living in senior living facilities.

Sample Files are used to create this micro-sample file for PopGen, Public Use Micro-sample (PUMS) files from Census are modified by recoding control variables to the defined categories. 2012-2016 5-year PUMS survey was used and joined into the 2010 iPUMS GIS files to build the relationships. The following households are identified as clustered university student households, or seniors in senior living facilities: any one- or two- person households in which all members are students or seniors.

Correspondence file is used to connect the Sample and Marginal files.

5.4.1.2 Run PopGen

Based on marginal control files, PopGen estimates household and person type constraints using the Iterative Proportional Fitting procedure. Then it estimates sample household weights that satisfy both household and person type constraints using the Iterative Proportional Updating (IPU) algorithm. The resulting outputs from PopGen are synthetic household and population files. Below are the table structures used by SACOG staff; PopGen requires some of these inputs but not all inputs. For more information on how to download, install and run PopGen, go to the Mobility Analytics PopGen webpage: *https://www.mobilityanalytics.org/popgen.html. Appendix G Run PopGen provides screenshots of the steps SACOG staff uses to run a PopGen scenario to generate a*

synthetic population. Appendix G also includes brief tips for troubleshooting installation using the Population installation instructions on the PopGen webpage.

5.4.1.2.1 PopGen SACOG Input File Descriptions

Table 5-19 Marginal File Population File Field Descriptions

Columns for Popgen	Type	Description
state	bigint	State # ID, California is 6
county	bigint	County # ID
tract	bigint	Census Tract ID
bg	bigint	Block Group ID, Left as 0 for SACOG (controlled at Tract Level)
pop0	bigint	# of Persons Age Group: <= 14 years
pop15	bigint	# of Persons Age Group: 15 - 34 years
pop35	bigint	# of Persons Age Group: 35 - 64 years
pop65	bigint	# of Persons Age Group: >= 65 years
eth_w_nh	bigint	# of Persons General Ethnicity Group: White, Non-Hispanic
eth_b_nh	bigint	# of Persons General Ethnicity Group: Black, Non-Hispanic
eth_h	bigint	# of Persons General Ethnicity Group: Hispanic
eth_a_o	bigint	# of Persons General Ethnicity Group: Asian and Other

Source: SACOG 2020.

Table 5-20 Marginal File Household File Field Descriptions

Columns for Popgen	Type	Description
state	bigint	State # ID, California is 6
county	bigint	County # ID
tract	bigint	Census Tract ID
bg	bigint	Block Group ID, Left as 0 for SACOG (controlled at Tract Level)
hhsize1	bigint	# of Households Group: 1 Person Households
hhsize2	bigint	# of Households Group: 2 Person Households
hhsize3	bigint	# of Households Group: 3 Person Households
hhsize4	bigint	# of Households Group: 4+ Person Households
worker0	bigint	# of Households Group: 0 Worker Households
worker1	bigint	# of Households Group: 1 Worker Households
worker2	bigint	# of Households Group: 2 Worker households
worker3	bigint	# of Households Group: 3+ Worker Households
hh_inc_1	bigint	# of Households Group: Household with Annual Income <15k
hh_inc_2	bigint	# of Households Group: Household with Annual Income >=15k and <30k
hh_inc_3	bigint	# of Households Group: Household with Annual Income >=30k and <50k
hh_inc_4	bigint	# of Households Group: Household with Annual Income >=50k and <75k
hh_inc_5	bigint	# of Households Group: Household with Annual Income >=75k
hh_hd_1	bigint	# of Households Group: Householder Age < 35
hh_hd_2	bigint	# of Households Group: Householder Age 35 - 64
hh_hd_3	bigint	# of Households Group: Householder Age >=65
univ1	bigint	# of Households Group: Students in Dorm Living Facilities
univ2	bigint	# of Households Group: Seniors in Senior Living Facilities

Source: SACOG 2020.

Table 5-21 Marginal File Group Quarter File Field Descriptions

Columns for Popgen	Type	Description
state	bigint	State # ID, California is 6
county	bigint	County # ID
tract	bigint	Census Tract ID
bg	bigint	Block Group ID, Left as 0 for SACOG (controlled at Tract Level)
type1	bigint	# of College/University Students
type2	bigint	# of Seniors in Senior or Assisted Living Facilities

Source: SACOG 2020.

Table 5-22 Correspondence File Group Quarter File Field Descriptions

Columns for Popgen	Type	Description
county	bigint	State # ID, California is 6
tract	bigint	County # ID
bg	bigint	Census Tract ID
state	bigint	Block Group ID, Left as 0 for SACOG (controlled at Tract Level)
pumano	bigint	Public Use Micro-sample ID
stateabb	bigint	State Abbreviation, "CA"
countyname	text	County Name

Source: SACOG 2020.

The Sample file used is the Public Use Microdata Sample (PUMS) data 2012-2016 5-year average sample for PUMA within the SACOG region, available to download from the Census webpage.

5.4.1.3 Post-processing PopGen Outputs

The synthetic population file generated from PopGen lists the frequencies of micro-sample in each census tract, and only include control variables in categories. The population file used in SACSIM requires per-person per record and actual values instead categories for the variables. Also the SACSIM input file requires some additional variables such as gender, student status etc. Therefore, the output from PogPen is post-processed to meet SACSIM's requirements.

5.4.1.4 Allocate Household/population to Parcels

Finally, a separate, customized program is used to allocate the representative population households to parcels within tracts considering characteristics of the parcel such as land use type, distance to transit, distance to schools etc. See Chapter 4 for more information on how to setup and run the household and population to parcel allocation process.