California Statewide Travel Demand Model, Version 2.0

Transportation Analysis Zones

final

report

prepared for

California Department of Transportation

prepared by

Cambridge Systematics, Inc.

and

HBA Specto, Inc.

June 2014

www.camsys.com

final report

California Statewide Travel Demand Model, Version 2.0

Transportation Analysis Zones

prepared for

California Department of Transportation

prepared by

Cambridge Systematics, Inc.
555 12th Street, Suite 1600
Oakland, CA 94607

date

June 2014

Table of Contents

1.0 Introduction **1-1**

2.0 TAZ System Development **2-1**

2.1 Modifying the TAZ System 2-1

Compatibility with 2010 Census Boundaries 2-1

Compatibility with COABDIS Boundaries 2-3

Other Edits 2-5

2.2 External Zones 2-8

3.0 TAZ System Summary **3-1**

4.0 File Names and Descriptions **4-1**

List of Tables

Table 2.1 External Zones for the CSTDM 2-8
Table 3.1 Distribution of Year 2010 TAZs 3-1

List of Figures

Figure 2-1 Before TAZ Modification (Census Boundary) 2-2

Figure 2-2 After TAZ Modification (Census Boundary)..... 2-2

Figure 2-3 Zone Splits along CARB COABDIS Boundaries, Northern California..... 2-4

Figure 2-4 Zone Splits along CARB COABDIS Boundaries, Southern California..... 2-4

Figure 2-5 Before TAZ Modification (Air Basin Boundary) 2-5

Figure 2-6 After TAZ Modification (Air Basin Boundary) 2-5

Figure 2-7 Before TAZ Modification (Random Anomalies) 2-6

Figure 2-8 After TAZ Modification (Random Anomalies) 2-6

Figure 2-9 Coastline Anomalies 2-7

1.0 Introduction

This report documents the development of the Transportation Analysis Zones (TAZs) for use in the California Statewide Travel Demand Model (CSTDM), Version 2.0. The updated TAZ system for the CSTDM Version 2.0 is based on the zonal system developed by ULTRANS Institute of Transportation Studies at University of California, Davis, and HBA Specto, Inc. for the future year forecasts (CSTDM2009_FutureTAZ_2009.shp), developed in 2011. Documentation on previous CSTDM zonal systems are included in Appendix A and B. Some modifications of the zonal system were required to be consistent with 2010 U.S. Census block groups and the California Air Resources Board (CARB) Air Basin District boundaries.

While the ideal scenario is for the TAZs to be compatible with Census blocks, CARB districts, counties, and MPOs TAZs, these boundaries do not always align. As such, some TAZs cross CARB Air Basin District boundaries. However, these issues only occur in rural areas, and do not significantly impact travel demand or greenhouse gas emissions calculations.

The CSTDM09 TAZ system included different numbers of TAZs in future years compared to the base year model. Zones were split in future years to better represent zonal detail in high growth areas. For CSTDM Version 2.0, the base and future year zone systems have been standardized to improve ease and consistency of analysis.

2.0 TAZ System Development

2.1 MODIFYING THE TAZ SYSTEM

Compatibility with 2010 Census Boundaries

Two key goals in updating the zone system were that TAZs should be aggregations of Year 2010 Census boundaries, and that TAZs should also respect County boundaries.

A multi-step union and processing procedure between the CSTDM Version 2.0 TAZs and 2010 Census blocks was implemented to identify TAZs which should be manually edited. The process focused on where TAZ boundaries and Census blocks did not precisely match up. For example, locations were identified where between 5 and 95% of the Census block was contained within the TAZ, instead of the block being 100% contained within the TAZ.

Over 2,000 unique TAZs were identified as containing irregular Census blocks. In a number of cases, the boundary differences were slight, and TAZ boundaries were adjusted to match Census block boundaries.

The list of TAZs with irregular Census blocks was filtered for those where the sum of the irregular areas was greater than 0.25 square miles or where the total percent of the TAZ area affected was greater than 10%[1]. This filtering eliminated areas where minor differences in boundaries occurred but were very small TAZs or the blocks were very small areas. Some of these very small alignment discrepancies were confirmed low-priority edits because the alignments lie along a freeway corridor, for example, where TAZ modification would provide no difference in model data inputs or performance. Figures 2.1 and 2.2 illustrate the types of modifications that were applied to the CSTDM TAZs.

[1] As time permitted, TAZs within 5-10% were also corrected.

Figure 2-1 Before TAZ Modification (Census Boundary)

Figure 2-2 After TAZ Modification (Census Boundary)

Compatibility with COABDIS Boundaries

The other main goal in updating the TAZs was that they should conform to CARB District Boundaries (referred to as COABDIS). As shown in Figures 2-3 and 2-4, some additional zone splits were added along those COABDIS boundaries.

Figure 2-3 Zone Splits along CARB COABDIS Boundaries, Northern California

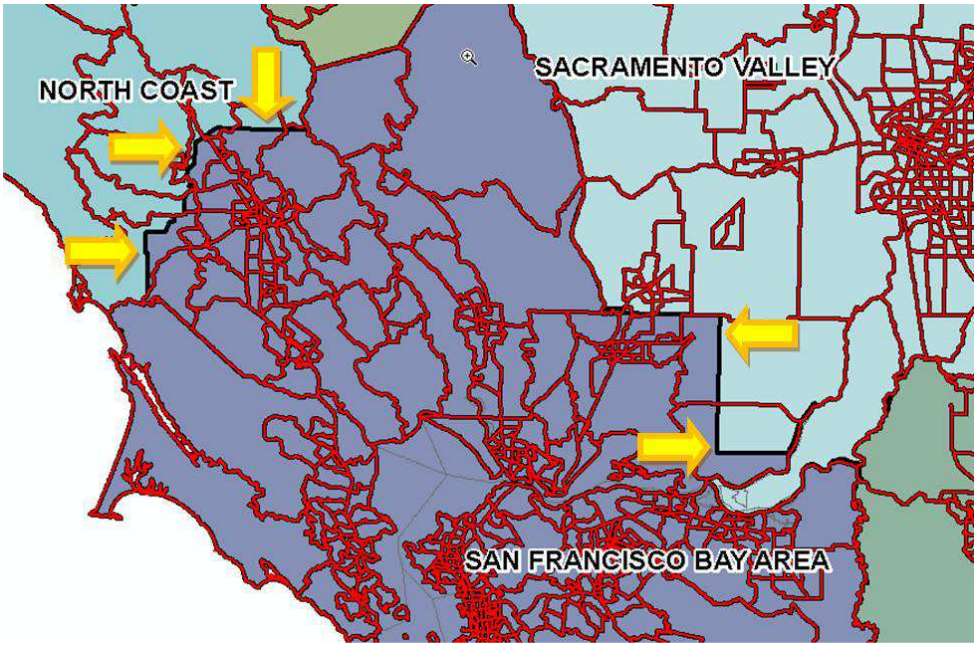
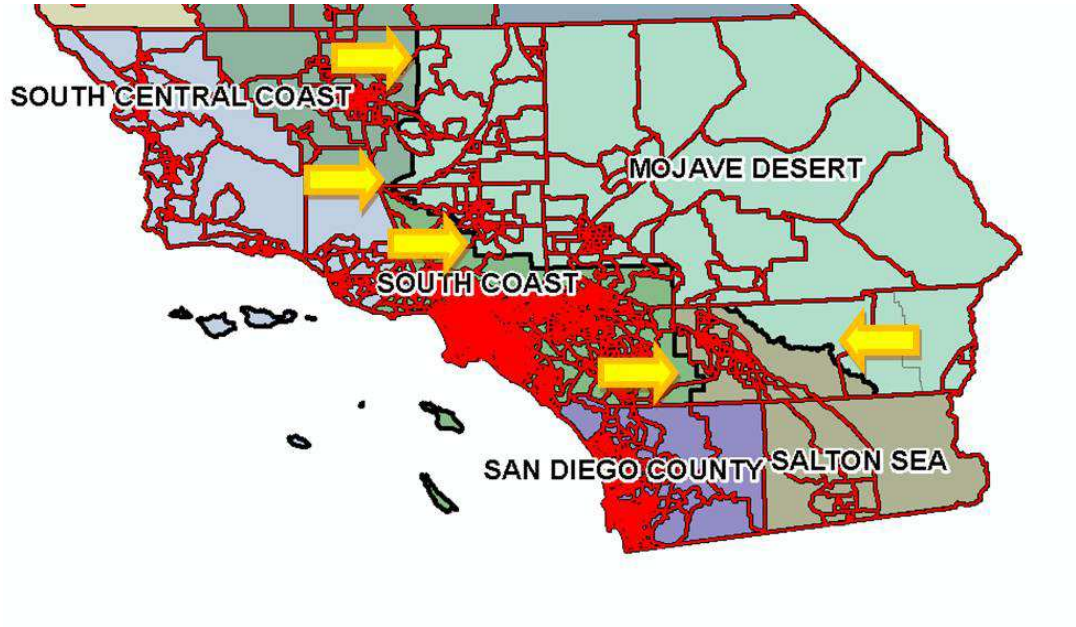


Figure 2-4 Zone Splits along CARB COABDIS Boundaries, Southern California



TAZs were split where COABDIS boundaries did not follow TAZ boundaries. However, COABDIS boundaries did not always respect Census boundaries; therefore, the nearest Census Tract boundary was used for TAZ delineation near the COABDIS boundary. The areas in which the TAZ boundary was off relative to the COABDIS boundary were in rural areas, with little to no population within

the gaps between TAZ and COABDIS boundaries (see Figures 2-5 and 2-6). Furthermore, additional TAZs were split along Federal 8-hour Non-Attainment Boundaries, as requested by CARB staff. These locations and proposed TAZ splits were reviewed and approved by CARB and Caltrans staff.

Figure 2-5 Before TAZ Modification (Air Basin Boundary)

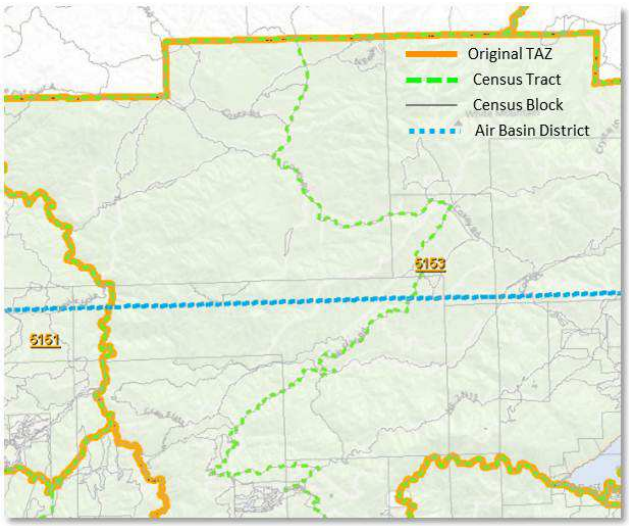
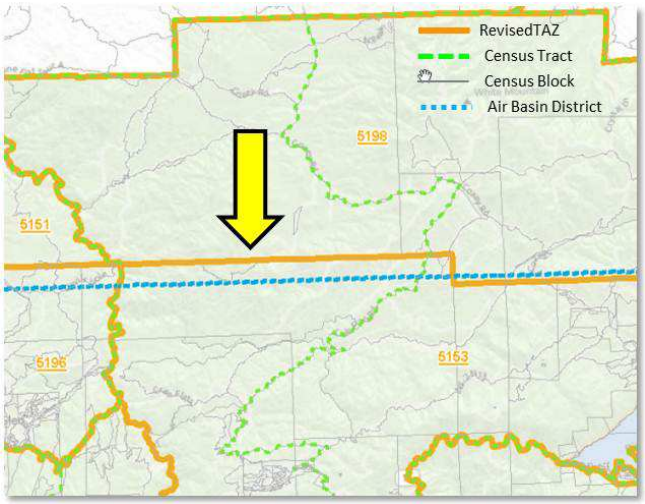


Figure 2-6 After TAZ Modification (Air Basin Boundary)



Other Edits

Random editing anomalies were also corrected, when encountered. An example is provided in Figures 2.7 and 2.8. The modifications resulted in more logical zone boundaries – rather than boundaries that include arbitrary zig-zags.

Figure 2-7 Before TAZ Modification (Random Anomalies)

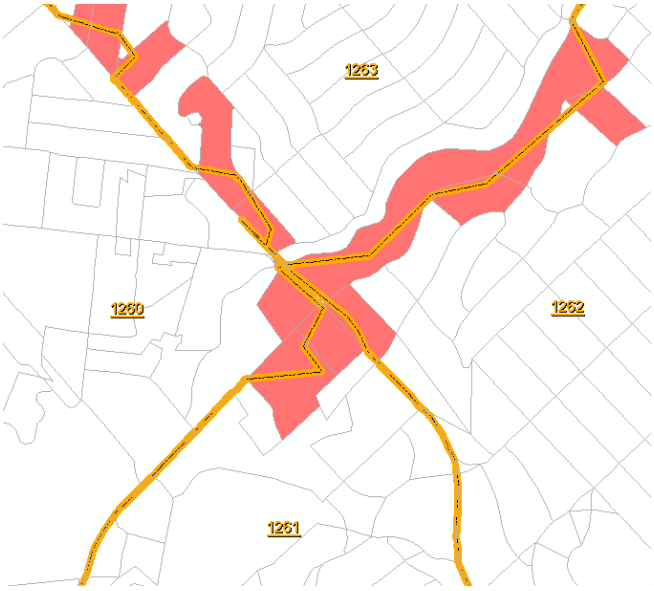
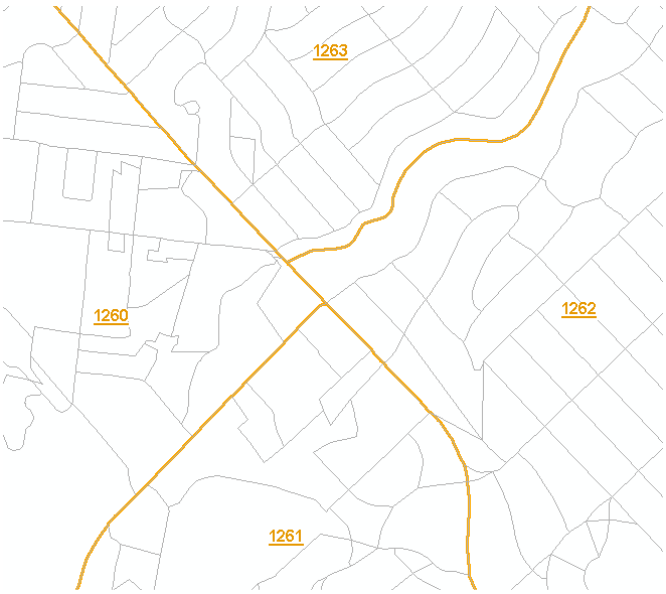


Figure 2-8 After TAZ Modification (Random Anomalies)



There were additional discrepancies along the coastline between TAZs and Census blocks, as shown in Figure 2.9. These coastal TAZs were not modified but a processing method allocated 100% of the Census block areas to the appropriate TAZs. To find the appropriate TAZ, a routine was written to find the nearest TAZ centroid for all the unallocated blocks. These centroid assignments were plotted for reasonableness and modified if need be.

Figure 2-9 Coastline Anomalies

2.2 EXTERNAL ZONES

The external TAZs were not modified from the TAZ system developed for CSTDM09. The existing external zones use node numbers 1 through 53, as listed in Table 2.1. For more information on the external zones, please refer to the External Travel Model Report.

Table 2.1 External Zones for the CSTDM

TAZ	External District	Name
1	Oregon	US 101 – N. Smith River
2	Oregon	US 199 – N. Patrick Creek
3	Oregon	Indian Creek Road
4	Oregon	I- 5 – N. Hilt
5	Oregon	US 97 – N. Dorris
6	Oregon	Cal 139 – N. Tulelake
7	Oregon	Cal 50 – S. Main (Ore)
8	Oregon	Muldoon Rd / Westside Rd
9	Oregon	US 395 – N. New Pine Creek
10	Nevada North	Surprise Valley Road
11	Nevada North	Cal 299 – E. Cedarville
12	Nevada North	Cal 81 – S. Eagleville
13	Nevada North	Wendel Rd / Summers Rd E. Wendel / Herlong
14	Nevada North	US 395 – N. Lemmon Valley (Nev)
15	Nevada North	Henness Pass Road/ Sierra County Road S860
16	Nevada North	I- 80 – W.Reno (Nev)
17	Nevada North	CAL 28 – E. Kings Beach
18	Nevada North	US 50 – N. South Lake Tahoe
19	Nevada North	Cal 88 – N. Paynesville
20	Nevada North	US 395 – S.Topaz Lake (Nev)
21	Nevada North	Cal 182 – N.Bridgeport
22	Nevada North	Cal 167 – E.Mono
23	Nevada North	US 6 – N.Benton
24	Nevada South	Cal 266 – N.Oasis
25	Nevada South	Cal 266 – E.Oasis
26	Nevada South	N Hwy (Cal 267) – Death Valley
27	Nevada South	Daylight Pass Rd (CAL 374) – Death Valley

TAZ	External District	Name
28	Nevada South	Cal 127 – N.Death Valley Junction
29	Nevada South	Ash Meadows Rd – E.Death Valley Junction
30	Nevada South	Cal 178 – W.Pahrump (Nv)
31	Nevada South	I- 15 – N.Mountain Pass
32	Nevada South	Nipton Rd (Cal 164) – E.Nipton
33	Nevada South	US 95 – N.Arrowhead Junction
34	Nevada South	US 95 – E. Laughlin!
35	Arizona	I- 40 – W.Topock (Nev)
36	Arizona	Parker Dam (N. Parker)
37	Arizona	Cal 62 – Earp
38	Arizona	Agnes Wilson Rd E. 95 (S. of Earp)
39	Arizona	I- 10 – E.Blythe
40	Arizona	E. Laguna Rd at Imperial Reservoir (Imperial Dam)
41	Arizona	S. 4th Ave – Yuma
42	Mexico	Cal 186 – Mexican Border Yuma
43	Mexico	Cal 111 – Calexico
44	Mexico	Cal 188 – Tecate
45	Mexico	I- 5 – San Diego
46	Arizona	I- 8 – Yuma
47	Mexico	Collina del Sol (Cal 905) – San Diego
48	Mexico	Cal 7 – Calexico
49	Major Seaport	Long Beach (Port)
50	Major Seaport	Los Angeles (Port)
51	Major Seaport	Oakland (Port)
53	Mexico	Otay Mesa East (opens 2016)

3.0 TAZ System Summary

The final zonal system included 5,454 internal TAZs, adding 33 new TAZs, and has the broad properties reported in Table 4.1. The highest density TAZs include special downtown TAZs that were already very small in geographic area. Another special exception included very high employment in one TAZ in San Diego that represents Camp Pendleton military base, where no detailed Census information can be obtained to split that zone.

After finalizing the CSTDM zone system, a correspondence file between CSTDM TAZs and COABDIS districts was provided to CARB.

Table 3.1 Distribution of Year 2010 TAZs

Total number of TAZs	5,454
TAZs with 2010 Pop below 1,000	392
TAZs with 2010 Pop above 10,000	1,000
TAZs with 2010 Pop above 15,000	76
Mean TAZ Population	6,830
Minimum TAZ Population	0
Maximum TAZ Population	37,491
Standard Deviation Population	3,649
TAZs with 2010 Employment below 100	96
TAZs with 2010 Employment above 10,000	224
TAZs with 2010 Employment above 15,000	64
TAZs with 2010 Employment above 25,000	19
Mean TAZ Employment	2,996
Minimum TAZ Employment	0
Maximum TAZ Employment	72,056
Standard Deviation Employment	3,776

4.0 File Names and Descriptions

All files are located in the GIS directory within the installation of CSTDM:

- **CSTDM_v2.0_TAZ.shp:** TAZ, County, Area (square miles)

A. CSTDM09 TAZ Documentation

CSTDM09 - California Statewide Travel Demand Model

Model Development

Transportation Analysis Zones

System Documentation: Technical Note

ULTRANS
Institute of Transportation Studies, UC Davis
Davis, California

HBA Specto, Incorporated
Calgary, Alberta

May 2011

Table of Contents

1. Introduction 5

2. TAZ System Development 5

2.1 Modifying the TAZ System 5

2.2 External Zones 8

3. Implementation Process 10

3.1 Numbering Systems 12

3.2 TAZs Exceeding Recommended Targets 13

4. Results 13

5. File Names and Descriptions 24

Tables:

Table 1: Identification of HSR TAZs to be Split or Merged for CSTDM..... 6
Table 2: Example of Split-Merge Identification Procedure 6
Table 3: External Zones for the CSTDM. 9
Table 4: Distribution of TAZ by Population and Employment Size..... 14
Table 5: List of Districts..... 17

Figures

Figure 1: HSR TAZ Split Merge Codes Displayed by Color 7

Figure 2: TAZs with Population > 15,000 and/or Employment > 10,000 15

Figure 3: Transportation Analysis Zones..... 22

Figure 4: TAZ Districts..... 23

1. Introduction

This technical note documents the development of the Transportation Analysis Zones (TAZs) for use in the California Statewide Travel Demand Model (CSTDM). The TAZ system for the CSTDM drew on the TAZ used in High Speed Rail (HSR) model networks; however, major modifications were made to those TAZs and networks to suit the specific needs of the CSTDM.

In delineating TAZs for the CSTDM, the effects of TAZ size on both the accuracy and practicality of the model were considered. If TAZs are too large, then forecast accuracy can be compromised due to (1) non-homogenous land uses, (2) too many trips loaded to the network in a small area, (3) loading points that are not representative of the zone, and (4) too many intrazonal trips. On the other hand, numerous small zones can reduce the practicality of the model system by significantly increasing total model run time both in the matrix calculation and trip assignment phases.

Another consideration in the development of the TAZ system was the consistency with both U.S. Census block groups and the zones used in the statewide land use model (California PECAS), which is currently under development. In general, LUZs nest within Counties, TAZs nest within PECAS Land Use Zones (LUZs), and block groups nest within TAZs.

2. TAZ System Development

2.1 Modifying the TAZ System

The modifications of the TAZ system for the CSTDM were guided by a number of criteria. Each TAZ should contain less than 10,000 employees and between 6,000 and 15,000 residents based on the 2030 projections used in the HSR study. These criteria were used to identify the HSR TAZs that needed to be split or expanded. As described above, TAZs should be aggregations of year 2000 census block groups and should respect existing LUZ boundaries.

Table 1 describes the procedure used to identify zones that needed to be split or merged. Using the 2030 zonal total population and employment totals from the HSR study along with the targets stated above (a maximum of 15,000 residents or 10,000 employees), a score value was calculated for each zone. This value was the greater of two figures: total zonal population or employment multiplied by 1.5. Then, based on the score value, each zone was assigned a code (1 to 5), which corresponded to a split or merge decision. TAZs with codes of 4 and 5 were split into an appropriate number of zones. Codes 1 and 2 were combined. An example (from San Diego) of the split-merge identification procedure is shown in Table 2. The split-merge codes assigned to all HSR TAZs in the state are displayed by color in Figure 1.

Table 1: Identification of HSR TAZs to be Split or Merged for CSTDM

Score Value =	Max of 2030 Population OR 1.5 * 2030 Employment	
Code 5	Score > 30,000	Split (into score/15,000)
Code 4	Score between 15,000 & 30,000	Split (into 2)
Code 3	Score between 6,000 & 15,000	No Change
Code 2	Score between 2,000 & 6,000	combine (2 into 1)
Code 1	Score < 2,000	combine (more than 2 into 1)

Table 2: Example of Split-Merge Identification Procedure

County	TAZ	LUZ	Area	2000 Population	2030 Population	2000 Employment	2030 Employment	Score	Code	Split	Num new	Combine
SAN DIEGO	1140	352	27	2,281	24,109	16,592	38,293	57,440	5	4	3	0
SAN DIEGO	1142	325	92	1,628	5,041	360	2,120	5,041	2	0	0	0.5
SAN DIEGO	1143	326	453	5,690	9,949	2,052	4,182	9,949	3	0	0	0
SAN DIEGO	1144	325	70	2,561	2,881	369	552	2,881	2	0	0	0.5
SAN DIEGO	1145	326	157	2,026	2,442	491	1,098	2,442	2	0	0	0.5
SAN DIEGO	1146	309	683	585	1,468	66	177	1,468	1	0	0	0.75
SAN DIEGO	1147	309	575	5,891	8,828	2,075	3,699	8,828	3	0	0	0
SAN DIEGO	1148	309	395	2,498	7,500	1,490	3,306	7,500	3	0	0	0
SAN DIEGO	1149	327	89	4,029	5,111	721	2,133	5,111	2	0	0	0.5
SAN DIEGO	1150	324	6	28,331	31,653	4,119	5,669	31,653	5	3	2	0
SAN DIEGO	1151	354	8	31,764	34,259	3,487	5,086	34,259	5	3	2	0
SAN DIEGO	1152	354	5	41,624	42,789	6,124	11,352	42,789	5	3	2	0
SAN DIEGO	1153	354	6	44,847	49,119	11,037	15,531	49,119	5	4	3	0
SAN DIEGO	1154	353	20	19,667	90,774	6,063	26,938	90,774	5	7	6	0
SAN DIEGO	1155	324	3	2,207	2,469	624	1,312	2,469	2	0	0	0.5
SAN DIEGO	1156	324	6	5,439	9,135	4,475	15,233	22,850	4	2	1	0
SAN DIEGO	1157	357	6	49,560	57,086	19,559	29,209	57,086	5	4	3	0
SAN DIEGO	1158	351	22	53,119	66,033	9,560	23,264	66,033	5	5	4	0
SAN DIEGO	1159	325	118	9,749	28,833	2,499	7,352	28,833	4	2	1	0
SAN DIEGO	1160	324	5	1,499	2,446	23,108	27,859	41,789	5	3	2	0
SAN DIEGO	1161	357	2	26,245	30,303	9,367	10,350	30,303	5	3	2	0
SAN DIEGO	1162	350	11	68,863	79,399	10,098	16,968	79,399	5	6	5	0

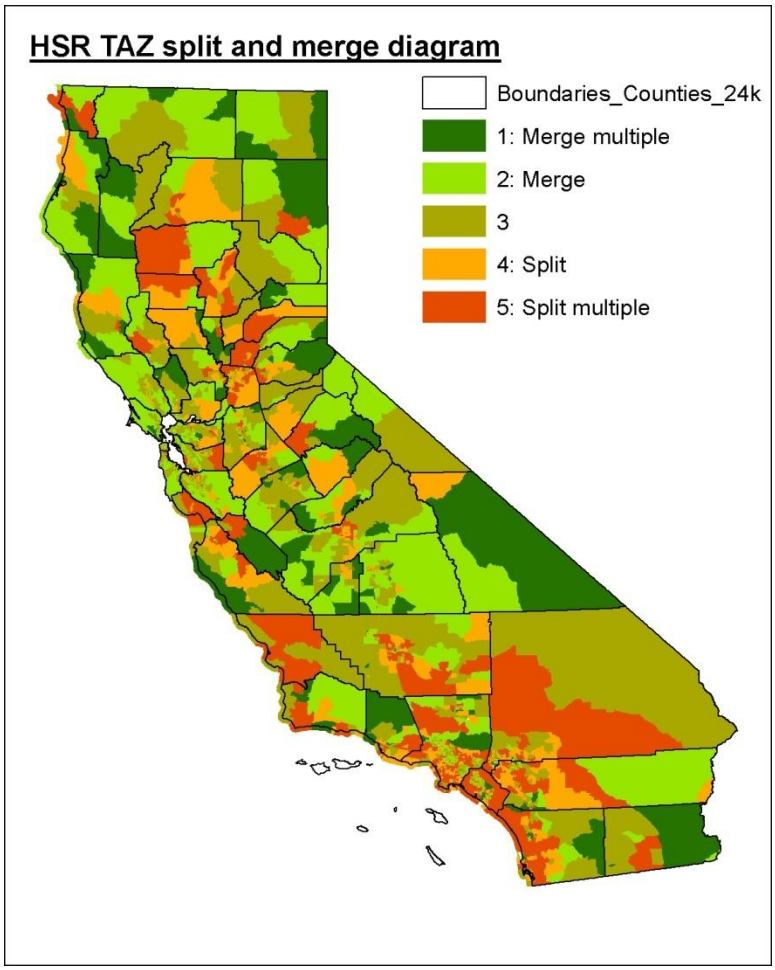


Figure 1: HSR TAZ Split-Merge Codes Displayed by Color

New TAZs were then assembled using year 2000 census block groups (as rereleased by the U.S. Census Bureau in 2008). The choice of blocks to group for TAZs was made based on characteristics of the local geography, transportation network, as well as year 2000 population and employment size. Year 2000 employment was obtained from the Census Transportation Planning Products (CTPP).

In areas that were fully developed, the population totals were allowed closer to 15,000 residents based on the 2000 census. In rural areas likely to be developed (e.g., adjacent to cities), the year 2000 TAZ populations were kept below the 15,000 threshold to allow for population increases. In rural areas and areas distant from high speed rail

stops, revisions to the HSR TAZs were based on accessibility to the road network and total projected population and employment.

The TAZs used in the HSR model appeared to be from two different sources. Some urban regions and areas surrounding proposed HSR stops were likely derived from local Metropolitan Planning Organizations (MPO) travel model TAZs. The rest of the state appears to have originated from the Statewide Travel Model (2004). In both cases, the HSR TAZs were individually reviewed and edited.

The regions that look as if they were from MPO model TAZs are all of the SCAG and MTC counties, downtown Sacramento, downtown Bakersfield, downtown San Diego, La Jolla and Escondido. In many of these areas, the TAZ system met the needs of the CSTDM, but in other areas they did not. In the San Diego region, the TAZs were too dense in the core of downtown, but very rapidly became too large as the distance from the core increased. In the San Francisco Bay Area, the TAZs were approximately well suited. In the Sacramento region the TAZs were slightly too dense in the downtown, but became far too large outside of the core downtown area. In the Los Angeles region, there were many zones that needed to be split. However, there were exceptions to these regional trends.

The areas of the state not mentioned above appear to have originated from the Statewide Travel Model (2004). Broadly stated, many of the TAZs were large in geographic area and varied between the extremes of very low populations and populations that warranted splitting the TAZ into multiple new zones. Regions that were already urban in character but that were based on the Statewide Travel Model (2004) TAZs required extensive splitting of TAZs – notably in Sacramento, the major cities of the San Joaquin Valley, and the suburban areas of San Diego.

2.2 External Zones

The external zones in the HSR network were largely unchanged with the exception of three new external zones, which were added to represent ports in Long Beach, Los

Angeles, and Oakland. The external zones used in the CSTDM are documented in Table 3. The existing external zones use node numbers 1 through 48, and nodes 49 to 51 are the new external zones.

Table 3: External Zones for the CSTDM.

TAZ	External District	Name
1	OR	United States Highway 101 - N. Smith River
2	OR	United States Highway 199 - N. Patrick Creek
3	OR	Indian Creek Road
4	OR	Interstate 5 - N. Hilt
5	OR	United States Highway 97 - N. Dorris
6	OR	State Highway 139 - N. Tulelake
7	OR	State Highway 50 - S.Main (Ore)
8	OR	Muldoon Rd / Westside Rd
9	OR	United States Highway 395 - N. New Pine Creek
10	NV_N	Surprise Valley Road
11	NV_N	State Highway 299 - E.Cedarville
12	NV_N	State Highway 81 - S.Eagleville
13	NV_N	Wendel Rd / Summers Rd E. Wendel / Herlong
14	NV_N	United States Highway 395 - N. Lemmon Valley (Nev)
15	NV_N	Henness Pass Road/ Sierra County Road S860
16	NV_N	Interstate 80 - W.Reno (Nev)
17	NV_N	State Highway 28 - E.Kings Beach
18	NV_N	United States Highway 50 - N. South Lake Tahoe
19	NV_N	State Highway 88 - N.Paynesville
20	NV_N	United States Highway 395 - S.Topaz Lake (Nev)
21	NV_N	State Highway 182 - N.Bridgeport
22	NV_N	State Highway 167 - E.Mono
23	NV_N	United States Highway 6 - N.Benton
24	NV_S	State Highway 266 - N.Oasis
25	NV_S	State Highway 266 - E.Oasis
26	NV_S	N Hwy (State Highway 267) - Death Valley
27	NV_S	Daylight Pass Rd (State Highway 374) - Death Valley
28	NV_S	State Highway 127 - N.Death Valley Junction
29	NV_S	Ash Meadows Rd - E.Death Valley Junction
30	NV_S	State Highway 178 - W.Pahrump (Nv)
31	NV_S	Interstate 15 - N.Mountain Pass
32	NV_S	Nipton Rd (State Highway 164) - E.Nipton
33	NV_S	United States Highway 95 - N.Arrowhead Junction

TAZ	External District	Name
34	NV_S	United States Highway 95 - E. Laughlin!
35	AZ	Interstate 40 - W.Topock (Nev)
36	AZ	Parker Dam (N. Parker)
37	AZ	State Highway 62 - Earp
38	AZ	Agnes Wilson Rd E. 95 (S. of Earp)
39	AZ	Interstate 10 - E.Blythe
40	AZ	E. Laguna Rd at Imperial Reservoir (Imperial Dam)
41	AZ	S. 4th Ave - Yuma
42	MEX	Highway 186 - Mexican Border Yuma
43	MEX	State Highway 111 - Calexico
44	MEX	State Highway 188 - Tecate
45	MEX	Interstate 5 - San Diego
46	AZ	Interstate 8 - Yuma
47	MEX	Collina del Sol (State Highway 905) - San Diego
48	MEX	State Highway 7 - Calexico
49	POLB	Long Beach (Port)
50	POLA	Los Angeles (Port)
51	POOak	Oakland (Port)

3. Implementation Process

Assembling the new TAZ system was completed as documented in the following discussion. The 2008 release of the 2000 census block groups was prepared by adding a joining year 2000 population count aggregated from the block level and employment from the CTPP package to each block group. The HSR TAZ system was also joined to this dataset by census block centroid. In most cases, this provided a good indication of the original HSR TAZ for each block group and eased the creation of TAZs when the HSR TAZs were acceptable. In conjunction with this dataset, a color coded version of the existing HSR TAZ system was used to visually identify which HSR TAZs needed to be split, merged, or left alone. A year 2004 road network was also used as a reference network.

Two custom tools were built in ArcGIS to assist in creating the new TAZ dataset. The first tool grouped selected block groups together and assigned them the TAZ number of an existing HSR TAZ for use when a TAZ was going to be kept intact. The second

grouped selected block groups and assigned them a temporary new TAZ identification number.

A standard working procedure included selecting a county and exporting the block groups for that county into a new dataset, on which all editing was conducted. Block groups were aggregated into TAZs using the procedure described as follows.

1. Existing Boundaries. County boundaries were always respected. All TAZs were aggregations of block groups with exceptions in only one area (see the note on exceptions below). Wherever possible, LUZ boundaries were maintained so that TAZs did not cross them.
2. Population and Employment Totals. As referenced above, the targets were 2030 populations below 15,000 and employment counts below 10,000. Since there were no projections by block group for 2030, the 2000 population and employment counts were used for reference. In existing developed areas, the population and employment totals from 2000 were allowed to approach the upper limits, but in less developed areas the population totals were generally kept below 10,000, and in areas that could expect large growth the totals were kept below 6,000. In several cases, individual block groups had population and/or employment totals that exceeded the targets. These block groups were assigned to a TAZ made up of only the single block group.
3. Access Methods to Road Network. Block groups were assembled into TAZs that had a common access point to the road network.
4. Land Use Homogeneity. Where possible, TAZs had land uses that were homogeneous. In most cases, this goal was achieved naturally by selecting similar adjacent block groups.
5. Regular Shape of Resulting TAZ. When possible, TAZs were assembled to create a regular shape that avoided having a length to width ratio greater than 3 to1. Block group geographies did not always permit this guideline to be followed.

Following a first round of TAZ creation for the entire state, a review session was held with both the HBA Specto and ULTRANS research teams. As a result of these

discussions, some TAZ boundaries were redefined to better match probable growth areas and to remedy some irregular TAZ boundary issues. The final changes to the TAZ system occurred following a California PECAS Peer Advisory group meeting that included a minor modification to the boundary between Merced and Fresno Counties, which reflected the annexation of a small portion of rural Fresno County into Merced County. At the same time the TAZs in the El Dorado County portion of the Lake Tahoe Basin were adjusted to allow the Tahoe Basin to be separated into a distinct set of LUZs.

LUZ boundaries were defined through matching the TAZs to the LUZ boundaries used in the demonstration version of the California PECAS model. In locations where maintaining the integrity of the earlier LUZ system was not possible, TAZs were assigned to the closest possible LUZ based on the TAZ's geographic centroid. The Tahoe Basin was assigned to a set of LUZs to represent the jurisdictional boundary of the Tahoe Regional Planning Agency.

3.1 Numbering Systems

HBA Specto proposed a district system for assigning the TAZs identification numbers (IDs). It was decided that the thousands and hundredths digits of each TAZs ID would be representative of the district it is found in. TAZs were assigned to districts based on their geographic location and with the intention that there be a few potential IDs remaining in each district’s number block to allow for the potential of adding new TAZs. Within each district the tens and ones digit of each TAZ was taken from the Object ID field in the GIS dataset. This did not result in a logical numbering sequence for TAZs, but all TAZs are members of districts that each have less than 100 TAZs and are grouped geographically. These districts are intended to ease locating TAZs for transportation modeling purposes and do not explicitly follow LUZ boundaries.

LUZ IDs were assigned through ordering county centroids north to south, assigning a sequential number to the county, and numbering LUZs within the county based on the order in their centroids going from south to north.

3.2 TAZs Exceeding Recommended Targets

In the initial TAZ development, several TAZs emerged that were comprised of single block groups with either population, employment, or both totals exceeding the recommended targets.

After reviewing these TAZs, all of the TAZs that exceeded the totals based solely on population were left unchanged. In almost all cases, these TAZs exceeded the recommended values only slightly, and had either significant geographic constraints, which made splitting the TAZ unlikely to improve the fidelity of the model, or had other effects that made it logical to consider the area as a single TAZ. TAZs between 10,000 and 25,000 employees were left as individual block groups to ease the development of demographic data. All TAZs that had employment exceeding 25,000 employees were individually reviewed and addressed specifically. These were separated into sub-block group level geographies that were aggregations of year 2000 census blocks. Employment totals for the blocks were assumed based on year 2002 OnTheMap 4 synthetic employment data. The employment totals for the block assemblages were targeted at between 10,000 and 15,000, but sometimes totals above or below these targets were accepted due to geographic constraints.

4. Results

Work on delineating the TAZs from block groups began on August 11, 2009, a preliminary review was conducted on August 26, 2009, and a formal review was held on September 2, 2009. The TAZs were provisionally accepted following edits proposed on September 2nd and minor edits that became necessary following the removal of large water bodies from the area covered by TAZs during the LUZ assignment process the first week of October 2009. Further changes to the TAZ system were prompted by the CalPECAS Peer Advisory Meeting November 18, 2009, and were fully implemented on December 7, 2009. The final changes were made in the last week of January 2010, which included dividing some very large employment TAZs into smaller TAZs.

LUZs were initially assigned in the first week of October 2009, and final adjustments, which included the separation of the Lake Tahoe Basin into distinct LUZs, were completed on December 7, 2009.

The resulting TAZ/LUZ system has 5,191 TAZs and 526 LUZs. For a summary of the data please see Table 4 below. There are 31 TAZs with employment totals below 100, and 146 with employment totals above 10,000. The oversized TAZs are almost entirely single block groups in high density downtown areas, industrial parks, airports, prisons, or military bases (Figure 2). Many of the largest single block group TAZs were split into smaller TAZs using the year 2000 census block boundaries.

Table 4: Distribution of TAZ by Population and Employment Size

Total number of TAZs	5,191
TAZs with 2000 Pop below 1,000	294
TAZs with 2000 Pop above 10,000	734
TAZs with 2000 Pop above 15,000	3
Mean TAZ Population	6,524
Minimum TAZ Population	0
Standard Deviation Population	3,150
Maximum TAZ Population	36,146
TAZs with Employment below 100	31
TAZs with Employment above 10,000	146
TAZs with Employment above 15,000	55
TAZs with Employment above 25,000	2
Mean TAZ Employment	2,749
Minimum TAZ Employment	0
Maximum TAZ Employment	34,645
Standard Deviation Employment	3,060

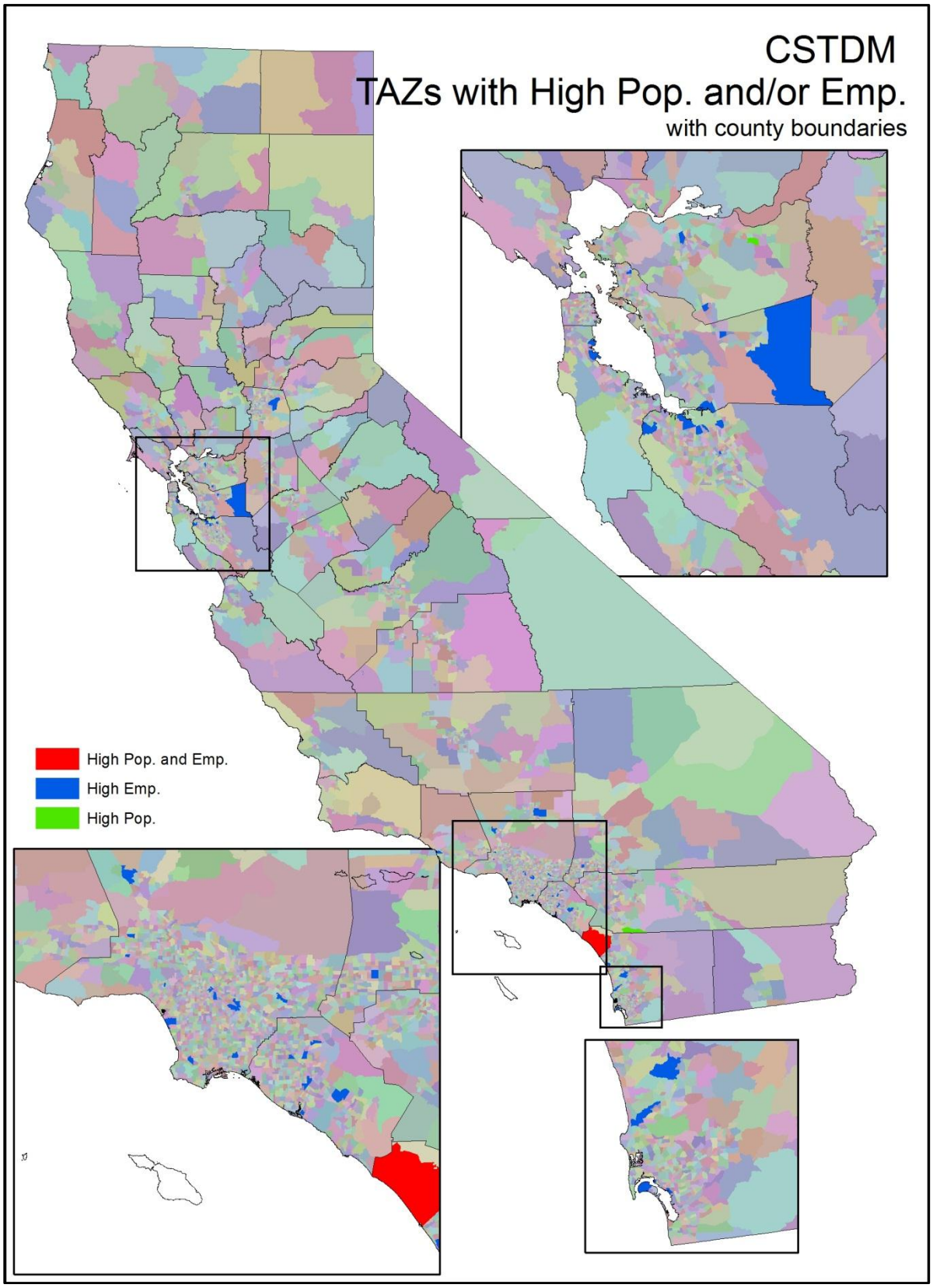


Figure 2: TAZs with Population > 15,000 and/or Employment > 10,000

There are also a few special cases. In Riverside county six block groups were split into multiple TAZs by census block. The six block groups have multiple areas/polygons that are not contiguous and for which grouping all of the sections of the block group into a single TAZ would have severely violated other guidelines for TAZ creation. These disjointed block groups are a result of the 1862 and 1884 Pacific Railroad Acts land grants. Population was assigned to the sections of the block groups through totaling the population by census blocks, and the ratio of section population to block group population was used to assign employment to each section.

Another special case is block group 060730099029 in San Diego Bay, which contained almost zero land area with very low population (63) and moderate employment (685). This block group quite clearly excluded docks and other land areas.

The resulting TAZ system, with 5191 zones, has the broad properties reported in Table 4 and Figure 3. The majority of the TAZs fall within our guidelines for population size, employment totals, shape, and access routes with a few outliers as described above. Because of the large number of TAZs, they are further grouped into 70 Districts (Table 5, Figure 4) for easily identifying where in the state any particular TAZ may be found. The TAZs are also aggregated into 526 Land Use Zones (LUZ) for integration with the California PECAS model (see Figure 5).

Table 5: List of Districts

District	Number of TAZs	Low TAZ	High TAZ	Region	HSR Region	Superdistrict	District
0				Externals	Externals	Externals	Externals
1	74	100	173	North	North	Far North	Del Norte, Humboldt, Trinity, Siskiyou, Shasta, Modoc, Lassen counties
2	85	200	284	North	North	Mid North	Mendocino, Lake, Glenn, Tehama, Colusa, Butte, Plumas, Sierra, Nevada counties
3	48	300	347	SACOG	SACOG	North SACOG	Sutter, Yolo and Yuba counties
4	66	400	465	SACOG	SACOG	North SACOG	Placer and El Dorado counties
5	46	500	545	SACOG	SACOG	Sacramento	Within ~ 5 mi of HSR station (leave room for extra zones)
6	78	600	677	SACOG	SACOG	Sacramento	Sacramento County North of American River
7	74	700	773	SACOG	SACOG	Sacramento	Sacramento County South of American River
8	98	800	897	ABAG	ABAG	North Bay	Northern Sonoma, Napa and Solano counties
9	100	900	999	ABAG	ABAG	North Bay	Marin and Southern Sonoma counties
10	72	1000	1071	ABAG	ABAG	Peninsula	Central San Francisco
11	90	1100	1189	ABAG	ABAG	Peninsula	Rest of SF County and northern San Mateo County
12	93	1200	1292	ABAG	ABAG	Peninsula	Southern portion of San Mateo County
13	92	1300	1391	ABAG	ABAG	East Bay	Eastern Contra Costa County
14	82	1400	1481	ABAG	ABAG	East Bay	Oakland
15	85	1500	1584	ABAG	ABAG	East Bay	West Contra Costa and north Alameda
16	100	1600	1699*	ABAG	ABAG	East Bay	South Alameda County (San Lorenzo,

District	Number of TAZs	Low TAZ	High TAZ	Region	HSR Region	Superdistrict	District
							Hayward, Fremont)
17	99	1700	1798	ABAG	ABAG	South Bay	Northwest Santa Clara (Palo Alto, Sunnyvale, Mountain View)
18	78	1800	1877	ABAG	ABAG	South Bay	Central San Jose
19	98	1900	1997	ABAG	ABAG	South Bay	South and West Santa Clara County
20	47	2000	2046	Central Valley	San Joaquin	San Joaquin	Stockton urban area
21	63	2100	2162	Central Valley	San Joaquin	San Joaquin	Rest of San Joaquin county
22	76	2200	2275	Central Valley	Stanislaus	Stanislaus	Stanislaus county
23	41	2300	2340	Central Valley	Merced	Merced	Merced county
24	96	2400	2495	Central Valley	Fresno Madera	Fresno Madera	Fresno urban area
25	68	2500	2567	Central Valley	Fresno Madera	Fresno Madera	Rest of Fresno county and Madera county
26	27	2600	2626	Central Valley	Kings/Tulare	Kings/Tulare	Kings county
27	79	2700	2778	Central Valley	Kings/Tulare	Kings/Tulare	Tulare county
28	88	2800	2887	Central Valley	Kern	Kern	Bakersfield urban area
29	50	2900	2949	Central Valley	Kern	Kern	Rest of Kern county
30	35	3000	3034	Sierras	Sierras	Sierras	Alpine, Amador, Calaveras, Inyo, Mariposa, Mono and Tuolumne counties
31	42	3100	3141	Coast	Coast	North Coast	Santa Cruz and San Benito counties
32	64	3200	3264	Coast	Coast	North Coast	Monterey county
33	88	3300	3388	Coast	Coast	South Coast	San Luis Obispo and Santa Barbara counties
34	62	3400	3462	SCAG	SCAG	Ventura	North and west Ventura county (inc. Ventura and Oxnard)
35	42	3500	3541	SCAG	SCAG	Ventura	East Ventura county (Thousand Oaks, Simi

District	Number of TAZs	Low TAZ	High TAZ	Region	HSR Region	Superdistrict	District
							Valley)
36	90	3600	3697	SCAG	SCAG	LA North	Lancaster, Palmdale and San Gabriel mountains
37	74	3700	3795	SCAG	SCAG	LA North	North San Fernando valley
38	95	3800	3895	SCAG	SCAG	LA North	South San Fernando valley and Santa Monica mountains
39	96	3900	3997	SCAG	SCAG	LA North	Burbank and Glendale area
40	82	4000	4081	SCAG	SCAG	LA Central	Downtown Los Angeles (inside freeway loop; room for more zones)
41	99	4100	4198	SCAG	SCAG	LA Central	East LA (Monterey Park, Montebello, Commerce)
42	96	4200	4296	SCAG	SCAG	LA Central	South Central area (Compton, Lynwood)
43	71	4300	4372+	SCAG	SCAG	LA Westside	West Central LA inc Beverly Hills, Hollywood
44	72	4400	4471	SCAG	SCAG	LA Westside	West LA inc Santa Monica, Culver City
45	78	4500	4577	SCAG	SCAG	LA South Bay	Inglewood, Lawndale, Manhattan Beach area
46	77	4600	4676	SCAG	SCAG	LA South Bay	Torrance, Lomita area
47	80	4700	4779	SCAG	SCAG	LA Gateway	Long Beach / Seal Beach area
48	95	4800	4899	SCAG	SCAG	LA Gateway	Lakewood, Norwalk, Whittier
49	97	4900	4996	SCAG	SCAG	LA San Gabriel	Pasadena area
50	99	5000	5099	SCAG	SCAG	LA San Gabriel	San Gabriel Valley (Baldwin park, Covina, San Dimas)
51	80	5100	5179	SCAG	SCAG	San Bernadino	North and East San Bernadino
52	90	5200	5289	SCAG	SCAG	San Bernadino	West San Bernadino (Ontario area incl.

District	Number of TAZs	Low TAZ	High TAZ	Region	HSR Region	Superdistrict	District
							industrial areas around I-15 Ontario freeway)
53	97	5300	5399	SCAG	SCAG	San Bernadino	Central San Bernadino (Fontana, San Bernadino, Redlands)
54	95	5400	5496	SCAG	SCAG	Riverside / Imperial	Northwest Riverside county (Riverside, Norco)
55	98	5500	5598	SCAG	SCAG	Riverside / Imperial	Southwest Riverside county (Temecula, Perris, San Jacinto)
56	87	5600	5688	SCAG	SCAG	Riverside / Imperial	Coachella valley (Palm Springs, Indio) and east; Imperial county
57	94	5700	5799*+	SCAG	SCAG	Orange	North Orange County (Orange, Anaheim, Fullerton, Yorba Linda)
58	97	5800	5899	SCAG	SCAG	Orange	West Orange County (SW of Santa Ana Freeway & Santa Ana River; Buena Park, Huntington Beach, Westminster)
59	107	5900	5999*	SCAG	SCAG	Orange	Central Orange county (Newport Beach, Irvine, Santa Ana, Tustin)
60	56	6000	6055	SCAG	SCAG	Orange	East Orange county (Mission Viejo, San Clemente, Laguna Hills)
61	37	6100	6136	SANDAG	SANDAG	San Diego	East county (mountain areas)
62	97	6200	6296	SANDAG	SANDAG	San Diego	North Coast (Oceanside-Escondido corridor)
63	106	6300	6399*	SANDAG	SANDAG	San Diego	Central coast (University City, La Jolla, Miramar)
64	72	6400	6472	SANDAG	SANDAG	San Diego	Central and north

District	Number of TAZs	Low TAZ	High TAZ	Region	HSR Region	Superdistrict	District
							San Diego
65	87	6500	6586	SANDAG	SANDAG	San Diego	Eastern San Diego suburbs (Lemon Grove, El Cajon)
66	63	6601	6665	SANDAG	SANDAG	San Diego	South San Diego suburbs (Chula Vista, Imperial Beach)
67	1	6700	6700	ABAG	ABAG	Overflow	Bay Area overflow
68	8	6800	6807	SCAG	SCAG	Overflow	SCAG overflow
69	6	6900	6905	Other	Other	Overflow	Overflow for remainder of state

*: The district's TAZ range extends into one of the overflow districts 67-69

+: TAZ numbers are not consecutive due to TAZs that were created and numbered but retired from use because they did not meet minimum criteria.

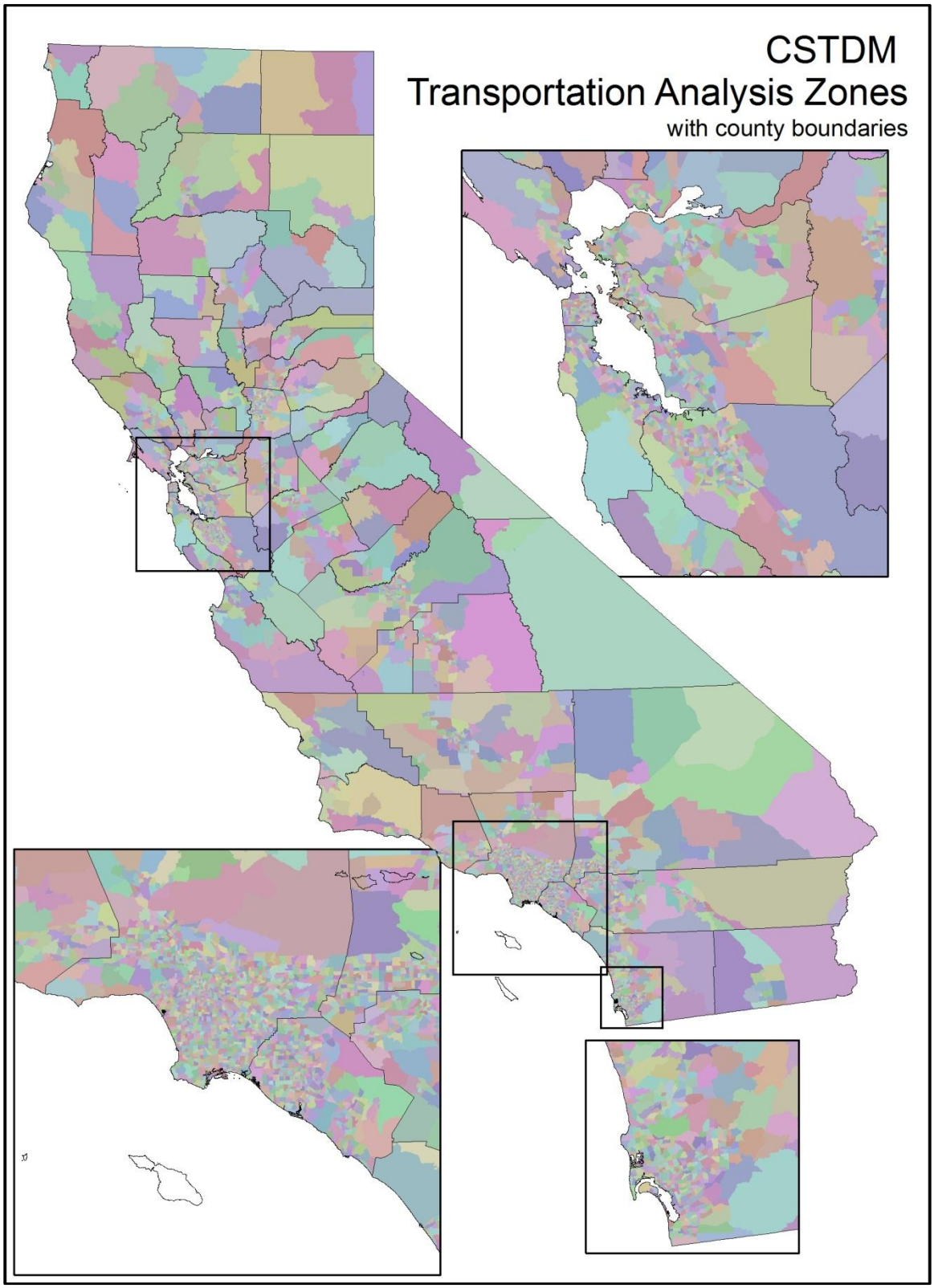


Figure 3: Transportation Analysis Zones

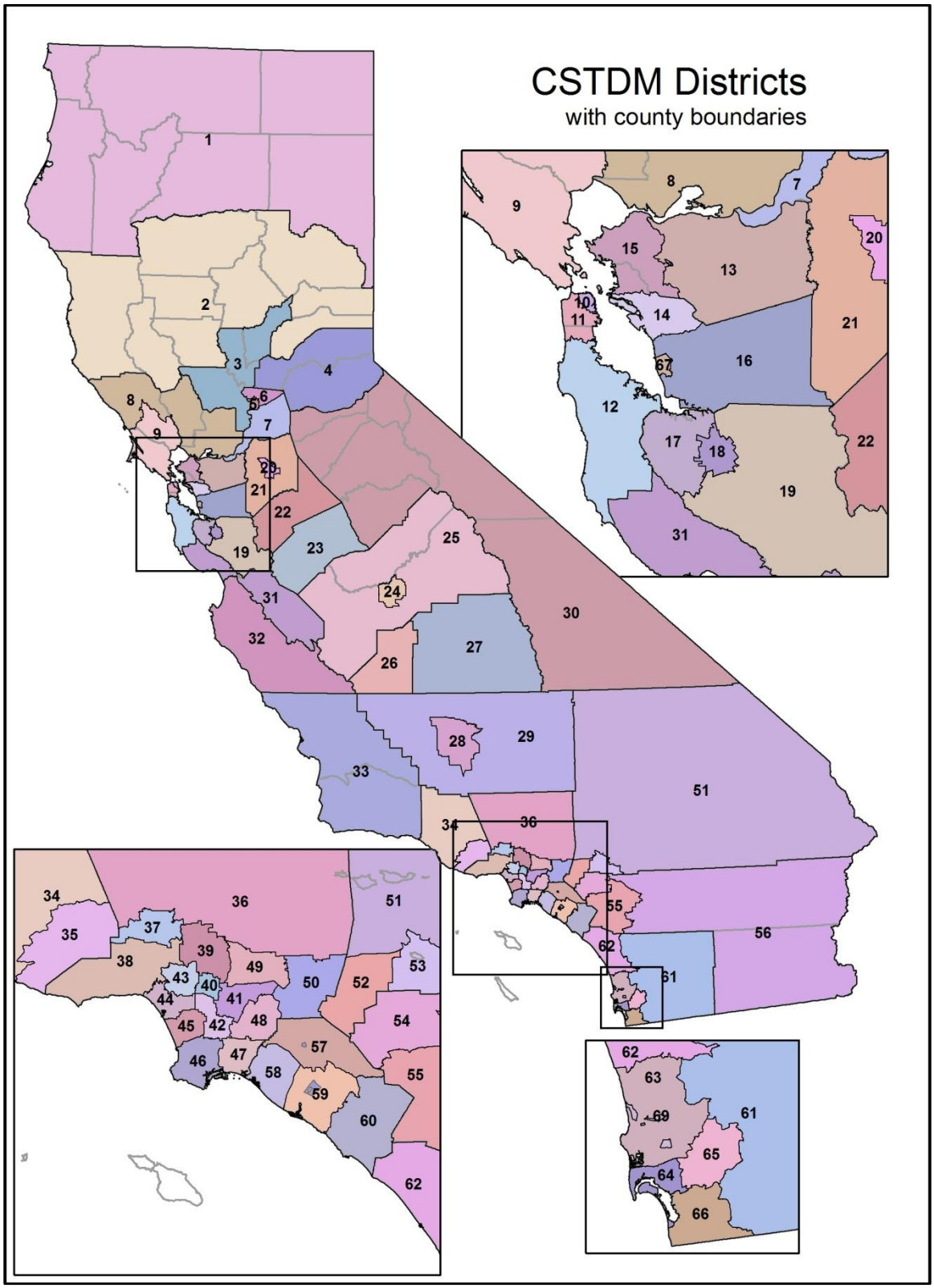


Figure 4: TAZ Districts

5. File Names and Descriptions

All files are located in the GIS directory within the installation of CSTDM:

CSTDM2009_TAZ.shp: TAZ, LUZ, District, County. Fields with a summation of CorrPop are based on block population sums for 2000 and CorrEmp are based either on the 2000 CTPP or the 2002 OnTheMap4 data where block groups were divided. **CSTDM2009_blocks_withTAZ.shp**: These are the 2008 rerelease of the 2000 US Census Blocks with a listing of which TAZ and LUZ they belong to.

CSTDM2009_DIST: District boundaries.

CSTDM2009_TAZCentroids.shp: This includes centroids extracted from the CSTDM TAZ system. These are the centroids used within the transportation model, not geographic centroids of the TAZ polygons.

B. CSTDM Future Year TAZ Documentation

California Statewide Travel Demand Model

Development of Future Population and Employment

Draft Documentation

ULTRANS
Institute of Transportation Studies,
UC Davis
Davis, CA

In association with
HBA Specto, Incorporated
Calgary, Alberta

January, 2012

Table of Contents

Tables **4**

Figures **5**

1. Introduction **6**

2. MPO and RTPA Forecasts for 2020 and 2035 ..... **6**

2.1 Spatial Correspondence **7**

2.1.1 Identification of Nested TAZ **8**

2.1.2 Splitting Non-Nested TAZ **12**

2.1.3 Aggregating Forecast Data to CSTDM TAZ **15**

2.1.4 CSTDM TAZ Splitting for the TAZs that Exceeded Maximum Thresholds **17**

2.2 Employment by Industry Correspondence **21**

2.3 Population from Households **23**

2.4 Time Horizon Correspondence **25**

2.4.1 County-level Population Interpolation and Extrapolation **26**

2.4.2 Distributing County Population to TAZ **27**

2.4.3 Interpolating TAZ Employment by Industry **27**

3. California Areas without Local (MPO/RTPA) Forecasts for 2020 and 2035 ..... **28**

3.1 Population **29**

3.2 Employment **31**

4. Forecasts for 2050 ..... **36**

4.1 County-Level Population and Employment Totals for 2050 ..... **36**

4.2 Allocation of 2050 Population and Employment to CSTDM TAZ **37**

4.3 Employment Industry Categories **39**

5. Final Population and Employment **39**

5.1 Population Synthesis **39**

5.1.1 Creation of Population Synthesizer Inputs **41**

5.1.2 Running Population Synthesizer **44**

5.1.3 Combining Regional Synthetic Populations **44**

5.2 Employment by Occupation **45**

References **56**

Appendix A: Reference List of Forecast Data Files and Reports (MPO/RTPA) .. **58**

..... Error! Bookmark not defined.

Appendix B: List of Forecast Data by Attributes and Time Horizon Years Provided by California MPOs and RTPAs **59**

Appendix C: Python Scrip for Aggregating Primary Forecast Data to CSTDM TAZ **63**

Appendix D: Python Scrip for Aggregating Primary Forecast Data to CSTDM TAZ **70**

Appendix E: Employment Industry Category Cross-Classification **76**

Appendix F: California Department of Finance Forecasts **84**

Interim Projections **84**

Projection Distributions **86**

Appendix G: Moody's and TREDIS Forecasts **87**

Appendix H: Employment Floorspace by Industry **91**

Appendix I: Projections Household by Size, Income, Auto Ownership, and Worker Categories **104**

Tables

Table 1: Number and Percentage of Nesting and Non-Nesting TAZs by MPO/RTPA .. 10
Table 2: Explanation of the Split Type Codes 13
Table 3: Counties where CSTDM TAZs were Split 20
Table 4: CSTDM Industry Categories..... 21
Table 5: Household Size Correspondence – US Census and CSTDM 24
Table 6: Time Horizons of Forecasts Provided by MPOs/RTPAs 25
Table 7: Non-Residential Floorspace Categories..... 33
Table 8: Crosswalk for CSTDM and California PECAS Framework Space Types 34
Table 9: Share of Non-Residential Floorspace by CSTDM Industry Category by TAZ for Calaveras County with Six TAZs..... 35
Table 10: Sample Targets File (showing a subset of typical columns)..... 42
Table 11: Weight Multipliers Example 43
Table 12: Statewide Distribution Employment by Industry and Occupation 46
Table 13: Employment Intensity Types 48

Figures

Figure 1: Example of MPO/RTPA TAZ Nesting and Not Nesting within CSTDM TAZ 8
Figure 2: Overview of Geo-processing Tool Use for Spatial Correspondence 11
Figure 3: Example of Working Directory Structure and the Spatial Correspondence Geo-processing Tool 11
Figure 4: Examples of the 3 Split Type Codes 13
Figure 5: An Example of a CSTDM TAZ that was Split due to High Employment in 2035 19
Figure 6: California Areas without Local (MPO/RTPA) Forecasts 29
Figure 7: Stages to Create the Synthetic Population 40
Figure 8: Samples File 44
Figure 9: Proportion of Office Support Workers by Industry 47
Figure 10: Employment Intensity Types (Sacramento Area) 49
Figure 11: Employment Intensity Types (Central San Joaquin Area) 49
Figure 12: Employment Intensity Types (Los Angeles Area) 50
Figure 13: Intensity Classification Areas for Intensity Type 8 52
Figure 14: Intensity Classification Areas for Intensity Type -3 53

1. Introduction

This report documents the preparation of the future population and employment data developed as inputs to the California Statewide Travel Demand Model (CSTDM) for the years 2020, 2035, and 2050. Complete documentation of this model can be found at http://ultrans.its.ucdavis.edu/doc/cstdm-documentation. The CSTDM was developed by the Urban Land Use and Transportation Center (ULTRANS) at the University of California Davis, in association with HBA Specto, for the California Department of Transportation (Caltrans). The model was designed to forecast travel between major regions in California and to provide external travel forecasts for the regional models operated by Metropolitan Planning Organizations (MPOs). As a result, the primary source for future population and employment data was obtained from California's MPOs and, when possible, from Rural Transportation Planning Agencies (RTPAs). This data was largely supplemented, when necessary, with forecasts from the California Department of Finance.

2. MPO and RTPA Forecasts for 2020 and 2035

Requests for future population and employment forecasts were made via E-mails and phone communications from ULTRANS staffs to MPO staff and RTPA staff from the period of May to August 2011. The following MPOs provided forecast data by geographic units of analysis known as Traffic Analysis Zone (TAZ) that are used in local travel demand models: Association of Monterey Bay Area Governments (AMBAG), Butte County Association of Governments (BCAG), Fresno Council of Governments (FCOG), Kings County Association of Governments (KCAG), Kern Council of Governments (KCOG), Merced County Association of Governments (MCAG), Madera County Transportation Commission (MCTC), Metropolitan Transportation Commission (MTC), Sacramento Area Council of Governments (SACOG), San Diego Association of Governments (SANDAG), San Joaquin Council of Governments (SJCOG), San Luis Obispo Council of Governments (SLOCOG), Santa Barbara County Association of Governments (SBCAG), Shasta County Regional Transportation Planning Agency (SCRTPA), Southern California Association of Governments (SCAG), Stanislaus Council of Governments (StanCOG), and Tulare County Association of Governments

(TCAG). The following RTPAs provided forecast data by TAZ: Amador CTC (County Transportation Commission), Tuolumne CCAPC (County/Cities Area Planning Council), Lake and Mendocino. The forecasts for Lake and Mendocino were obtained from the Wine County Travel Demand Model (WC-TDM), which was developed by Parsons Brinckerhoff and submitted to Caltrans and Wine County Inter-Regional Partnership in February of 2011. This model covered four counties: Lake, Napa, Sonoma, and Mendocino. MTC had already provided us with data for Napa and Sonoma counties, so we only utilized the data for Lake and Mendocino counties. In total all 17 MPOs and four RTPAs were able to provide their latest population and employment forecasts as of August 2011 to be used in this project. See Appendix A for a reference list of forecast data files and reports provided by MPOs and RTPAs. See Appendix B for a list of the forecast data by attributes and time horizon years provided by the MPOs and RTPAs.

2.1 Spatial Correspondence

In this section, the methods and data used to develop a correspondence between the MPO/RTPA TAZ and the CSTDM TAZ will be explained. Population and employment forecasts were available from the MPOs/RTPAs by the geographic TAZ that are used by their local travel demand models. We here describe how population and employment forecasts provided by TAZ from MPO/RTPA models were nested and/or aggregated into CSTDM TAZs.

Since population and employment forecast data from MPOs/RTPAs were generally provided on a smaller spatial scale than the CSTDM uses, a methodology was created to aggregate their data into the CSTDM TAZ system. In Section 2.1.1, the method of identifying MPO/RTPA TAZs that nest or do not nest within CSTDM is outlined. Figure 1 illustrates nesting (orange) and non-nesting (green) TAZ. In Section 2.1.2, the method of splitting non-nesting TAZs into CSTDM TAZ is described. In section 2.1.3, the process of calculating final forecast totals from the MPO/RTPA TAZs to the CSTDM TAZs is described.

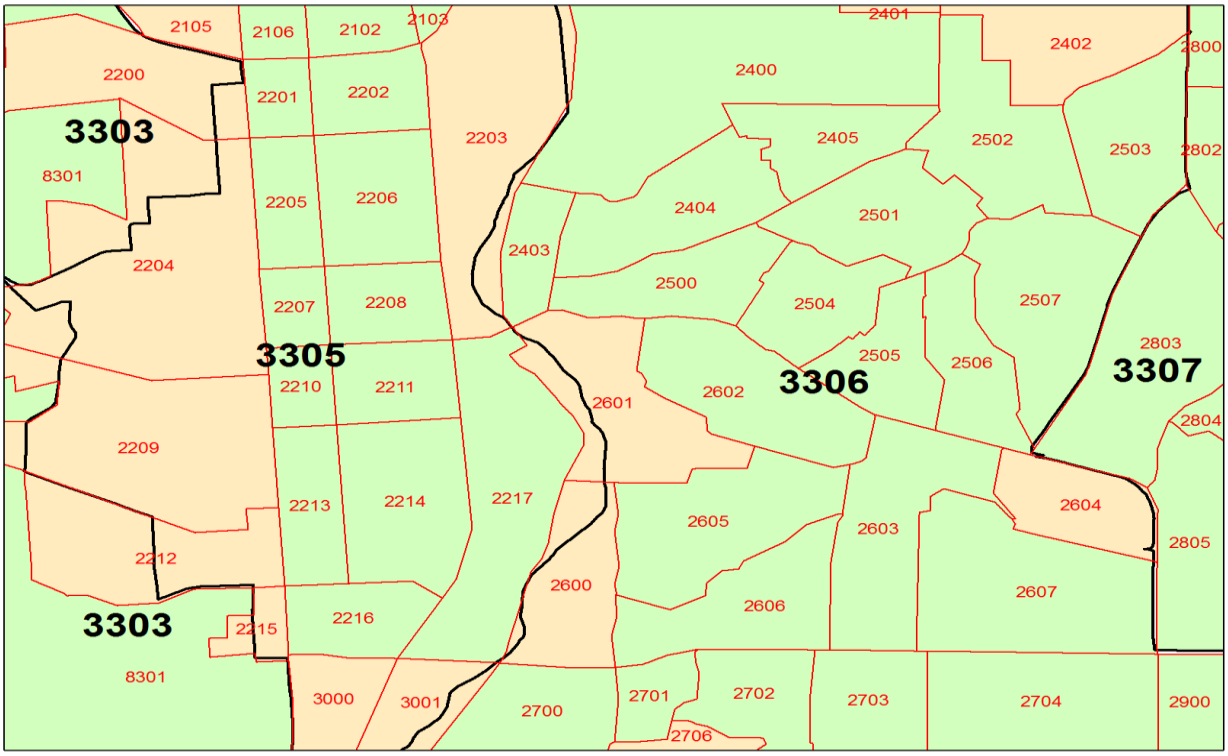


Figure 1: Example of MPO/RTPA TAZ Nesting and Not Nesting within CSTDM TAZ

(CSTDM TAZ boundaries are shown in black lines and MPO/RTPA TAZ boundaries are shown in red lines. Nesting MPO/RTPA TAZs are green and non-nesting ones TAZ are orange.)

2.1.1 Identification of Nested TAZ

A geo-processing tool was developed to identify the MPO/RTPA TAZs that do not nest completely within the CSTDM TAZ. This tool was run separately for each MPO or RTPA. The MPO/RTPA TAZ and CSTDM TAZ layers are required inputs to this tool. Additionally, the tool accepts a Protected Areas layer (California Protected Areas Database 1.3, ©August 2009 GreenInfo Network - www.calands.org) as another input. This layer was clipped to the extent of MPO or RTPA and then copied to the output directory for reference during the manual splitting step (see Section 2.1.2). The main outputs of this tool are two geo-database feature classes representing nesting and non-nesting MPO/RTPA TAZ. The non-nesting TAZ are split along CSTDM TAZ boundaries, and have a field identifying their proportion, by area, of the original MPO/RTPA TAZ.

After the input data is specified, the first processing step is to intersect the MPO/RTPA TAZs and CSTDM TAZs, to determine their minimum common extent. Their common extent is then used to clip the input MPO/RTPA TAZs and CSTDM TAZs. This eliminates slivers along the edges due to slightly different boundaries and discrepancies between the TAZ systems, such as bodies of water which were included in one system, but excluded in the other. It also ensures that the split percentages of each MPO zone will add up to 100 percent.

Next, fields are added to the clipped layers to contain the area of each TAZ, which is then calculated. The MPO/RTPA layer and CSTDM layer are then overlaid in a “Union” operation, to effectively split the MPO/RTPA TAZs along the CSTDM TAZ boundaries. At this point, the proportion of the total MPO/RTPA TAZ can be calculated for each polygon. The proportion is calculated by dividing the area of the split polygon by the area of the entire MPO/RTPA TAZ. The polygons that represent at least 98 percent (≥ 0.98) of the original MPO/RTPA TAZ zone are considered to nest within that CSTDM TAZ (see Table 1). The 98 percent threshold was selected so that small slivers along a common boundary would not prevent a polygon from being identified as nesting.

Table 1: Number and Percentage of Nesting and Non-Nesting TAZs by MPO/RTPA

MPO/RTPA	Total # of TAZs	# nesting	# non-nesting	% nesting	% non-nesting
ACTC (Amador)	208	197	11	94.7%	5.3%
AMBAG	1682	1433	249	85.2%	14.8%
BCAG	912	794	118	87.1%	12.9%
Fresno	1571	1176	395	74.9%	25.1%
Kern	1692	1429	263	84.5%	15.5%
Kings	646	492	154	76.2%	23.8%
Madera	414	286	128	69.1%	30.9%
Merced	521	296	225	56.8%	43.2%
MTC	1454	558	896	38.4%	61.6%
SACOG	1502	1279	223	85.2%	14.8%
San Joaquin	786	509	277	64.8%	35.2%
SANDAG	4670	3391	1279	72.6%	27.4%
SBCAG	268	200	68	74.6%	25.4%
SCAG	11267	10276	991	91.2%	8.8%
Shasta	900	773	127	85.9%	14.1%
SLOCOG	1207	970	237	80.4%	19.6%
Stanislaus	2242	1711	531	76.3%	23.7%
Tulare	1332	997	335	74.8%	25.2%
Tuolumne	507	373	134	73.6%	26.4%
Lake & Mendocino	992	865	127	87.2%	12.8%
TOTAL:	**34773**	**28005**	**6768**	**80.5%**	**19.5%**

MPO/RTPA TAZs that do not have a split polygon representing at least 98 percent of the original area are considered non-nesting. All of the polygons for these zones are copied to a new file, which is used to manually split zone attributes (see section 2.1.2). This file has fields added to the attribute table, which are specific to the categories of data available from each MPO/RTPA. The inputs, processing steps, and outputs are summarized in Figure 2.

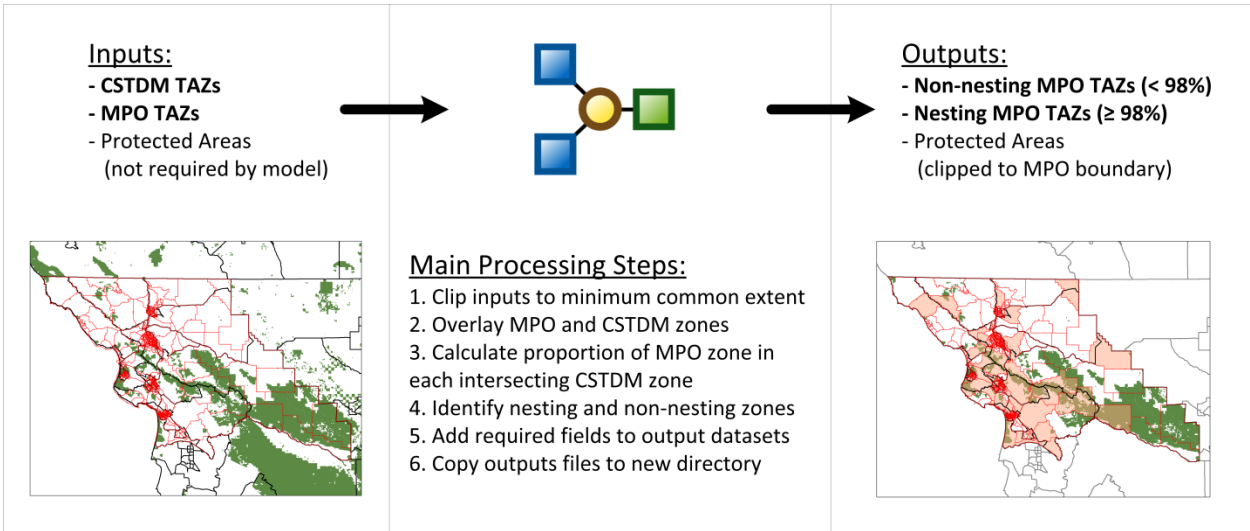


Figure 2: Overview of Geo-processing Tool Use for Spatial Correspondence

This geo-processing tool was then applied to each set of MPO/RTPA TAZ data (see Figure 3), with minor modifications to accommodate the set of available attributes. The geo-processing tool (called 'Zonal Properties Model' in Figure 3) is applied to each set of MPO/RTPA TAZ data, to identify nesting and non-nesting TAZs.

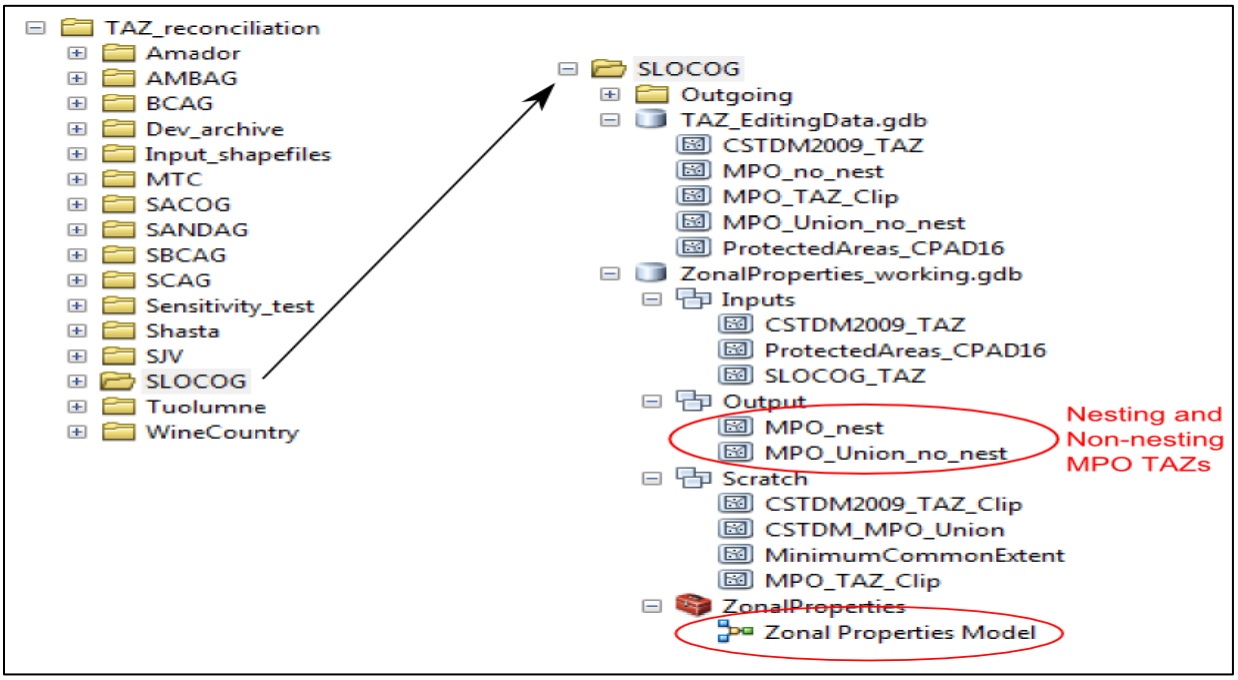


Figure 3: Example of Working Directory Structure and the Spatial Correspondence Geo-processing Tool

2.1.2 Splitting Non-Nested TAZ

Each MPO/RTPA TAZ that did not nest within a CSTDM TAZ was manually reviewed so that population and employment could be properly assigned to the CSTDM TAZs in which it fell. A process was developed to manually split the MPO/RTPA population and employment totals to CSTDM TAZs. It began by adding the “MPO_Union_no_nest” output from the geo-processing tool described in section 2.1.1 (see Figure 3) to ArcMap. The ESRI World Imagery Map and the Protected Areas layer (clipped by the geo-processing tool) were also added to ArcMap for reference. The World Imagery Map allowed the editor to see the location of current population and employment locations. It also aided in the identification of the likely location of future population and employment. The Protected Areas layer was used to identify areas that are unlikely to be developed in future years.

Once the reference layers were added to ArcMap, the MPO/RTPA data categories (population, employment, and school enrollment, if available) were ready to be split. For each data category the MPO/RTPA provided by their TAZ geographies, the geo-processing tool created four fields in the “MPO_Union_no_nest” shapefile. These four fields were: Split Type Base Year, Split Percentage Base Year, Split Type Future Years, and Split Percentage Future Years. Each non-nesting MPO/RTPA TAZ was edited sequentially. Once the editor had selected a MPO/RTPA TAZ, they determined what type of split was appropriate for both the base (2008) and the future years. The split type codes are described in Table 2 and shown in Figure 4.

The MPO/RTPA TAZ is highlighted in blue; the CSTDM TAZ boundary is yellow. All of the base and future population and employees should be assigned to the largest section of the MPO/RTPA TAZ -- on the right. Therefore, the "A" Split Type code is used for both the base year and the future years. The editor does not need to fill in the Split Percentage fields for the base year or future year because a post-processing script will recognize the "A" Split Type code and assign the largest section 100% of the base and future population and employees.

The MPO/RTPA TAZ is highlighted in blue; the CSTDM TAZ boundary is pink. Since both sections of the MPO/RTPA TAZ are uniformly developed, all of the base and future population and employees should be assigned to both sections based on the proportion of the MPO/RTPA zone that is in each CSTDM TAZ. Therefore, the "P" Split Type code is used for both the base year and the future years. The editor does not need to fill in the Split Percentage fields for the base year or future year because a post-processing script will recognize the "P" Split Type code and assign percentages to each section based on the proportion of the MPO/RTPA TAZ that is in each CSTDM TAZ.

The MPO/RTPA TAZ is highlighted in blue; the CSTDM TAZ boundary is yellow. The right section on the MPO/RTPA TAZ will have the majority of the population and employment in the base year, the left section will have the majority of the future population and employment. Therefore, the "D" Split Type code is used for both the base year and future years. With the "D" Split type code the editor needs to fill in the Split Percentage fields. For the base year this would be 10%/90% (left section/right section); for the future years this would be 99%/1%.

Figure 4: Examples of the 3 Split Type Codes

Table 2: Explanation of the Split Type Codes

Code	Description	Split Method Used
A	All-or-Nothing	All of the population or employment is assigned to the CSTDM TAZ that contains the greatest proportion of the MPO zone.
P	Proportion	Population or employment is allocated based on the proportion of the MPO zone that is in each CSTDM TAZ
D	Developable	The editor manually determines the population or employment that is allocated to each CSTDM TAZ

The first code (A) was used to assign all the data numbers to the largest CSTDM TAZ in the MPO zone. The second code (P) was used to assign the data numbers based on the proportion field. This field was used to determine the percentage of the MPO/RTPA TAZ that was in each CSTDM TAZ. The third code (D) was used to assign the percentages manually based on the discretion of the person doing this task using visual imagery and the protected area layer. The code system was faster than entering the percentages in manually for every non-nested zone.

The Split Percentage field was automatically calculated if the A or P Split Type codes were used. If the A Split Type code was used, the CSTDM TAZ with the largest area received 100 percent of the data while the other CSTDM TAZs that overlapped with the zone got 0 percent. When the P code was used, the Split Percentage field was calculated as equal to the proportion field. The editor manually entered in values to the Split Percentage field when the D Split Type code was used.

Split Type and Split Percentage fields were created for both the base year and the future year. This allowed for the differentiation of built and developable land. For example, if one of the TAZs in a non-nesting zone is completely developed in the base year and the other TAZs are vacant, the Split Percentage in the base year would be assigned a value of 100%. However, for the future, that same TAZ should have a Split Percentage equal to 1% (accounting for infill) and the other TAZs that were vacant in the base year should have the majority of the future employment and population.

After the MPO zone splitting task had been completed and all the polygons had been assigned a Split Percentage, an error check was performed. Rounding errors or incorrect manual inputs were remedied by checking that the Split Percentages added up to 100% for all non-nesting MPO zones. Once this was completed, the shape files were used for the next task described in section 2.1.3 below.

2.1.3 Aggregating Forecast Data to CSTDM TAZ

After the non-nesting MPO/RTPA TAZs were manually reviewed and assigned split percentages for the available data categories, all of the MPO/RTPA data was aggregated to CSTDM TAZs. This procedure was implemented in a Python script so that it could be applied quickly and consistently to each of the MPO and RTPA datasets.

The first section of the Python script identifies the working directory, input files and fields, and an output file name. These values are different for each MPO/RTPA, and are modified each time the script is run, to match the values specific to that MPO. The working directory is generally set up with an input directory and an output directory, where files are read from and written to. The basic set of inputs to the script is as follows.

1. Input directory (the location of the input data)
2. Input data files (the names of the files):
    a. Non-nesting MPO TAZ layer (the end product of the manual zone splitting process explained in section 3.1.2)
    b. Nesting MPO TAZ table (a table containing the nesting MPO TAZs, their demographic attributes, and the CSTDM TAZ they nest within)
    c. CSTDM TAZ layer covering the MPO area
3. MPO and CSTDM TAZ ID fields
4. Demographic data fields (specific field names from the input files):
    a. Population (base and future years)
    b. Employment (base and future years)
    c. School enrollment (base and future years broken down by school level, if available)
    d. Any other demographic fields relevant to that MPO or RTPA
5. Split percentage fields (fields specifying the proportion of the demographic data to allocate to this polygon, for non-nesting MPO or RTPA TAZs only):
    a. Population percent (base and future years)
    b. Employment percent (base and future years)

c. School enrollment percent (base and future years broken down by school level, if available)
d. Split percent for any other demographic fields

The only output variable that must be specified is an output filename. This file, which is created by the script, will contain the final demographic totals, after being aggregated to CSTDM TAZs. See Appendix C for the full text of the Python script, including all of the input variables. Comments within the script (text following a "#" character) describe each input variable and most of the processing code.

After all of the input variables were been specified, the script began reading and processing the data. First, the data from the non-nesting (split) MPO/RTPA TAZs were read. The split percentages were used to allocate a portion of the MPO/RTPA demographic values to the split polygons, but this was done differently for base year data and future year data. For base year data, the values were multiplied by the base year split percentage. This represents the current allocation of demographic attributes or categories within the MPO/RTPA TAZ. For each future year, the difference was calculated relative to the base year value. If the difference was positive, then the MPO/RTPA TAZ was forecast to experience growth. In this case, the future split percentage (which represented developable area within the MPO/RTPA TAZ) was multiplied by the amount of growth that was expected to occur. This value was added to the base year value to get a total value for each future year. If the difference was negative, then the MPO/RTPA TAZ was expected to experience a decline in terms of that demographic attribute. In this case, the base year split percentage was multiplied by the future year value. This effectively represented an even decline across the currently developed areas of the MPO/RTPA TAZ.

After the demographic attributes had been split for the non-nesting TAZs, an error checking algorithm was applied to the split values. It compared the sum of the split values to the original value for each MPO/RTPA TAZ, to make sure that they match. Small rounding errors may have been introduced as a result of splitting integer values

(people, jobs, etc.), and this algorithm corrected any of these minor discrepancies by adjusting the split values so that they exactly matched the original total value for the MPO/RTPA TAZ. At this point, the script outputs a copy of the split MPO/RTPA TAZs for reference.

Once the non-nesting MPO/RTPA TAZ attributes had been split and checked, they were aggregated to CSTDM TAZs. The nesting MPO/RTPA TAZ file was then read and the attributes for these zones were also aggregated to CSTDM TAZs. Next, the CSTDM TAZ file was copied to the output directory with the specified output name, and fields were added to contain each of the demographic attributes. Finally, the nesting and non-nesting TAZ attributes were added together for each CSTDM TAZ and these values were written to the output CSTDM TAZ file.

Many MPOs also provided detailed demographic data, such as population by age group or the number of jobs by employment category. These values needed to be aggregated to CSTDM TAZs in the same manner as the main demographic attributes (population, number of jobs, etc.). A second Python script (see Appendix D) was developed to handle this splitting and aggregation process. It follows almost exactly the same procedure as the first script, but was modified to accept additional subclass tables containing demographic subcategories as input. An appropriate split percentage field is specified, which is then applied to each subcategory. For example, the population split percentage was used to split all of the population by age group categories.

2.1.4 CSTDM TAZ Splitting for the TAZs that Exceeded Maximum Thresholds

A procedure to split CSTDM TAZs was developed once it was recognized that TAZs existed with forecasted values for population above 25,000 and/or above 15,000 for employment in future years (2020 and 2035). This resulted in too many trips originating and terminating from these TAZs. In order to alleviate this problem, these TAZs were split into smaller units with population and employment totals below the threshold values (see Figure 5). Exceptions were made for CSTDM TAZs where the population or

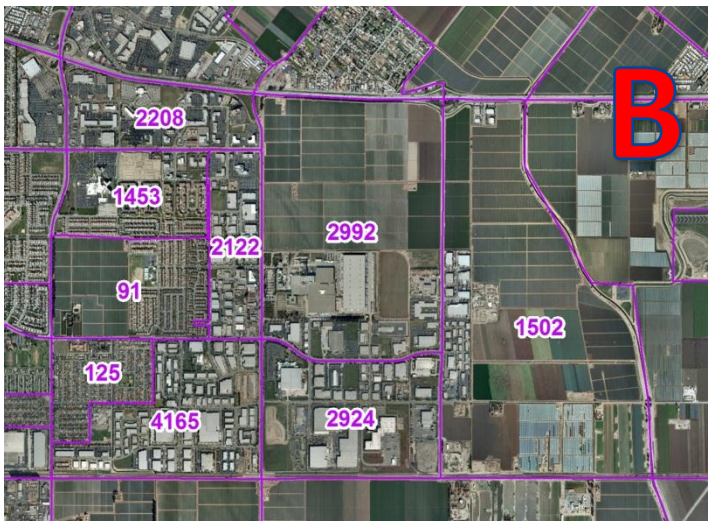
employment threshold was exceeded but these people or employees were constrained to a single building or business complex (examples: military bases, airports, business centers). Two methods were used to make these splits: the aggregation and cut polygon methods.

In order to split CSTDM TAZs into smaller polygons, we used future year projections to guide where these splits should occur. Since the future year projections were given to us by MPO TAZs and these TAZs are smaller than CSTDM TAZs, we began by aggregating MPO TAZs to create new CSTDM TAZs. This process often resulted in shape files that had topology errors. In these cases, we recreated the split CSTDM TAZs by splitting the original CSTDM TAZ. Both of these methods are described below.

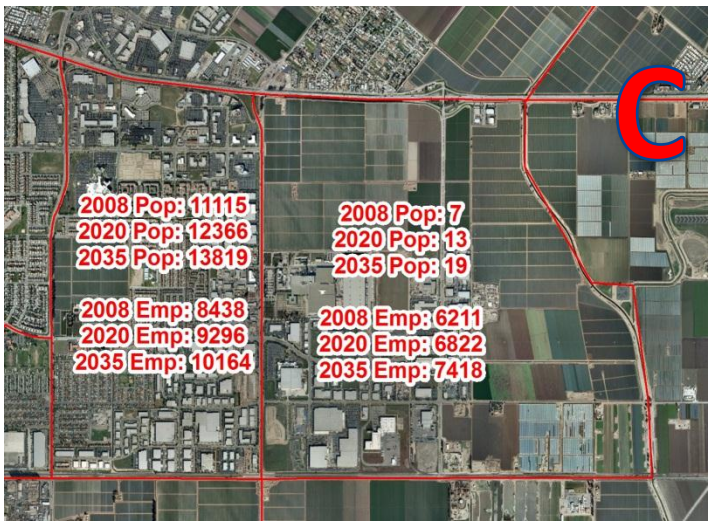
The aggregation method uses MPO TAZs and aggregates (combines) them to create new CSTDM TAZs. We began by splitting MPO TAZs that cross the boundaries of CSTDM TAZs. Then using population and employment totals from the MPO TAZs and the ESRI World Imagery Map, we combined them until they formed CSTDM TAZs with population and employment totals below the maximum threshold. The ESRI World Imagery Map was used to help identify natural barriers that could be used in splitting CSTDM TAZs, such as railroads, rivers/canals, freeways, mountains, or anything else that creates a barrier to travel.

When the aggregate method created topological errors in the CSTDM TAZ shape file, a second method was implemented: the cut polygon method. This method consisted of using the “Cut Polygon” tool in ArcGIS to split CSTDM TAZs with population and employment totals that exceeded the maximum thresholds. They were cut along the same boundaries determined in the aggregate method. This method was used rather than trying to fix the topological errors because it was faster than trying to fill in all gaps or remove areas where polygons overlapped (the two most common topological errors).

A. This is a CSTDM TAZ with a 2035 employment of 17,582. Since the maximum number of employees we want in a TAZ is 15,000 this TAZ should be split.



B. These are the MPO TAZs within the CSTDM TAZ. Displayed is the 2035 employment. MPO TAZs are used to aid in determining where the CSTDM TAZ should be split.



C. The aggregation method was used to combine the right 3 MPO TAZs into a CSTDM TAZ and the left 6 MPO TAZS into a second CSTDM TAZ. The result is 2 CSTDM TAZs with employment totals under 15,000.

Figure 5: An Example of a CSTDM TAZ that was Split due to High Employment in 2035

Splitting the CSTDM TAZs due to the original ones exceeding population and/or employment totals resulted in a new CSTDM TAZ system. The original CSTDM TAZ system consisted of 5191 zones, the new one has 5421. The counties where these 230 TAZs were added are displayed in Table 3.

Table 3: Counties where CSTDM TAZs were Split

County	Original Number of TAZs	New Number of TAZs	TAZs Added
Alameda	228	238	10
Butte	27	29	2
Contra Costa	132	140	8
El Dorado	27	28	1
Fresno	139	146	7
Imperial	17	19	2
Kern	138	151	13
Kings	27	29	2
Lake	9	10	1
Los Angeles	1301	1320	19
Madera	25	32	7
Merced	41	44	3
Monterey	64	66	2
Orange	354	370	16
Placer	39	43	4
Riverside	263	286	23
Sacramento	198	210	12
San Bernardino	267	282	15
San Diego	462	482	20
San Francisco	146	159	13
San Joaquin	110	115	5
San Mateo	109	112	3
Santa Barbara	55	58	3
Santa Clara	275	297	22
Solano	64	65	1
Stanislaus	76	81	5
Tulare	79	85	6
Tuolumne	11	12	1
Ventura	104	107	3
Yolo	27	28	1
		Total:	**230**

2.2 Employment by Industry Correspondence

The CSTDM requires forecasts of employment industry categories, which describe the type of activity at a person's place of work. The CSTDM uses eight North American Industrial Classification System (NAICS)-based industry categories, shown in Table 4, which is the federal government's standard industry classification system that groups establishments into industries based on the activities in which they are primarily engaged.

Table 4: CSTDM Industry Categories

NIACS Codes	Short Name	Description
11,21,23, 31-33	Prim_Sec	Primary and Secondary
42	Whole	Wholesale
22,48-49	TranU	Transportation and Utilities
44-45	Retail	Retail
51-56,92	Office	Office
61-62	EdMed	Education and Health
71-72	LeisHops	Leisure and Hospitality
81	OthServ	Other Service
N/A	Military	Military

Each of the 21 MPOs/RTPAs use employment by TAZ as inputs to their travel demand models. MPOs/RTPAs typically classify employment in either industry categories or land use categories, which needed to be transformed into the eight plus military CSTDM industry categories. This was not entirely straightforward because of differences in the definitions of employment categories. With the exception of San Diego, which provided military employment forecasts, military employment was assumed to remain unchanged from the base year. Some MPO/RTPAs have employment categories that correspond to the NAICS classification systems, which enabled a clear crosswalk between CSTDM industry categories and are also consistent with NAICS classification system. Other MPOs/RTPAs provided descriptions of categories sufficient to develop relationships to the CSTDM categories. As an additional check of category correspondence, MTP/RTPA and CSTDM base year totals by category were compared.

There were four possible structures of the correspondence between categories:

(1) One-to-one: the MPO/RTPA and CSTDM employment category share the same definition. For example, MPO/RTPA has a retail category, whose definition corresponds exactly to that of the CSTDM retail category.

(2) Many-to-one: the MPO/RTPA has two or more categories with definitions that combine to be equivalent to a single CSTDM category. For example, an MPO/RPTA has an education category as well as a medical category, which can be combined to make the CSTDM education/medical category.

(3) One-to-many: the MPO/RTPA has a category with a definition that is equivalent to two or more CSTDM categories. For example, the MPO/RTPA may have a non-office service category that includes employment in the CSTDM's primary/secondary, transport/utilities, and wholesale categories.

(4) Many-to-many: the MPO/RTPA TAZ has more than one category whose definitions enable a combination into more than one combination of CSTDM categories. For example, the MPO/RTPA has retail and leisure/hospitality categories that map well to the CSTDM, but then have categories, such as manufacturing/other, government, and general service, which don't nest neatly to any CSTDM categories without some overlap.

In case 3 and case 4, the category distributions for the 2008 CSTDM TAZ were applied to split these larger categories (case 3) or groups of categories (case 4) into the full set of CSTDM categories. Some zones with future employment have no employment in the 2008 CSTDM, so the county-level distribution were used instead.

Once the employment category correspondences were determined (see Appendix E) from the base year, the future employment was converted to the CSTDM categories for each zone for each year modeled by the MPO/RTPA.

A few MPO/RTPAs provided the amount of land use by type rather than employment by industry. In these cases, the base-year relationships between land use quantities and

CSTDM employment by TAZ were used to calculate the future employment. The future employment for each industry category was calculated according to equation 1.

(1) $$Emp_{k,i,\text{Future}} = Emp_{L,i,\text{Future}} \left(\frac{Emp_{L,i,\text{Base}}}{Emp_{k,i,\text{Base}}} \right)$$

where Emp is the number of employees for TAZ i, industry category k, or corresponding land use category L, for a given year.

2.3 Population from Households

Some MPO/RTPAs did not provide population forecasts by TAZ, because they had household-based travel models. CSTDM requires population by TAZ in order to generate the synthetic population. All MPO/RTPAs that lacked population data did have the number of households by type in each TAZ. The availability of zonal data is summarized in Appendix B. After transforming the household categories spatially into the CSTDM TAZ from the MPO/RTPA TAZ, as described in Section 2.1, it was necessary to convert households by type into persons.

Rates of persons per household by type were calculated for each county based on the 2000 US Census Summary File 3. The census provides household size by tenure and unit type in table “HCT003: TENURE BY HOUSEHOLD SIZE BY UNITS IN STRUCTURE [83] - Universe: Occupied housing units.” Table 5 shows how the CSTDM/MPO/RTPA household types correspond to the census categories.

Table 5: Household Size Correspondence – US Census and CSTDM

Tenure	Census Unit Type Categories	CSTDM HH Types	
Owner Occupied	1, detached or attached	Sfd	SF
	2 to 4	Att	MF
	5 to 19		
	20 to 49		
	50 or more		
	Mobile home	Mh	SF
	Boat, RV, van, etc.		
Renter Occupied	1, detached or attached	Sfd	
	2 to 4	Multi	MF
	5 to 19		
	20 to 49		
	50 or more		
	Mobile home	Mh	SF
	Boat, RV, van, etc.		

Then number of households by number of persons in each category was used to create an average rate of persons per household for each CSTDM household type in each county. Because, the census has a top code category of five or more persons per household, the average size of these large households was computed for each county. After computing the number of people in households with four-or-fewer people, the number of persons in five-or-more households is the remainder of people in the county based on the census's household population by tenure (rent or own) table "QT-H3 - Total population: Household Population and Household Type by Tenure: 2000." Then, the average size of the five-or-more households for a county is the number of people in five-or-more households divided by the number of five-or-more households. Taking the number of persons per household for each type and aggregating to the CSTDM types provides the rate of persons per household by type for each county. The rates are then applied to the number of household by type at the TAZ-level for all counties that did not provide zonal population directly. The only exception to this process is SACOG, which does not directly provide zonal population, but provides persons per household data by TAZ. This data was similarly used to create a rate for each TAZ to compute the population for the TAZ.

2.4 Time Horizon Correspondence

Many of the population and employment forecasts provided by MPOs/RTPAs were available for a range of time horizons (referred to hereafter as MPO/RTPA time horizons), some of which coincided with the 2020 and 2035 time horizons required for (referred to hereafter as CSTDM time horizons) and some of which did not. A summary of MPO/RTPA time horizons by TAZ can be found in Table 6. Methods of interpolation growth rates (estimated from available MPO/RTPA) were used to develop population and employment forecasts for the CSTDM time horizons from the MPO/RTPA time horizons when necessary. At this point in the data development process, all of MPO/RTPA TAZ were aggregated to CSTDM TAZ and thus MPO/RTPA forecasts by time horizon years were in CSTDM TAZ form. First, the county totals for population and employment were developed for each CSTDM time horizon, and then the county totals were distributed to CSTDM TAZs.

Table 6: Time Horizons of Forecasts Provided by MPOs/RTPAs

Region	Years Available
MTC	2010, 2020, 2035
SANDAG	2008, 2020, 2035, 2050
AMBAG	2005, 2035
SLOCOG	2008, 2020, 2035
BCAG	2006, 2035
Tuolumne	2010, 2020, 2030, 2040, 2050
Shasta	2005, 2010, 2020, 2030, 2050
ACTC	2010, 2030, 2050
Fresno	2008, 2035
Kings	2010, 2015, 2020, 2025, 2030, 2035
Kern COG	2006, 2020, 2035
Madera	2005, 2020, 2035, 2050
Merced	2010, 2020, 2035
San Joaquin	2010, 2020, 2030
Stanislaus	2006, 2030
Tulare	2007, 2020, 2030, 2035
SACOG	2008, 2035
SBCAG	2000, 2030
SCAG	2008, 2020, 2035
Lake & Mendocino	2009, 2020, 2030

2.4.1 County-level Population Interpolation and Extrapolation

If a given MPO/RTPA provided forecasts for one or more years beyond a CSTDM time horizon year, then forecasts for the CSTDM time horizon year were interpolated using annual compound growth rates calculated from the interval of the closest available MPO/RTPA horizon year forecasts. For example, Shasta County's model has time horizons 2020, 2030 and 2050, but not 2035. The Shasta County annual compound growth rate $\left(AnnualPopGrowthRate_{county,2030,2050} \right)$ for the interval 2030 to 2050 was then calculated using equation 2 below based on county total population $(PopTotal_{\text{county},2030}\ and\ PopTotal_{\text{county},2050})$.

(2) $$AnnualPopGrowthRate_{county,2030,2050} = \left(\frac{PopTotal_{\text{county},2050}}{PopTotal_{county,2030}} \right)^{\left(\frac{1}{2050-2030} \right)} - 1$$

The growth rate calculated above would then be applied to the 2030 population for five years to obtain the 2035 population forecasts in equation 3 below.

(3)

$$PopTotal_{county,2035} = PopTotal_{county,2030} * \left(1 + AnnualPopGrowthRate_{county,2030,2050} \right)^{(2035-2030)}$$

If an MPO/RTPA has no time horizon of 2035 or later, then the 2035 CSTDM population was obtained through a process of extrapolation that used annual compound growth rates$\left(AnnualPopGrowthRate_{county,2030,2050} \right)$ calculated from the California Department of Finance (DOF) 2050 forecasts, which are described in Appendix F. This was the case for San Joaquin, Stanislaus, Santa Barbara, Lake, and Mendocino County. The $AnnualPopGrowthRate_{county,2030,2050}$ is computed for each county, as in equation 2, using the DOF population forecasts by horizon year in place of MPO/RTAP population forecasts by horizon year. The county $AnnualPopGrowthRate_{county,2030,2050}$ is applied to the last available MPO/RTPA horizon year population forecast, as in equation 3. For example, 2030 population was available for San Joaquin County, but not 2035. The

average annual average growth rate was computed using DOF population forecasts for San Joaquin County for the horizon years 2030 and 2035 using equation 2. The county level annual growth rate was then applied in equation 3 to San Joaquin County's 2035 population forecast using a 2030 to 2035 time interval to obtain the San Joaquin County 2035 population forecast. This method was also used to develop 2050 county population totals for all MPO/RTPAs that did not have a 2050 time horizon.

2.4.2 Distributing County Population to TAZ

After developing county population totals for each CSTDM time horizon, these totals were disaggregated into TAZ based on the portion of change in population between MPO/RTPA time horizons for that zone relative to the county that it is in. Equation 4 was used to calculate the population for a TAZ in 2035, when the MPO/RTPA time horizons are 2030 and 2050.

(4) $$PopTotal_{TAZ,2035} = PopTotal_{TAZ,2030} + (PopTotal_{county,2035} - PopTotal_{county,2030})\left(\frac{\Delta PopTotal_{TAZ,(2050-2030)}}{\Sigma\Delta PopTotal_{TAZ,(2050-2030)}}\right)$$

where $PopTotal_{county}$ and $PopTotal_{TAZ}$ represents the county-total and TAZ-level population, respectively.

Other population-based socio-demographic categories, for example households by type, were interpolated using the same proportions. For example, using equation 4, this is effectively replacing $PopTotal_{TAZ,2030}$with the number of single-family households TAZ for 2030, but using the same ratio of change in population.

2.4.3 Interpolating TAZ Employment by Industry

At this point in the process, employment for each TAZ had been developed by industry categories, but only for the MPO/RTPA time horizons. It was necessary to estimate this employment data for the CSTDM time horizons. Based on the assumption that population and employment will grow at a similar pace for a given county, the CSTDM county-level population forecasts were used to estimate the employment for CSTDM time horizons. As shown in equation 5, these values were calculated by applying a

"portion-of-growth" factor (POG) to the change in employment between the bounding MPO/RTPA time horizons for each TAZ by industry category. The POG factor is simply the amount of change in population in an initial interval relative to an extended interval.

(5)

$$EmpTotal_{TAZ,industry,2035} = EmpTotal_{TAZ,industry,2030} + POG * \Delta EmpTotal_{TAZ,\text{industry},(2050-2030)}$$

where

$$POG = \left(\frac{\Delta PopTotal_{county,(2035-2030)}}{\Delta PopTotal_{county,(2050-2030)}}\right)$$

and $EmpTotal_{TAZ,\text{industry},2030}$ represents the 2030 employment for a certain TAZ, and a certain industry category.

3. California Areas without Local (MPO/RTPA) Forecasts for 2020 and 2035

As indicated in Section 2, local forecasts were not available from all RTPAs. These areas are the green areas in Figure 6 and include the counties of Alpine, Calaveras, Colusa, Del Norte, Glenn, Humboldt, Inyo, Lassen, Mariposa, Modoc, Mono, Nevada, Plumas, Sierra, Siskiyou, Tehama, Trinity, and the parts of El Dorado and Placer County that are not covered by the SACOG region (near Lake Tahoe). In these areas forecasts of population and employment are not available below the county level. As a result, procedures were developed to disaggregate county level population and employment projections to CSTDM TAZs for 2020 and 2035.

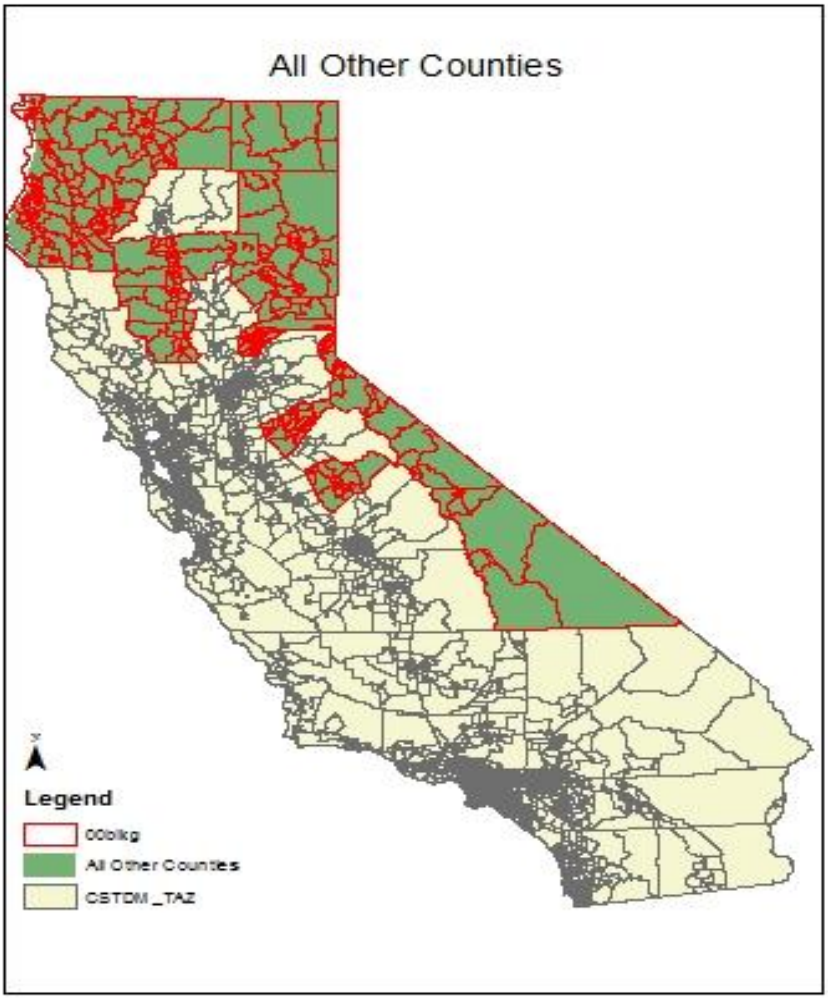


Figure 6: California Areas without Local (MPO/RTPA) Forecasts

3.1 Population

County level population projections were developed from California Department of Finance (DOF) Projections (see full discussion of data sources and transformation in Appendix F) for the years 2020 and 2035. The US Census Bureau provides population totals by Census Block Group (BLKG) for year 2000 and year 2010. Census block group systems are different for 2000 and 2010 census statistics. Each census block group GIS layer was projected on the CSTDM TAZ layer in order to generate TAZ-to-

census-block-group (BLKG) crosswalk. Two crosswalk tables, respectively for 2000 and 2010, were compared geographically to crosswalk the BLKGs for year 2000 and year 2010.

Year 2000 census BLKGs were kept as the basis layer for future steps. For each 2000 census BLKG, there was a pair of population statistics (referred to 2000 and 2010). Based on this information annual Census population growth rate was calculated at the Year 2000 census BLKG level as shown in equation 6.

(6) $$AnnualPopGrowthRate_{BLKG,2000,2010} = \sqrt[10]{\frac{PopTotal_{BLKG,2010}}{PopTotal_{BLKG,2000}}} - 1$$

So that the relation between the total population in 2010 and 2000 is:

(7) $$PopTotal_{BLKG,2010} =$$
$$= PopTotal_{BLKG,2000} * (1 + AnnualPopGrowthRate_{BLKG,2000,2010})^{(2010-2000)}$$

Population totals at year 2020 and 2035 at the BLKG level were then obtained by multiplying 2010 US Census Bureau BLKG population totals by growth rate at BLKG obtained from equation 6. The calculation is illustrated in equations 8.

(8)
For 2020:
$$PopTotal_{BLKG,2020} =$$
$$= PopTotal_{BLKG,2010} * (1 + AnnualPopGrowthRate_{BLKG,2000,2010})^{2020-2010}$$

For 2035:
$$PopTotal_{BLKG,2035} =$$
$$= PopTotal_{BLKG,2010} * (1 + AnnualPopGrowthRate_{BLKG,2000,2010})^{2035-2010}$$

The results of equations 8 are then scaled to match the total DOF population targets for 2020 and 2035 at the county level, as shown in equations 9.

(9)

For 2020:

$$AdjustedPopTotal_{BLKG,2020} = \frac{PopTotal_{DOF,county,2020}}{\sum_{BLKG}^{county} PopTotal_{BLKG,2020}} * PopTotal_{BLKG,2020}$$

For 2035:

$$AdjustedPopTotal_{BLKG,2035} = \frac{PopTotal_{DOF,county,2035}}{\sum_{BLKG}^{county} PopTotal_{BLKG,2035}} * PopTotal_{BLKG,2035}$$

Finally, the estimates that we obtained at the BLKG level were aggregated at the CSTDM TAZ level. Year 2000 BLKG geographies nest exactly to CSTDM TAZ geographies. Total population by CSTDM TAZ for 2020 and 2035 was then calculated by summing the total population of all BLKGs within the TAZ for each horizon year (2020 and 2035), as shown in equation 10.

(10)

For 2020:

$$FinalPopTotal_{TAZ,2020} = \sum_{BLKG}^{TAZ} AdjustedPopTotal_{BLKG,2020}$$

For 2035:

$$FinalPopTotal_{TAZ,2035} = \sum_{BLKG}^{TAZ} AdjustedPopTotal_{BLKG,2035}$$

3.2 Employment

The following sources of data were used to develop employment forecasts for RTPAs that did not have local forecast.

(1) County level employment forecasts by NAIC industry categories produced by the TREDIS model, which uses Moody's employment forecasts. The TREDIS model walks Moody's employment data through time using IMPLAN to produce employment forecasts by industry categories (see Appendix G for a more detailed description of this dataset).

(2) Non-residential floorspace by space type in the 50 by 50 meter grid cells for the State of California in the year 2000 was obtained from a floorspace synthesizer developed by Urban Land Use and Transportation Center (ULTRANS) at the University of California, Davis (see Appendix H for a more detailed description) for use in a spatial economic model for the State of California that is based on the PECAS framework.

(3) CSTDM population by county, i.e. the incremental population forecasts from the California Department of Finance (DOF) for year 2020 and year 2035.

The development of the employment forecasts required the development of a number of crosswalks between data categories and geographical information, as described below. The space types for non-residential floorspace in the 50 by 50 meter grid cells for the State of California (see Appendix H for a description) are listed in Table 7.

Table 7: Non-Residential Floorspace Categories

Space Type ID	Space Type Name
95	vacant space
101	park space
102	forest space
103	mining and extraction space
104	grazing space
105	non-irrigated agricultural space
106	irrigated agricultural space
107	light industrial space
108	heavy industrial space
109	warehouse space
110	highway retail space
111	downtown retail space
112	mall and big box retail space
113	neighborhood retail space
114	low density office space
115	high density office space
116	developed amusement parks space
117	hospital space
118	secondary education space
119	primary K-12 education space
120	religious space
121	government operations space
122	military space
123	fishing dock space
124	depot space

A crosswalk was developed to link non-residential floorspace types to CSTDM industry categories (see Table 8). All but one non-residential floorspace type could be assigned

to a single CSTDM industry category: space type 112 (mall and big box retail space) was split between Whole (wholesaled trade) and Retail (retail trade) categories.

Table 8: Crosswalk for CSTDM and California PECAS Framework Space Types

Employment by Industry	**Space Type**			
CSTDM	**California PECAS framework**			
PrimSec	104	105	106	123
	103			
	107	108		
Whole	50% of 112			
Retail	110	111	50% of 112	113
Tran_U	124			
	109			
Office	114	115		
	121			
EduMed	118	119		
	117			
LeisHosp	116			
OthServ	120			
Military	122			

Next, the datasets and crosswalks were used to calculate employment by CSTDM industry types by CSTDM TAZ. Non-residential floorspace by space type in the 50 by 50 meter grid cells for the year 2000 was associated with CSTDM industry categories using the crosswalk described in Table 6, as shown in equation 11.

(11) $$FloorspaceSQFT_{industry,cell} = \sum_{spaceType \in industry} FloorspaceSQFT_{spacetype,cell}$$

Non-residential floorspace by CSTDM industry categories in the 50 by 50 meter grid cells were then summed at the CSTDM TAZ level, as shown in equation 12.

(12) $FloorspaceSQFT_{industry,TAZ} = \sum_{cell \in TAZ} FloorspaceSQFT_{industry,cell}$

The share of non-residential floorspace by CSTDM industry categories by TAZ within a county was calculated by dividing non-residential floorspace by CSTDM industry category for each TAZ in a county by the sum of the non-residential floorspace by CSTDM industry category for all TAZs in that county, as shown in equation 13. As an example, Table 9 shows the shares for a small county, Calaveras County, with only six TAZs: the column headings, 3007-3012, represent the CSTDM TAZ numbers, and the rows are the CSTDM industry categories.

(13) $FloorspaceSQFTratio_{industry,TAZ} = \frac{FloorspaceSQFT_{industry,TAZ}}{\sum_{TAZ \in county} FloorspaceSQFT_{industry,TAZ}}$

Table 9: Share of Non-Residential Floorspace by CSTDM Industry Category by TAZ for Calaveras County with Six TAZs

(columns are TAZ and rows are CSTDM industry categories)

TAZ / Industry	3007	3008	3009	3010	3011	3012	Total
EduMed	1.81%	0.41%	15.41%	0.41%	50.91%	31.06%	100.00%
LeisHosp	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%
Military	20.52%	0.00%	0.00%	0.00%	22.40%	57.08%	100.00%
Office	22.81%	5.12%	22.89%	16.30%	10.80%	22.08%	100.00%
OthServ	5.43%	10.86%	0.00%	0.00%	0.00%	83.71%	100.00%
PrimSec	6.31%	13.84%	13.36%	36.73%	15.23%	14.53%	100.00%
Retail	43.89%	6.32%	19.60%	12.67%	5.81%	11.71%	100.00%
Tran_U	41.66%	0.00%	7.92%	2.88%	23.77%	23.77%	100.00%
Whole	45.28%	3.39%	17.20%	14.17%	4.61%	15.34%	100.00%
Total	6.33%	13.84%	13.37%	36.72%	15.22%	14.53%	100.00%

County level socio-economic datasets were integrated to this step, including TREDIS employment by industry, Moody's population, and CSTDM incremental DOF population. Equation 14 shows how the TREDIS employment data are scaled by the ratio of CSTDM population over Moody's Data.

(14) $EmpTotal_{CSTDM,industry,county} = PopTotal_{CSTDM,county} * \frac{EmpTotal_{TREDIS,industry,county}}{PopTotal_{Moody's,county}}$

For these regions, which present rather low densities and not very intensive economic activities when compared to other areas of California, we assume that each industry consumes the same ratio of floorspace across all study regions. Therefore, the spatial distribution of employment activities is then applied to disaggregate county level CSTDM employment totals into TAZs:

(15)

$$EmpTotal_{industry,TAZ} = EmpTotal_{CSTDM,industry,county} * FloorspaceSQFTratio_{industry,TAZ}$$

The result from equation 15 is the total employment by CSTDM industry at the CSTDM TAZ level.

4. Forecasts for 2050

Amador, Madera, San Diego, Shasta, and Tuolumne counties provide 2050 population and employment data, which was transformed for use in the CSTDM as described in Section 2. This section describes the process to develop socio-economic data for all other counties.

4.1 County-Level Population and Employment Totals for 2050

For counties within MPO/RTPAs, when population forecasts were not available for year 2050 from the MPO/RTPAs, they were estimated using extrapolation of the latest available year. The MPO/RTPA county population forecasts were extrapolated using annual compound growth rates computed from DOF forecasts. This process is described in more detail in Section 2.31. For areas outside of MPO/RTPA, the 2050 population was taken from the adjusted DOF forecasts described in Appendix F.

When time horizon 2050 was not available from MPO/RTPAs, 2050 employment was developed assuming the ratio of population to employment for each county was the same for 2050 as 2035. The ratio of population in 2050 to 2035 was applied to the 2035 employment as shown in the formula below.

$$Emp_{2050} = Emp_{2035}\left(\frac{Pop_{2050}}{Pop_{2035}}\right)$$

4.2 Allocation of 2050 Population and Employment to CSTDM TAZ

For counties that do not have zonal population and employment data from MPO/RTPAs, the following method was used to allocate the 2050 county level population and employment (described above) to CSTDM TAZ.

The spatial patterns of an urban growth model were used to disaggregate 2050 population totals to CSTDM TAZ geographies. Uplan is a rule-based urban growth model (http://ice.ucdavis.edu/doc/uplan). It was used in a project that looked at impacts to dispersal corridors, fire threats and climate-sensitive agriculture across California for six policy scenarios for the California Energy Commission (Thorne, In review). Of the six scenarios, we chose to use the "business-as-usual" one. This policy assumes that the percentage breakdown at the county level for housing densities is the same in 2050 as it was in 2000. There are four residential density categories: R20 (20 units per acre), R5 (5 units per acre), R1 (1 unit per acre) and R.1 (one unit on a 10 acre lot). For example, if the percentage breakdown in 2000 for County A was 15% R20, 30% R5, 40% R1 and 15% R.1 and there are 1000 households added between 2000 and 2050 in that county, then 150 households would be R20, 300 would be R5, 400 would be R1 and 150 would be R.1. There are three employment categories: commercial-high, commercial-low, and industrial. The three employment categories did not match well with CSTDM categories, so the total was used for distributing employment to CSTDM TAZ.

Uplan allocates future urban growth based on a net attraction raster and a general plan. The net attraction raster is created by subtracting things that are discouragements to development from things that are attractions to development. Examples of attractors are proximity to existing urban areas and transportation facilities. Examples of discouragements are 100-year floodplains, farmland, and certain habitat types such as vernal pool complexes. The model also has exclusions areas where development cannot occur, such as rivers, public/open space, and existing built-out urban areas.

Once the net attraction raster is created and the exclusionary areas are removed, land uses are allocated based on a general plan raster.

The data we received was a statewide raster dataset (30m x 30m) that represented where the model predicts development is likely to occur between 2000 and 2050. Each raster cell had a value that represents one of four residential categories or one of three employment categories. We summarized the number of cells within each CSTDM TAZ for each residential type. The number of units per acre was then converted to number of units per CSTDM TAZ. Using each county's average number of people per household (an input into the Uplan model), we converted units per CSTDM TAZ into people per CSTDM TAZ. This summary resulted in the number of people added to each CSTDM TAZ between 2000 and 2050.

We have total population and employment per CSTDM TAZ for the year 2000, because 2000 was a base year run for CSTDM. We added the growth in population and employment in each CSTDM TAZ between 2000 and 2050 (from Upaln) to year 2000 population and employment totals (from CSTDM). This gave us total population and employment (as predicted by Uplan) in year 2050 by CSTDM TAZ.

We wanted to use the population and employment totals described in Section 4.1, not Uplan so we needed to treat the 2050 Uplan totals as proportions, not absolute numbers. To achieve this, on a per CSTDM TAZ basis, we computed the difference between 2035 population and employment (Sections 2 and 3.1) and Uplan 2050 population and employment. Any TAZ that had a population or employment decrease between 2035 and 2050 was changed to 0 rather than having a negative number. This decision was made because we wanted to use Uplan to guide where 2050 development is likely to be (not take away from forecasted 2035 populations). For the majority of the state, the 2035 population and employment totals were derived from MPO/RTPA data so we wanted to retain the trends that local officials are predicting.

Once we had total Uplan predicted population increases between 2035 and 2050 for each CSTDM TAZ in a county, we divided this increase by the total predicted increase for the county. This resulted in proportions that represented the location (TAZ) and magnitude of population and employment increases between 2035 and 2050 as predicted by Uplan. These proportions were multiplied by the population or employment increase by county between 2035 and 2050 according to the county totals described in Sections 2.41 and 4.1. The last step was to add these increments to the 2035 values to get total population and employment by CSTDM TAZ for the year 2050.

4.3 Employment Industry Categories

After total employment for each TAZ was developed for time horizon 2050, it needed to be split into industry categories. The distribution of employment categories within a zone was assumed to remain similar from 2035 to 2050. For each zone, the proportion of each industry category in 2035 was applied to the total employment for that zone in 2050. If the zone had no employment in 2035, the county-level distribution was used instead.

5. Final Population and Employment

5.1 Population Synthesis

In the CSTDM, the synthetic population represents every person and housing unit in California; it is based on sampling U.S. Census Public Use Microdata Sample (PUMS) 5% person and household data to match targets that can be derived from sources such as Summary File 3 (SF3) and the American Community Survey (ACS). The population synthesizer works by combining a trial population of households and altering it by switching new possible households in; if the match with the targets (or observed, estimated, or forecast totals) improves, the new household is kept. A detailed description of the algorithms used in this process is part of the detailed documentation of the population synthesizer. The population synthesizer is capable of handling multiple nested geographies, of matching categorical totals or averages, and of weighting possible targets.

In this application, the population synthesizer uses a large number of targets (or forecast totals) that the synthesizer must match. These include forecast data at the (1) CSTDM TAZ level for all available population attributes from MPOs/RTPAs, and other areas (as described in Section 2), (2) county level age distributions from the DOF (see Appendix F), and (3) land use analysis zone (LUZ) level household category distributions (i.e. size, workers, income, and autos) available from the Ridership and Revenue Forecasting Study conducted for the CHSRA (Cambridge Systematics, 2007) (see Appendix I). The development data for the targets files is discussed in Section 2, 3, and 4. Weight multiplier files and samples files are also required by the Population Synthesizer. A weight multiplier file is used to identify the PUMAs (geographic unit of analysis for PUMS records) in and around TAZ from which to draw PUMS household records. Samples files contain the regional PUMS households. Due to regional differences in target categories, unique population syntheses using region specific population attributes were conducted. The results of the population syntheses for all the regions are combined into one statewide synthetic population for each horizon year. This process is illustrated in Figure 7 below.

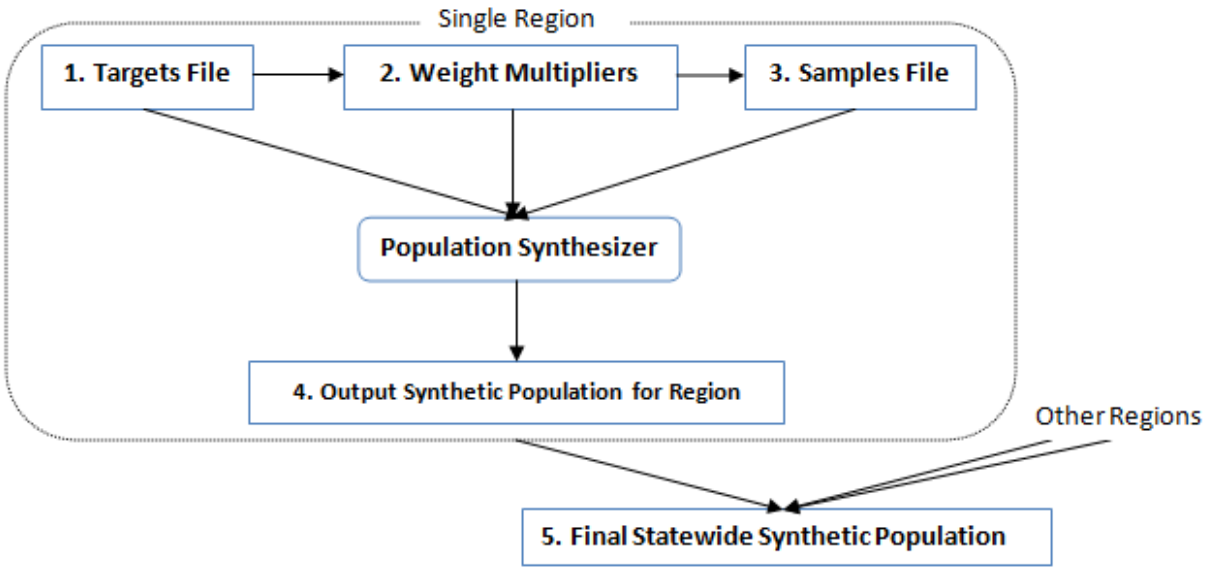


Figure 7: Stages to Create the Synthetic Population

5.1.1 Creation of Population Synthesizer Inputs

5.1.1.1 Targets File Creation

The targets file for each region (21 for all MPOs and RTPAs and one for the remaining areas of California) is a CSV table assembled with data from the various sources described above in Sections 2, 3, and 4. Each of the 21 target files include (1) CSTDM TAZ level population attributes for the region, (2) 20 population by age categories for the counties in the region from the DOF; and (3) 99 household categories for the LUZ in the region from the CHSRA. These household categories are combinations of the number of persons (1 to 4 or more) and workers (0 to 2 or more) in the household and the household income (low, medium, or high) and auto ownership (0 to 2 or more). The targets file also specifies the appropriate nesting of the smaller geographic units within the larger geographic units (i.e., TAZ by LUZ by county).

Table 10 shows a portion of the targets file for Amador County. Available population categories are listed across the top header row. In this example, they are zone number, population, households by type, person/worker/auto/income categories from the previous HSR model, and population by age. Row 2 contains a weight that is used to balance target categories. For example, population has a weight of 1, but the HSR categories, which have small fractional values, get a weight of 10,000. Rows 3 and higher contain the target value for each zone for each category. If a target is to be met at a larger geographic level, these cells contain a negative number that specifies the LUZ or county in which the TAZ is nested. For example, TAZ 3001 has a population of 8,160, and is part of LUZ 2601 and county 1. The row for a LUZ or county has zeros for the categories, which have targets at a finer geographic level. In the example below, 2.4% of the households in LUZ 2601 are in HSR category HHS1_NW0_I0 meaning that they have one person, zero workers, low income, and do not own any cars. County 1 has 2,213 people ages 5 through 9.

Table 10: Sample Targets File (showing a subset of typical columns)

Zone	Pop	type_sf	type_mf	HHS1_NW0_I0-A	HHS1_NW0_I7-A	age_0_4	age_5_9
0	1	1	1	10000	10000	1	1
3001	8160	3285	87	-2601	-2601	-1	-1
3002	6910	2701	183	-2601	-2601	-1	-1
3003	9980	3892	275	-2600	-2600	-1	-1
3004	11220	3842	1031	-2600	-2600	-1	-1
3005	6290	2158	571	-2600	-2600	-1	-1
3006	6730	1979	1057	-2600	-2600	-1	-1
2601	0	0	0	0.024332661	0.003130631	0	0
2600	0	0	0	0.013444112	0.002400909	0	0
1	0	0	0	0	0	2251	2213

5.1.1.2 Weights Multipliers Creation

The weights file is a CSV file that tells the population synthesizer which PUMA(s) to draw the PUMS household samples for each CSTDM TAZ. The file has three columns, TAZ, PUMA, and weightMultiplier. When the population synthesizer is drawing household samples from the PUMS records, it uses the weightMultipliers for each TAZ to determine from which PUMA to choose. For each TAZ, there are six possible PUMAs from which to sample households. These are the PUMA(s) in which the TAZ lie as well as the five nearest surrounding PUMAs. The PUMA(s) containing the TAZ has a weightMultiplier of 1, and the other five have a weightMultplier of 0.2.

The CSV file containing the targets is placed in the Prepare PopSynth Inputs folder and named “Targets_<SCENARIO>.csv”, where <SCENARIO> is the scenario for which the weights file is being created. The output weights file are called “weightMultipliers_<SCENARIO>. For example the targets file for Amador in model year 2020 would be named “Targets_Amador_2020.csv”, and the output will be named “weightMultipliers_2020.csv”. The command used to run the script would then be: python buildWeightsFromTargets.py Amador_2020. Table 11 is a snapshot of the weight multipliers file:

Table 11: Weight Multipliers Example

TAZ	puma	weightMultiplier
2828	3401	1
2828	3301	0.2
2828	3302	0.2
2828	3501	0.2
2828	3402	0.2
2828	3502	0.2
2818	3401	1
2818	3301	0.2
2818	3302	0.2
2818	3402	0.2
2818	3501	0.2
2818	3502	0.2

5.1.1.3 Samples File Creation

The samples file is a CSV database containing the PUMS household records. The population synthesizer draws household samples from this file and attempts to match the population attributes as represented in the targets. To generate the samples file, two sources of data are used: (1) the person and household records from PUMS and (2) the weight multipliers file generated for the specific area. Not all of the targets provided by regions exactly correspond to PUMS categories. For example, some regions provided population by age bands while the PUMS records contain exact ages. Thus, in some instances, the PUMS data categories were transformed into a form that would be consistent with regional target categories. Figure 8 provides an example of the samples file.

id	puma	weight	pop	hh	age_0_4	age_5_9	age_10_14	age_15_17	age_18_19	age_20_24	age_25_29
737	2001	16	2	1	0	0	0	0	0	0	0
4876	2001	10	2	1	0	0	0	0	0	0	0
12308	2001	0	1	1	0	0	0	0	0	0	1
14319	2001	32	2	1	0	0	0	0	0	1	1
15240	2001	5	2	1	0	0	0	0	0	0	0
20255	2001	16	1	1	0	0	0	0	0	0	0
21055	2001	19	2	1	0	0	0	0	0	0	0
32202	2001	21	2	1	0	0	0	0	0	1	1
33584	2001	18	2	1	0	0	0	0	0	0	0
38958	2001	26	4	1	0	0	2	0	0	0	0
42188	2001	24	2	1	0	0	0	0	0	0	0
51811	2001	13	1	1	0	0	0	0	0	0	0
57361	2001	21	3	1	0	0	0	0	1	0	0
58292	2001	23	4	1	0	0	2	0	0	0	0

Figure 8: Samples File (showing first 12 columns only)

A python script was developed to create the samples file. This script is run for each region for each time horizon, which is specified as an argument to the script.

5.1.2 Running Population Synthesizer

An instance or run of the population synthesizer exists for each region, and must be configured to use the appropriate input and output files. The samples, targets, and weightMultipliers files are placed in the InputOutput folder, and the Synth.properties file is adjusted with the names of the samples and targets files. The Synth.properties file is also used to specify the number of iterations and includes instructions on how to create the initial population. In all model year scenarios (2020, 2035, and 2050), 100 million iterations were performed. In the 2020 scenarios, the initialGenerationBasis parameter was set to “previous”, which means that it will use the 2008 population as a starting point. All subsequent years use the previous year as a starting point as well. After configuration is complete, the population synthesizer is run by executing RunSynth.bat in each of the regional folders.

5.1.3 Combining Regional Synthetic Populations

Finally, all of the MPO-level data are combined into a single table and inserted into the Synthetic Population database, which is called Cali.sqlite. This can be done

automatically with a python script, which was created for this purpose. The statewide synthetic population is now complete for the future years and can be queried and viewed in Cali.sqlite. This database can be used to analyze the synthetic population or as input to the CSTDM.

5.2 Employment by Occupation

The general principle for producing future employment by occupation given employment by industry is to assume that TAZs that do not change much into the future tend to retain their character (distribution of occupations in the industries present), while TAZs that do change will be similar to other TAZs of similar nature in the same part of the state.

The first step in this process was the production of a synthetic employment, using base year 2008 employment by industry and occupation, with each TAZ using the samples from the appropriate place-of-work PUMA to reflect the different employment mixes throughout the state. One notable example, for instance, is Santa Clara County, where primary and secondary industry workers are much more involved in managerial/business and professional/technical occupations than elsewhere. This reflects the nature of Silicon Valley, with many firms such as Apple that are classified as being in the manufacturing industry. Apple's design is done in Cupertino, but the production is done in China. In contrast, in low-density portions of the Central Valley, the majority of the primary and secondary industry group is agriculture with large proportions of blue collar agricultural workers.

This synthetic population provides a base distribution of employment by occupation for each industry in each zone. Thus, if one industry grows faster than others, the occupation growth should reflect that shift. For example, many downtown areas have a mix of office industries and leisure/hospitality industries. If the office industries increase more quickly, there will be more managerial, professional/technical and clerical workers in the future. If the Leisure/Hospitality sector increases, we would expect more Sales/Food/Entertainment workers. Table 12 below shows the breakdown of

occupations by industry statewide. It can be seen that most industries derive the bulk of their labor from one or two occupations.

Table 12: Statewide Distribution Employment by Industry and Occupation

Industry	Occupation							
	ManBus	ProfTech	Education	Health	SalesFE	Clerical	ServNS	BluCol
Prim_Sec	14.9%	8.9%	0.1%	0.2%	4.1%	9.3%	1.3%	61.2%
Tran_U	8.9%	3.5%	0.2%	0.1%	2.2%	28.8%	3.7%	52.7%
Whole	14.0%	2.8%	0.1%	0.2%	30.8%	18.7%	1.3%	32.1%
Retail	7.1%	2.0%	0.2%	2.1%	54.8%	15.4%	1.7%	16.6%
Office	23.4%	16.9%	0.7%	1.6%	14.7%	22.4%	12.4%	7.9%
EduMed	8.7%	7.1%	28.0%	25.5%	3.7%	14.0%	10.1%	2.7%
LeisHosp	12.1%	0.6%	0.8%	0.3%	61.7%	6.8%	12.7%	5.1%
OthServ	7.4%	8.1%	0.8%	2.3%	7.3%	9.9%	31.4%	32.9%

However, there is the potential for significant change in the future. If the nature of an area changes radically from the current situation (a second downtown rising out of a greenfield site, for instance), then expanding the current employment could be misleading. As a proxy for the nature of an area, the density can be used. Figure 9 below gives the proportion of "office support" (managerial/business, professional/technical and clerical workers) in different industries for the statewide average, for the densest and the least dense zones. There is a clear trend across all industries for companies in denser areas to perform more office support activity.

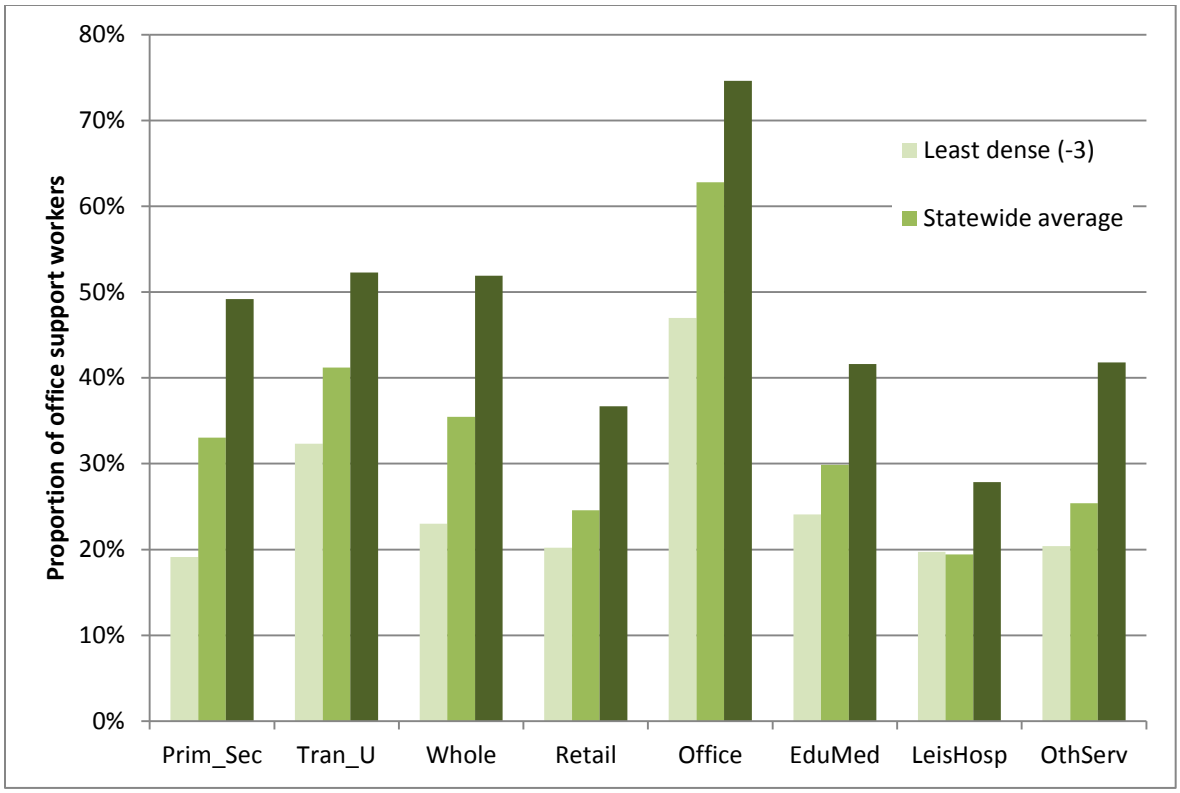


Figure 9: Proportion of Office Support Workers by Industry

To handle the possible change of an area into one of a very different character, every TAZ is classified into two dimensions. First, an Intensity Type describing the size of the employment center, and second, an Intensity Classification Area, which describes where the zone is in the state.

The intensity type is based on two measures: employment density and employment/population ratio, where

- Employment Density = Total number of jobs in zone / area of zone in square miles, and
- Employment/population ratio = number of jobs in zone / population of zone.

The density categories are based on the log of the employment density. Two TAZs with similar employment density may have very different distributions of industries and occupations if there is a significant residential presence in the same area (likely

indicating less noxious uses). TAZs are said to have high employment, if the employment/population ratio is greater than 0.75; in other words, if there are over 0.75 jobs for every resident. These zones are significant net employment areas, considering fewer than 50% of residents are workers. Certain uncommon intensity types are aggregated. The resulting 12 intensity types are shown in Table 13.

Table 13: Employment Intensity Types

Ln (emp. density)	Employment density	Normal zones	High emp. zones
-5 to 3	Under 20.1	-3	-3E
3 to 5	20.1 - 148.4	3	
5 to 7	148.4 - 1096.6	5	
7 to 8	1096.6 - 2981	7	7E
8 to 9	2,981 - 8,103	8	8E
9 to 10	8,103 - 22,027	9	9E
10 to 11.5	22,027 - 98,718	10E	
Over 11.5	98,718 or more	11.5E	

The maps in Figure 10, 11, and 12 below illustrate intensity (from light to dark) and normal (orange) versus high employment (blue) zones for the Sacramento, central San Joaquin Valley and greater Los Angeles areas.

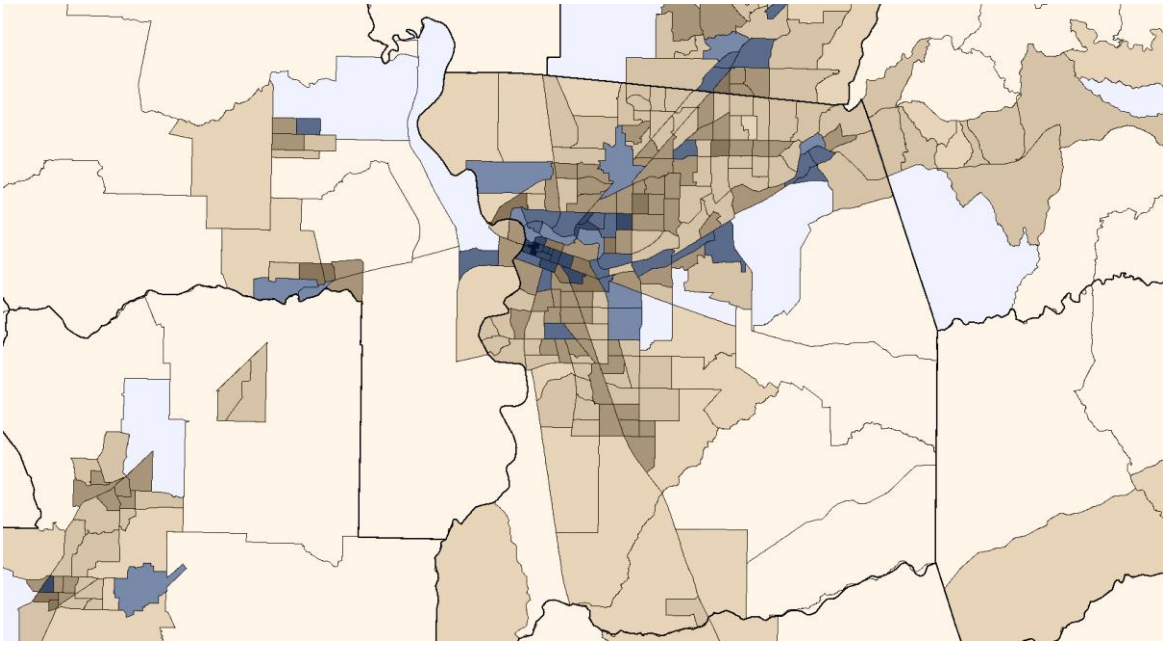


Figure 10: Employment Intensity Types (Sacramento Area)

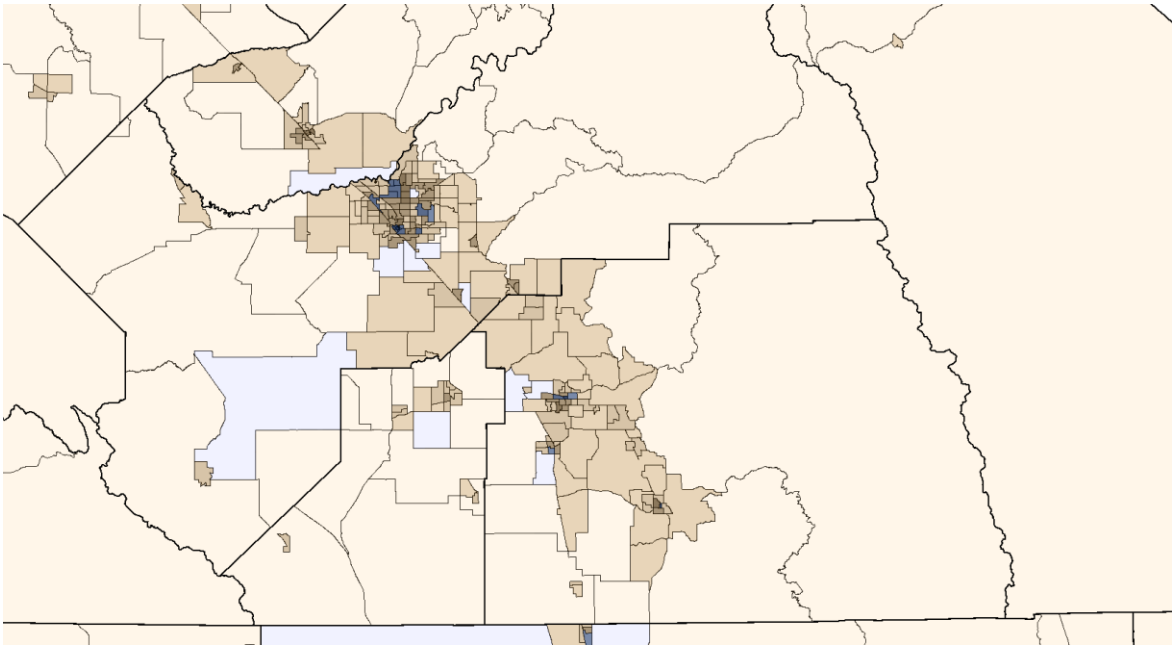


Figure 11: Employment Intensity Types (Central San Joaquin Area)

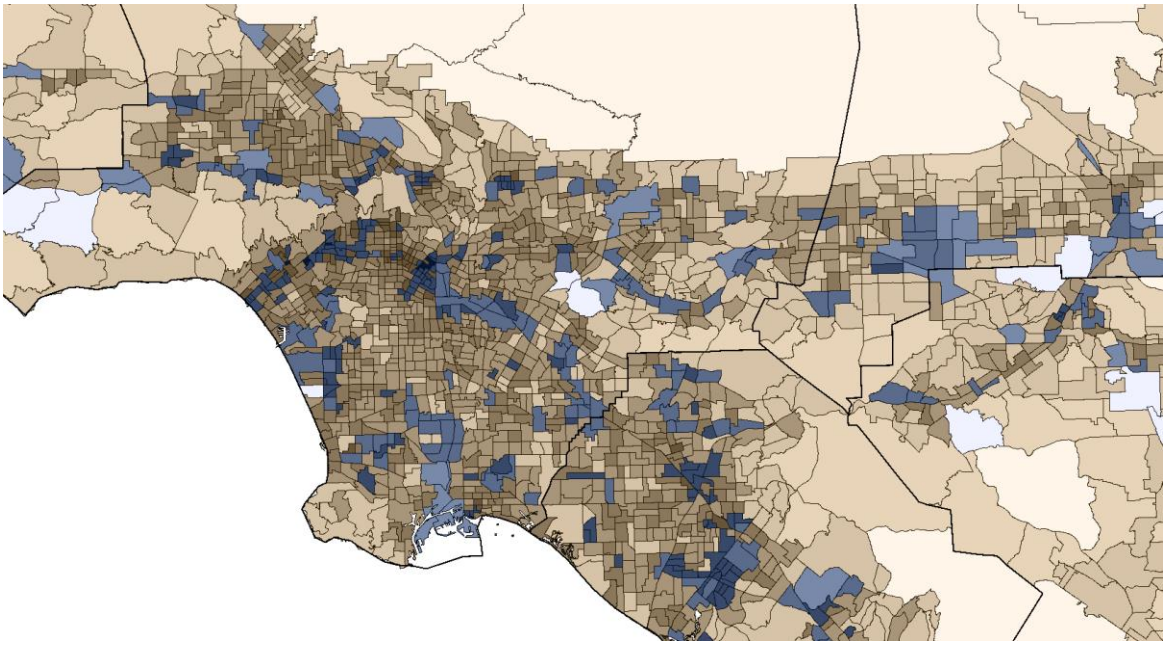


Figure 12: Employment Intensity Types (Los Angeles Area)

For the role of Intensity Classification Area, consider zone 2613, on the south side of Hanford in Kings County. In 2008, it is in Intensity Type 5, with an employment density of 530.6, and 0.18 jobs per resident (there are around 5000 residents). Let us assume the zone increases significantly in employment, from the base year 891 to 3000 workers. This is a significant increase in intensity, and the zone will be in a new Intensity Type, 7. The employment mix may be assumed to be shifting as the intensity increases, and the zone may function more like other Intensity Type 7 zones in Kings and Tulare counties (such as downtown Hanford, as well as central portions of other cities like Tulare and Visalia). A more extreme scenario could have the intensity increasing even more, say to 15,000 jobs. This would put the zone in Intensity Type 9E, and it could function more like the downtowns of large San Joaquin cities like Stockton, Modesto and Bakersfield.

In this way, the employment mix by intensity level can be subdivided into geographic areas, which can be aggregated and disaggregated, as data is available. In the far North, there are a large number of low-density zones, so the area can be subdivided into subareas for greater fidelity; the coastal counties may have different employment

distributions versus the mountain counties. For this area, however, there are no zones with very dense employment, so the denser employment distributions are taken from areas like SACOG or the northern Bay Area counties, which do have this data.

The following two maps (Figures 13 and 14) illustrate the Intensity Classification Area definitions for 8, a moderately intense use, and for -3, the least intense land use. The Intensity Classification Areas for 8 are finer in the urban areas, while the Intensity Classification Areas for -3 comprise more divisions in rural areas.

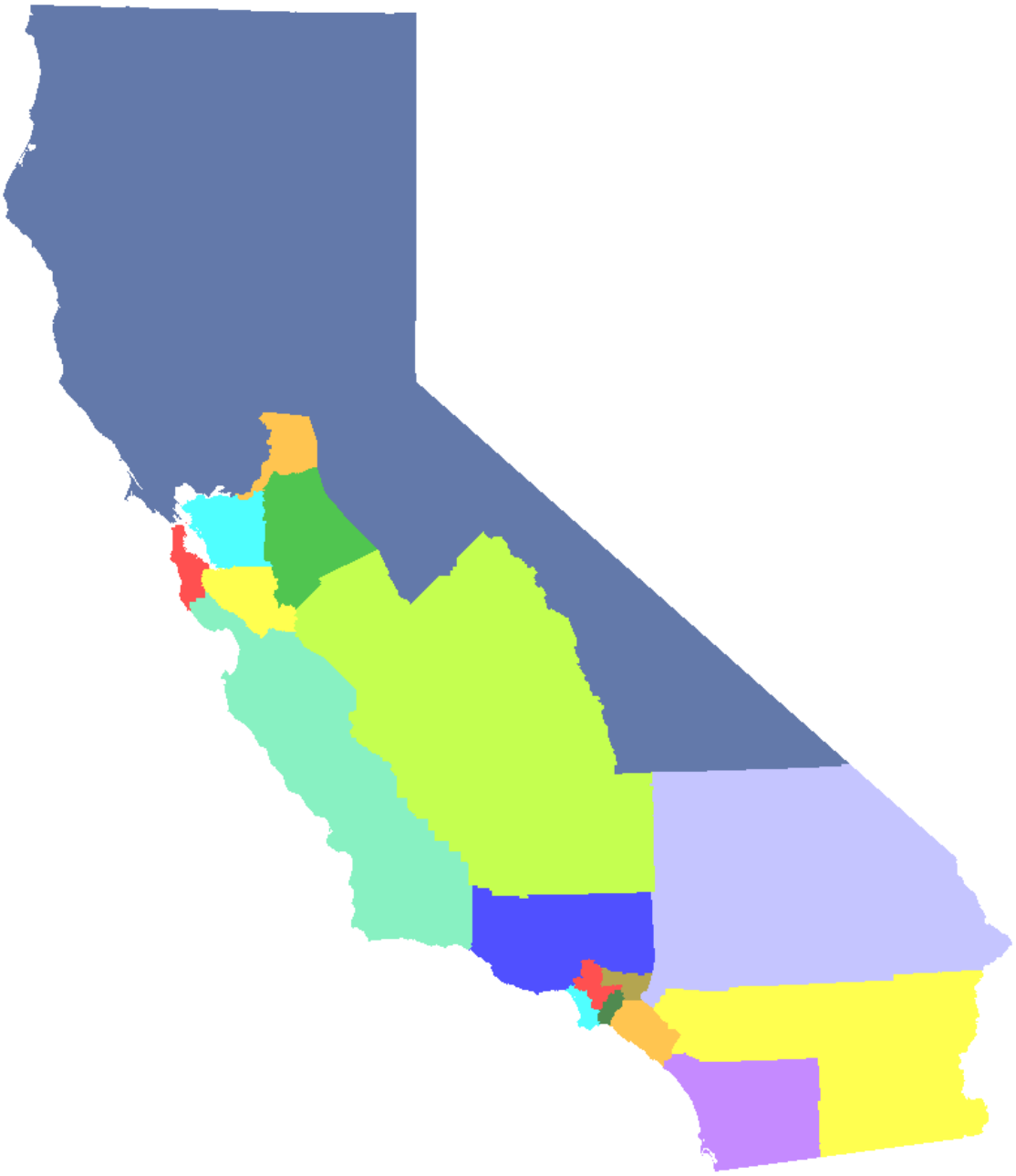


Figure 13: Intensity Classification Areas for Intensity Type 8

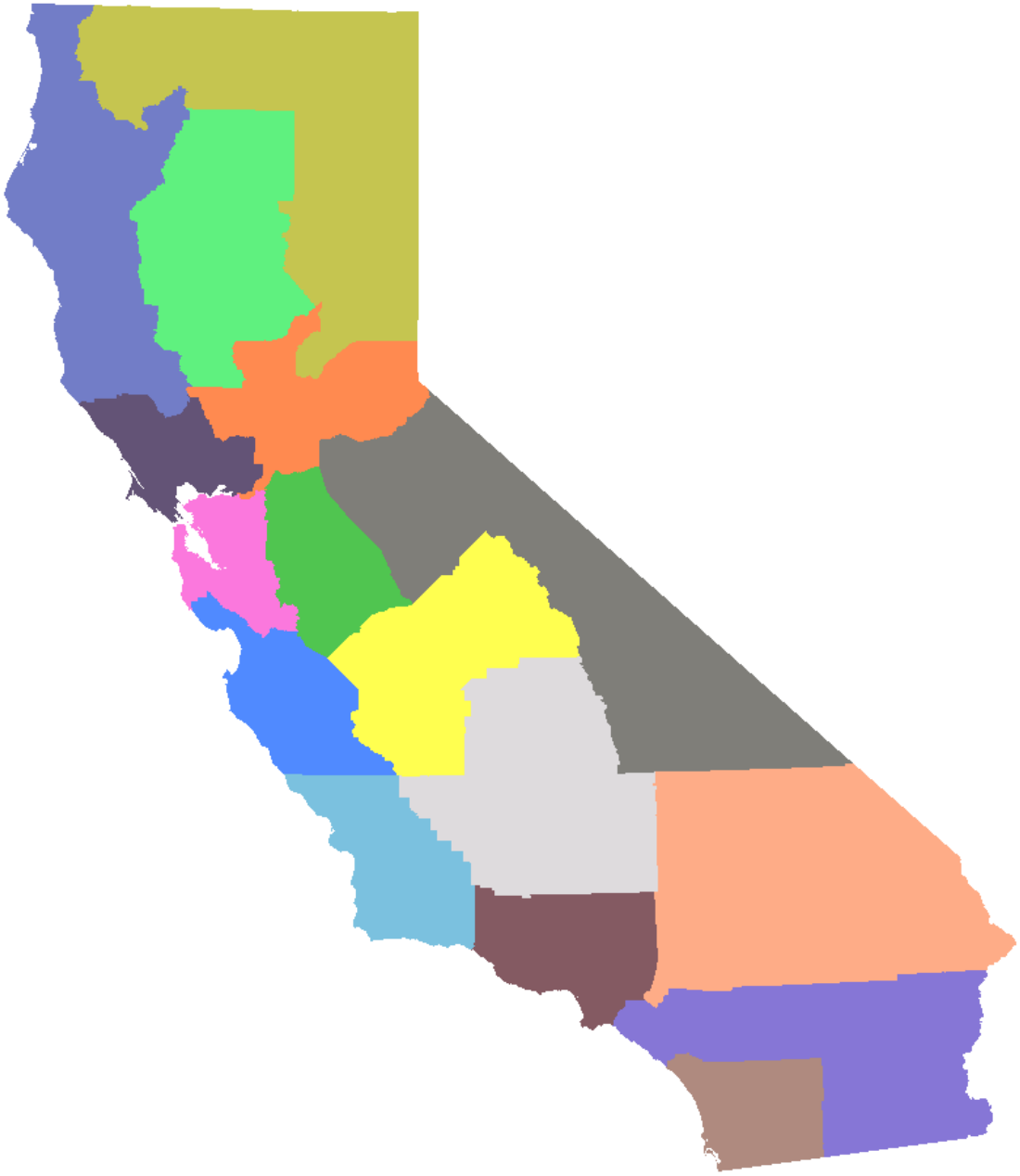


Figure 14: Intensity Classification Areas for Intensity Type -3

Of course, not every zone will grow significantly in the future. If the total growth is less than 50% of the base year, the existing employment breakdowns are used. If there are fewer than 100 workers in a given industry, then the occupation breakdown for that industry are "smoothed" by adding 100 workers distributed to occupations at the average rate for the zone's Intensity Type and Intensity Classification Area.

If the growth exceeds the 50% threshold and the zone changes to a new Intensity Type, then the employment breakdown for the new Intensity Type for the Intensity Classification Area with the zone are used.

If a zone loses employment in an industry, then the existing mix should be used as is precisely, to avoid negative employment in an occupation.

The method described above was implemented using a spreadsheet. First, a table of the occupation breakdowns by industry was created for each zone based on the 2008 synthetic employment. Next, groups of TAZ that have the same intensity type and intensity classification area were created, and a set of employment breakdowns was created based on the cumulative proportions of the TAZ in the group. When creating the occupation breakdowns, any industry that had fewer than 100 employees in a group, was “smoothed” as described above.

Next, the calculation spreadsheet was populated with a row for each TAZ with columns for population and employment by occupation for the base year and employment by industry for all time horizons. Each TAZ’s employment density and employment/population ratio were determined, which were then used to classify the TAZ with an intensity type as previously shown in Table 13. An intensity classification area was then attached to the TAZ based on its location.

After these preliminary values were assigned for each zone a series of conditions are tested to assign the appropriate occupation breakdown. If a TAZ’s employment change for an industry was negative (a reduction of that sector), the base breakdown for that

industry was used. If the total employment in a TAZ changed by more than 50% from the previous base or horizon year, and the intensity type changed, the zone will use the occupation distribution by industry from the group of zones with the new intensity type that are in the same intensity classification area. If either of those conditions is not met (total employment change is less than 50% or intensity type is the same), the zone will use the existing occupation breakdowns.

Now, having determined the appropriate occupation breakdowns for each TAZ by industry, the future employment by industry is multiplied by occupation distribution for that industry to yield the future jobs by occupation for each industry. Finally, all the jobs in each industry are summed by occupation for each zone to yield the employment by occupation for the zone.

Note that the military category is excluded from this process as it is both an industry and occupation category with a one-to-one relationship of industry to occupation.

References

California Department of Finance (2010). *Interim Data Projections.* Sacramento, CA: Department of Finance Demographic Research Unit. Retrieved from: http://www.dof.ca.gov/research/demographic/reports/projections/interim/

California Department of Finance (2007). *Population Projections by Race/Ethnicity for California and its Counties 2000-2050 (Report 06 P-1)*. Sacramento, CA: Department of Finance Demographic Research Unit.

Cambridge Systematics, Inc. (2007). *Bay Area/California High-Speed Rail Ridership and Revenue Forecasting Study: Ridership and Revenue Forecasts (Final Report)*. Sacramento, CA: California High Speed Rail Authority. Retrieved from: http://www.cahighspeedrail.ca.gov/library.aspx

Economic Development Research Group (2011). *TREDIS: California Import and Export Commodity Data.* Boston, MA.

GreenInfo Network (2009). *California Protected Areas Database (1.3*). Retrieved from: www.calands.org

Lehmer, Eric W. (2010). *Base Parcel Layer for California (Private Correspondence).* Urban Land Use and Transportation Center, University of California, Davis.

Moody's Analytics. (2011). *Demographic data 1987-2041.*

Parsons Brinckerhoff. (2011). *Wine County Travel Demand Model Project*: *Model Development Report*. Ukiah, CA: Lake County/City Area Planning Council. Retrieved from: http://www.lakeapc.org/docs/Model%20Development%20Report%20WCTDM%20FINAL%202011-02-01.pdf

Thorne, J.H., J. Bjorkman and N. Roth (2011). *Projecting Growth in California (2000 -2050) Under Six Alternative Policy Scenarios and Assessing Impacts to Future Dispersal Corridors, Fire Threats and Climate-Sensitive Agriculture (*CEC-500-2011 -XXX)*.* Sacramento, CA: California Energy Commission. In review.

Urban Land Use and Transportation Center (ULTRANS). (2011). *California Statewide Travel Demand Model (Final Documentation)*. University of California, Davis. Retrieved from: http://ultrans.its.ucdavis.edu/doc/cstdm-documentation

U.S. Census Bureau. (2010). *American Community Survey (ACS)*. Washington D.C. Retrieved from: http://www.census.gov/acs/www/

U.S. Census Bureau. (2000). *Public-Use Microdata Sample (PUMS)*. Washington D.C. Retrieved from: http://www.census.gov/main/www/pums.html

U.S. Census Bureau. (2000). *Summary File 3 (SF3)*. Washington D.C. Retrieved from: www.census.gov/prod/cen2000/doc/sf3.pdf

World Imagery. (2011). *Map Service by ESRI*. Retrieved from: http://services.arcgisonline.com/ArcGIS/rest/services/World_Imagery/MapServer

Appendix A: Reference List of Forecast Data Files and Reports (MPO/RTPA)

MPO/RTPA	Model Years	Data Files used
MTC	2010, 2020, 2035	tazData2010, tazData2020.csv, tazData2035.csv, bayarea_rtaz1454_rev1.shp
SANDAG	2008, 2020, 2035, 2050	emp_sr12_taz.xls, housing_sr12_taz.xls, income_sr12_taz.xls, mgra12.shp, TAZ12.shp
AMBAG	2005, 2035	AMBAG_TAZ_2005_2035.shp, AMBAG_TAZ_2005_2035.xlsx, 3_Model_Technical_Report_Draft_March2011.pdf
SLOCOG	2008, 2020, 2035	Final SLOCOG Model Development Report.pdf, DEMOGRAPHICS_2008.dbf, DEMOGRAPHICS_2020.dbf, DEMOGRAPHICS_2035.dbf, SLO-TAZ-vL.shp
BCAG	2006, 2035	BCAG Model Development Report_no_appendices.pdf, Projections_2006-2035report.pdf, BCAG_Traffic_Model_Land_Use_2006.shp, BCAG_Traffic_Model_Land_Use_2035.shp
Tuolumne	2010, 2020, 2030, 2040, 2050	LU_summaries_03_09_2011.xls, TAZ_Boundaries_Jan_2010.shp
Shasta	2005, 2010, 2020, 2030, 2050	SH10_LU.DBF, SH20_LU.DBF, SH30_LU.DBF, SH50_LU.DBF, 080424 Shasta Model Guide.doc, LU_Summary_080208.xls, Shasta_TAZ_060926.shp
ACTC (Amador)	2010, 2030, 2050	DRAFT_Model Development Report.doc, Amador_Demographics_2010.xls, Amador_Demographics_2030.xls, Amador_Demographics_2050.xls
Fresno	2008, 2035	FC_MASTER(9-13-10).zip, fcog2020mb.zip, FCOG2035mb.zip
Kings Co	2010, 2015, 2020, 2025, 2030, 2035	KCAG_LU_SJVformat_20110505.xls, TAZ_041211.shp, KCAG05_FINAL_LandUse_2010_04_29.xls, MASTER_1000TAZ_2007_RTP_2009_10_02.shp
Kern COG	2006, 2020, 2035	Kern 306_2006_Regional_Travel_Demand_Model.pdf, Modeling Parameters_050311_kern.xlsx, TAZ_MIP_Kern_050211.shp, KE06SOCIO.DBF, KE20SOCIO.DBF, KE35SOCIO.DBF
Madera	2005, 2020, 2035, 2050	Parameters_Madera_110729.xlsx, MaderaTAZ#Revised_04042011.shp, Madera2030TAZ.shp, Madera_TripGen_2005.xls, Madera_TripGen_2020.xls, Madera_TripGen_2035.xls
Merced	2010, 2020, 2035	TAZ_3_Final.shp, 3County_BaseYear_MIP_Land_Use.xls, LandUse2000.dbf, LandUse2010.dbf, LandUse2020.dbf, LandUse2035.dbf, TAZ.shp
San Joaquin	2010, 2020, 2030	TAZ_3_Final.shp, 3County_BaseYear_MIP_Land_Use.xls, 2006_LAND_USE.csv, 2020_LAND_USE.csv, 2030_LAND_USE.csv, taz_2006.shp
Stanislaus	2006, 2030	TAZ_3_Final.shp, 3County_BaseYear_MIP_Land_Use.xls, tppg2006.DBF, tppg2030.DBF
Tulare	2007, 2020, 2030, 2035	TAZ_Renumber_041211.shp, TCAG_LU_SJVformat_20110505.xls, TC20_LU.DBF, TC35_LU.DBF, TC07_LU.DBF
SACOG	2008, 2035	SACOG02003.pdf, 2008zbas.dbf, 2008hhmv.dbf, 2035zbas.dbf, 2035hhmv.dbf, TAZ07_w_tahoe.shp
SBCAG	2000, 2030	SBCAG ReginalGrowthforecastComplete Final.pdf, 2000 SB TAZ.shp, 2000 TAZData.xls, 2030 SB TAZ.shp, 2030 TAZData.xls,
SCAG	2008, 2020, 2035	T2_2008_control_RTP12PL94_042111.csv, T2_2020_control_RTP12PL94_071011.csv, T2_2035_control_RTP12PL94_071011.csv, TAZ_field_descriptions.xlsx, Tie2_taz_scag_2010.shp, MVS03_Chap02 - Socio-Demographic.pdf, SCAG TransCAD New Model File and Parameter Documentation.pdf
Lake & Mendocino (Wine Country)	2009, 2020, 2030	LAND_USE_2009.DBF, LAND_USE_2030.DBF, LAND_USE_2020.DBF, TAZ.shp, WC_TAZ.shp, Model Development Report WC-TDM FINAL 2011-02-02.pdf, WC-TDM Appendices FINAL 2011-02-02.pdf

Appendix B: List of Forecast Data by Attributes and Time Horizon Years Provided by California MPOs and RTPAs

This table shows the relevant available zonal data available from each of the California MPOs and RPTAs. This data is all originally at the MPO/RTPA TAZ-level. Section 2 describes how this data was transformed for use in the CSTDM. “Counties” lists which counties make up the region. “Base Year” and “Time Horizons” show what years are represented in the regional model. “Population” shows how the region categorizes data at the person-level, and “Households” how data is categorized at the household level. There is an additional table below that shows the full set of age categories for SANDAG and SBCAG. When “cross-class:…” is indicated this means that the following categories interact with one another to make a set of combined categories. For example, in Fresno, “Cross-Class: SF, MF, Autos (0,1,2+)” indicates the household categories include single-family with zero autos, single-family with one auto, single-family with two or more autos, multi-family with zero autos, and so on. Most MPOs provided income values in year-2000 dollars, which is the monetary unit of the CSTDM, with exception of SACOG, which has income in terms of 1990 dollars. “School enrollment” shows what school enrollment categories are available. “Employment” shows the categorization of employment in the regional model. Note that Appendix E shows more detailed information regarding the MPO/RTPA employment categories and how they relate to CSTDM categories. An additional table of abbreviations has been provided following the main table.

Summary of Zonal Data from MPOs and RTPAs

MPO/RTPA	Counties	Base Year	Time Horizons	Population
ACTC	Amador	2010	2030, 2050	N/A
AMBAG	Monterey, San Benito, Santa Cruz	2005	2035	Pop
BCAG	Butte	2006	2035	N/A
Fresno	Fresno	2008	2035	Pop
Kern COG	Kern	2006	2020, 2035	GQ
Kings Co	Kings	2010	2015, 2020, 2025, 2030, 2035	Pop
Madera	Madera	2005	2020, 2035, 2050	N/A
Merced	Merced	2010	2020, 2035	N/A
MTC	Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, Sonoma	2010	2020, 2035	Pop, Age (0-4,5-19,20-44,45-64,65+), GQ
SACOG	El Dorado, Placer, Sacramento, Sutter, Yolo, Yuba [excluding Tahoe]	2008	2035	N/A
San Joaquin	San Joaquin	2010	2020, 2030	N/A
SANDAG	San Diego	2008	2020, 2035, 2050	Pop, Age (x20), GQ
SBCAG	Santa Barbara	2000	2030	Pop, Age (x16)
SCAG	Imperial, Los Angeles, Orange, Riverside, San Bernardino, Ventura	2008	2020, 2035	Pop, Age (5-17,18-24,16-64,65+)
Shasta	Shasta	2005	2010, 2020,2030, 2050	Pop
SLOCOG	San Luis Obispo	2008	2020, 2035	N/A
Stanislaus	Stanislaus	2006	2030	Pop
Tulare	Tulare	2007	2020, 2030, 2035	N/A
Tuolumne	Tuolumne	2010	2020, 2030, 2040, 2050	N/A
Wine Country	Lake, Mendocino	2009	2020, 2030	N/A

MPO/ RTPA	Households	School Enrollment	Employment
ACTC	SF, MF	K-8, HS	Various land use categories in KSF and other units
AMBAG	Total	N/A	Retail, Service, Ag., Industrial, Const., Govt.
BCAG	Cross-Class: Rural, Suburban, Urban, SF, MF, MH	K-12, PSE	Various land use categories in KSF and other units
Fresno	Cross-Class: SF, MF, Autos (0,1,2+)	N/A	Retail, Service, Edu., Govt., Other
Kern COG	Total	K-8, HS, PSE	RHRET, RMRET, SCSER, SOSER, Basic, BWOTH
Kings Co	Cross-Class: SF, MF, Autos (0,1,2+)	N/A	Retail, Office, Edu., Industrial, Ag., Govt., Other
Madera	SF, MF	N/A	Retail, Office, Edu., Manuf., Govt., Other
Merced	SF, MF	Total	Retail, Office, Edu., Ag., Industrial
MTC	SF, MF, Workers (0,1,2,3+), Income (<$30k,$30-60k,$60-100k,>$100k)	HS, PSE	Retail, Finance/Prof., Health/Edu., Rec., Ag., Manuf., Other
SACOG	SF, MF, MH, Cross-class: Persons (1,2,3,4+), Workers (0,1,2,3+), Income (<$10,$10-20,$20-35,$35-50,>$50 in 1990 K$)	K-12, PSE	Retail, Office, Medical, Edu, Manuf., Other
San Joaquin	SF, MF	N/A	Retail, Service, Other
SANDAG	SF, MF, MH, Income (10)	N/A	Retail, Finance/Real Estate, IT, Prof./Bus., Leisure/Hosp., Other, Health/Edu., Military, Ag., Const., Manuf., Wholesale, Trans./Warehoue, Govt., Self Employed
SBCAG	Total, Income (x16)	N/A	Total
SCAG	SF, MF, Persons (1,2,3,4+), Workers (0,1,2,3+), Income (<$25k, $25-50k, $50-100k, >$100k)	K-12, PSE	Retail, IT, Finance/Real Estate, Prof., Art/Entertainment, Other Service, Edu., Ag., Const., Manuf., Wholesale, Trans., Public Administration
Shasta	SF, MF, MH, MH-park, Senior, Recreational	N/A	Retail, Retail-high, Retail Big Box, Office, Hotel, Restaurant, Restaurant-high, Institutional, Hospital, Medical Office, Residential Care, Child Care, School, College, Casino, Govt., Govt.-high
SLOCOG	SF, MF, MH, Rural	K-8, HS, PSE	Retail, Non-retail, various land uses in KSF
Stanislaus	SF, MF	N/A	Retail, Service, Edu., Govt., Other
Tulare	Cross-Class: SF, MF, Autos (0,1,2+)	N/A	Retail, Office, Service, Edu., Govt., Ag., Industrial
Tuolumne	SF, MF, MH, Condo, Multiple, Rural SF, Rural MF, Rual MH	K-12	Various land use categories in KSF and other units
Wine Country	SF, MF, Income (<$35k,35-100k,>$100k)	K-8, HS, PSE	Retail, Health, Ag., Leisure/Rec., Industrial, Other

Abbreviations used in the Appendix B Table

Household	
SF	Single Family
MF	Multi Family
MH	Mobile Home
School Enrollment	
HS	High School Students
PSE	Post Secondary Students
Employment	
IT	Information Technology
Prof.	Professional
Edu.	Education
Rec.	Recreational
Ag.	Agrigcultural
Manuf.	Manufacturing
Bus.	Business
Hosp.	Hospitality
Const.	Construction
Trans.	Transportation
Govt.	Government

Age Category Detail

MPO	Age Categories
SANDAG	0-4,5-9,10-14,15-17,18-19,20-24,25-29,30-34,35-39,40-44,45-49,50-54,55-59,60-61,62-64,65-69,70-74,75-79,80-84,85+
SBCAG	0-4,5-9,10-14,15-19,20-24,25-34,35-44,45-54,55-59,60-64,65-74,75-84,85+18+21+62+

Appendix C: Python Scrip for Aggregating Primary Forecast Data to CSTDM TAZ

```
import arcpy, os
##### Set local variables #####
# geoprocessor variables
arcpy.env.workspace = r'E:\Projects\HSRA_Forecasting\FinishedSplitting\SLOCOG_2\input'
arcpy.env.overwriteOutput = True

# other workspaces operations...
os.chdir(arcpy.env.workspace)
output_workspace = os.path.abspath('../output')          # the script assumes that the output directory already exists

# --------------- Input Data ---------------
MPO_Split = 'SLOCOG.shp'                                 # Non-nesting (Split) MPO TAZs
MPO_Nest = 'MPO_nest_table.dbf'                          # Nesting MPO TAZ table
CSTDM_TAZ = 'CSTDM_SLOCOG.shp'                           # CSTDM TAZs covering the MPO area

### input fields ###
# ********** 'Split MPO TAZ' shapefile **********
Spl_CSTDM_TAZ_F = 'TAZ'                                  # field containing the CSTDM TAZ number
Spl_MPO_TAZ_F = 'ZONE'                                   # field containing the MPO TAZ number

# list of fields containing data to be split
base_pop = ['Pop_2008']                                  # base population field
future_pop = [      'Pop_2020',                          # future population fields
                    'Pop_2035'
                    ]
base_emp = ['Emp_2008']                                  # base employment field
future_emp = [      'Emp_2020',                          # future employment fields
                    'Emp_2035'
                    ]
base_Elem = ['Elem_2008']                                # base employment field
future_Elem = [     'Elem_2020',                         # future employment fields
                    'Elem_2035'
                    ]
base_HS = ['HS_2008']                                    # base employment field
future_HS = [       'HS_2020',                           # future employment fields
                    'HS_2035'
                    ]
base_Coll = ['Coll_2008']                                # base employment field
future_Coll = [     'Coll_2020',                         # future employment fields
                    'Coll_2035'
                    ]
field_list = (base_pop + future_pop + base_emp + future_emp + base_Elem + future_Elem +
                    base_HS + future_HS + base_Coll + future_Coll
                    )                    # make the input fields into one list for later use
field_list_nest = ([base_pop] + [future_pop] + [base_emp] + [future_emp] + [base_Elem] + [future_Elem] + [base_HS] +
```

```
                    [future_HS] + [base_Coll] + [future_Coll]
                    )                    # make the input fields into a list, with nested sublists, also for later use
field_type = 'LONG'                      # the script generally assumes that it is processing integer values (people, jobs, etc.)

# split percentage fields
base_pop_split_perc = 'Base_Spl_1'                    # field containing the base year population split percentage
future_pop_split_perc = 'Fut_SplitP'                  # field containing the future growth population split percentage
base_emp_split_perc = 'Base_Spl_3'                    # field containing the base year employment split percentage
future_emp_split_perc = 'Fut_Spli_2'                  # field containing the future growth employment split percentage
base_Elem_split_perc = 'Base_Spl_5'                   # field containing the base year employment split percentage
future_Elem_split_perc = 'Fut_Spli_4'                 # field containing the future growth employment split percentage
base_HS_split_perc= 'Base_Spl_7'                      # field containing the base year employment split percentage
future_HS_split_perc = 'Fut_Spli_6'                   # field containing the future growth employment split percentage
base_Coll_split_perc = 'Base_Spl_9'                   # field containing the base year employment split percentage
future_Coll_split_perc = 'Fut_Spli_8'                 # field containing the future growth employment split percentage

perc_list = [base_pop_split_perc, future_pop_split_perc, base_emp_split_perc, future_emp_split_perc, base_Elem_split_perc,
                    future_Elem_split_perc,     base_HS_split_perc,     future_HS_split_perc,     base_Coll_split_perc,
                    future_Coll_split_perc]              # put the split percentage fields into a list for later use

# ********** 'Nesting MPO TAZ' table **********
Nest_CSTDM_TAZ_F = 'TAZ'                              # field containing the CSTDM TAZ number

nest_base_pop = ['Pop_2008']                          # base population field
nest_future_pop = ['Pop_2020',
                    'Pop_2035'
                    ]                                 # future population fields
nest_base_emp = ['Emp_2008']                          # base employment field
nest_future_emp = ['Emp_2020',
                    'Emp_2035'
                    ]                                 # future employment fields
nest_base_Elem = ['Elem_2008']                        # base employment field
nest_future_Elem = ['Elem_2020',
                    'Elem_2035'
                    ]                                 # future employment fields
nest_base_HS = ['HS_2008']                            # base employment field
nest_future_HS = [  'HS_2020',
                    'HS_2035'
                    ]                                 # future employment fields
nest_base_Coll = ['Coll_2008']                        # base employment field
nest_future_Coll = ['Coll_2020',
                    'Coll_2035'
                    ]                                 # future employment fields

nest_field_list = (nest_base_pop + nest_future_pop + nest_base_emp + nest_future_emp + nest_base_Elem +
                    nest_future_Elem + nest_base_HS + nest_future_HS + nest_base_Coll + nest_future_Coll
```

```
                )                                   # make the input fields into one list for later use

# ********** ' CSTDM TAZ' shapefile **********
CSTDM_TAZ_F = 'TAZ'                                 # field containing the CSTDM TAZ number


# --------------- Output Data ---------------
outfile = "CSTDM_SLOCOG_Total_temp.shp"             # name of output file, to be written to output directory

###############################
### Process data:
# build a dictionary containing the split MPO values, keyed by (CSTDM TAZ, MPO TAZ)
MPO_Raw_dict = {}                                   # initialize empty dictionary to contain raw (unsplit) MPO values
MPO_Split_dict = {}                                 # initialize empty dictionary to contain split MPO values
rows = arcpy.SearchCursor(MPO_Split, "", "",
                        Spl_CSTDM_TAZ_F + ';' + Spl_MPO_TAZ_F + ';' + ';'.join(field_list) + ';' + ';'.join(perc_list)
                        )                           # create cursor to read all rows in input file
for row in rows:                                    # iterate through the rows in the cursor
        CSTDM_ID = row.getValue(Spl_CSTDM_TAZ_F)    # get current CSTDM TAZ
        MPO_ID = row.getValue(Spl_MPO_TAZ_F)        # get current MPO TAZ
        split_values = []                           # initialize empty list to contain split percentages
        for perc in perc_list:
                split_values.append(row.getValue(perc))# get values for each split percentage field
        raw_values = map(lambda x: row.getValue(x), field_list)     # add the raw (unsplit) values to a list, then...
        MPO_Raw_dict[MPO_ID] = raw_values           # add that list to MPO_Raw_dict, keyed by MPO ID
        MPO_Split_dict[(CSTDM_ID, MPO_ID)] = []     # add keys to the split dictionary, pointing to empty lists
        n = 0                                       # initialize a counter
        for value in split_values:                  # iterate through the items in the split values list
                if n % 2 == 0:                      ### if it is in an even numbered position, it is assumed to be a
                                                    ### base year value <--- !!!!! IMPORTANT !!!!!
                        new_values = map(lambda x: int(round(row.getValue(x) * (0.01 * value))),
                                        field_list_nest[n]
                                        )           # multiply MPO values (pop, emp, etc.) by the base split
                                                    # percentage, then round and convert them to integers
                        base_temp = new_values[:]   # hold the base value for use in the next step
                else:                               ### if it is in an odd numbered position, it is assumed to be a
                                                    ### forecast year value <--- !!!!! IMPORTANT !!!!!
                        future_values = map(lambda x: row.getValue(x), field_list_nest[n])
                        base_values = map(lambda x: row.getValue(x), field_list_nest[n-1] * len(future_pop))
                        diff_values = map(lambda x,y: x - y,
                                        future_values,
                                        base_values
                                        )           # find the difference between the current and future
                                                    # values (the amount of growth/decline)
                        new_values = []             # create an empty list to hold the new calculated values
                        j = 0
```

```
                    for diff_value in diff_values:          # for each of the future values...
                        if diff_value >= 0:                 # if the difference is positive(growth) then...
                            split_diff_value = int(round(diff_value * (0.01 * value)))      # multiply the
                                                            # difference by the future split percentage,
                                                            # then round and convert it to an integer
                            new_values.append(base_temp[0] + split_diff_value)      # add the
                                                            # calculated growth to the base year number
                        elif diff_value < 0:                # if the difference is negative(decline) then...
                            split_diff_value = int(round(future_values[j] * (0.01 * split_values[n-1])))
                                                            # multiply the value by the base split percentage, then
                                                            # round and convert it to an integer
                            new_values.append(split_diff_value)
                        j += 1
            MPO_Split_dict[(CSTDM_ID, MPO_ID)].extend(new_values) # add the new values to the list in the
                                                                  # correct dictionary item
            n += 1                                                # advance the counter

# re-aggregate the split MPO values for use in the next section
MPO_SumCheck_dict = {}                  # initialize empty dictionary to contain split MPO values, which
                                        # have been re-aggregated by MPO TAZ
for key in MPO_Split_dict:
    CSTDM_ID = key[0]                   # get current CSTDM TAZ
    MPO_ID = key[1]                     # get current MPO TAZ
    if MPO_ID in MPO_SumCheck_dict:     # if there is already a record for the MPO TAZ then...
                                        # add the MPO value to the existing record
        MPO_SumCheck_dict[MPO_ID] = [(x + y) for x,y in zip(MPO_SumCheck_dict[MPO_ID],MPO_Split_dict[key])]
    else:                               # if not, then ...
        MPO_SumCheck_dict[MPO_ID] = MPO_Split_dict[key]     # create a record for that MPO TAZ

##### The following section of code compares the original values to the values that have been split and then #####
##### re-aggregated, to make sure that people (or households, jobs, etc.) have not been lost or added during #####
##### the splitting process. If there are differences found, then they are corrected by iteratively adding or #####
##### subracting 1 unit from the largest subsection. #####

log_file = open(os.path.join(output_workspace,'SplitError_log.txt'), 'w')
log_file.write('(CSTDM, MPO)\tNew Values\tRunning Total of Changes\tIncrement\n')
ErrorLog_dict = {}
for key in MPO_SumCheck_dict:
    ErrorLog_dict[key] = [0] * len(MPO_SumCheck_dict[key])

run = True                              # initialize run to True
while run == True:                      # this loop will run until no differences are found
    run = False                         # after entering the first loop, set run to False so that the loop will end after
                                        # the first time through if no differences are found
    for key in MPO_SumCheck_dict:                   # loop through all of the re-aggregated records
        for n in range(len(MPO_SumCheck_dict[key])): # loop through each value in the re-aggregated records
```

```
                    if MPO_SumCheck_dict[key][n] > MPO_Raw_dict[key][n]:    # if there is a (greater than) difference
                                                                            # between the re-aggregated and the
                                                                            # original value, then...
                        high_value = -9999
                        for k,v in MPO_Split_dict.items():                  # find the largest split area...
                            if k[1] == key and v[n] >= high_value:
                                high_value = v[n]
                                high_value_key = k[0]
                        MPO_Split_dict[(high_value_key, key)][n] -= 1       # and subtract 1 unit from it...
                        MPO_SumCheck_dict[key][n] -= 1          # also, subtract 1 unit from the same area in
                                                                # the re-aggregated dictionary, so that the next
                                                                # time it is checked, the correction will be
                                                                # represented
                        ErrorLog_dict[key][n] -= 1
                        log_file.write(str((high_value_key, key)) + '\t' + str(MPO_Split_dict[(high_value_key,
                                key)]) + '\t' + str(ErrorLog_dict[key]) + '\t' + str(n) + ' -1' + '\n')
                        run = True              # since a difference was found, set run to True, so that
                                                # every value will be checked again
                    elif MPO_SumCheck_dict[key][n] < MPO_Raw_dict[key][n]: # if there is a (less than) difference
                                                # between the re-aggregated and the original value,
                                                # then do the same as above, but add a unit...
                        high_value = -9999
                        for k,v in MPO_Split_dict.items():
                            if k[1] == key and v[n] >= high_value:
                                high_value = v[n]
                                high_value_key = k[0]
                        MPO_Split_dict[(high_value_key ,key)][n] += 1
                        MPO_SumCheck_dict[key][n] += 1
                        ErrorLog_dict[key][n] += 1
                        log_file.write(str((high_value_key, key)) + '\t' + str(MPO_Split_dict[(high_value_key,
                                key)]) + '\t' + str(ErrorLog_dict[key]) + '\t' + str(n) + ' +1' + '\n')
                        run = True
log_file.write('\nSummary of corrections made:\n')
for key in ErrorLog_dict:
        if sum(ErrorLog_dict[key]) != 0:
                log_file.write(str(key) + '\t' + str(ErrorLog_dict[key]) + '\n')

log_file.close()

##############################
##############################

# copy the split MPO TAZs to the output workspace
result = arcpy.Copy_management(os.path.abspath(MPO_Split), os.path.join(output_workspace, "SplitMPO_TAZ.shp"))
Split_TAZ_working = result.getOutput(0)
```

```
# add the fields from the input field list to the working Split TAZ shapefile
for field in field_list:
        arcpy.AddField_management(Split_TAZ_working, 's' + field, field_type)

# use the Split TAZ dictionary to update the fields in the Split TAZ file
rows = arcpy.UpdateCursor(Split_TAZ_working)                    # create cursor to update all rows in shapefile
for row in rows:
        CSTDM_ID = row.getValue(Spl_CSTDM_TAZ_F)                # get current CSTDM TAZ
        MPO_ID = row.getValue(Spl_MPO_TAZ_F)                    # get current MPO TAZ
        n = 0
        for field in field_list:
                row.setValue('s' + field, MPO_Split_dict[(CSTDM_ID,MPO_ID)][n])     # set the value in the current row and
                                                                                    # field to the value from the CSTDM
                                                                                    # TAZ dictionary
                n += 1
        rows.updateRow(row)                                                         # commit the change

del row                                                         # delete the row and...
del rows                                                        # cursor objects to free any locks on the table

# aggregate the split MPO values by CSTDM TAZ
MPO_Split_Sum_dict = {}                                         # initialize empty dictionary
for key in MPO_Split_dict:
        CSTDM_ID = key[0]                                       # get current CSTDM TAZ
        if CSTDM_ID in MPO_Split_Sum_dict:              # if there is already a record for the current CSTDM TAZ then...
                MPO_Split_Sum_dict[CSTDM_ID] = [(x + y) for x,y in zip(MPO_Split_Sum_dict[CSTDM_ID],
                                MPO_Split_dict[key])]           # add the split MPO values to the existing values
        else:                                                   # if there is no record for the current CSTDM TAZ then...
                MPO_Split_Sum_dict[CSTDM_ID] = MPO_Split_dict[key]     # create a new record for the split MPO values

# build a dictionary containing the summed values of the nesting MPO TAZs, keyed by CSTDM TAZ
MPO_Nest_dict = {}                                                      # initialize empty dictionary
rows = arcpy.SearchCursor(MPO_Nest, "", "", Nest_CSTDM_TAZ_F + ';' + ';'.join(nest_field_list))      # create cursor to read all
                                                                                                     # rows in shapefile
for row in rows:                                                        # iterate through the rows in the cursor
        CSTDM_ID = row.getValue(Nest_CSTDM_TAZ_F)                                                    # get current CSTDM TAZ
        nest_values = []
        for field in nest_field_list:
                nest_values.append(row.getValue(field))# get the values of all the fields in the nesting MPO TAZs
        if CSTDM_ID in MPO_Nest_dict:                   # if there is already a record for the current CSTDM TAZ then...
                                                        # add the MPO values to the existing values
                MPO_Nest_dict[CSTDM_ID] = [(x + y) for x,y in zip(MPO_Nest_dict[CSTDM_ID], nest_values)]
        else:                                           # if there is no record for the current CSTDM TAZ then...
                MPO_Nest_dict[CSTDM_ID] = nest_values           # create a new record for the MPO values

# copy the CSTDM TAZs to the output workspace
```

```
result = arcpy.Copy_management(os.path.abspath(CSTDM_TAZ), os.path.join(output_workspace, outfile))
CSTDM_TAZ_working = result.getOutput(0)

# add the fields from the input field list to the working CSTDM TAZ shapefile
for field in field_list:
        arcpy.AddField_management(CSTDM_TAZ_working, field, field_type)

# build a dictionary containing the summed values of the nested and non-nested TAZs, keyed by CSTDM TAZ
CSTDM_TAZ_dict = {}                                         # initialize empty dictionary
rows = arcpy.SearchCursor(CSTDM_TAZ_working, "", "", CSTDM_TAZ_F)      # create cursor to read all rows in shapefile
for row in rows:                                                       # iterate through the rows in the cursor
        CSTDM_ID = row.getValue(CSTDM_TAZ_F)                           # get current CSTDM TAZ
        if CSTDM_ID not in MPO_Nest_dict:                   # if there is no record in the nested dictionary, then...
                CSTDM_TAZ_dict[CSTDM_ID] = MPO_Split_Sum_dict[CSTDM_ID]   # just use the list in the split dictionary
        elif CSTDM_ID not in MPO_Split_Sum_dict:            # if there is no record in the split dictionary, then...
                CSTDM_TAZ_dict[CSTDM_ID] = MPO_Nest_dict[CSTDM_ID]    # just use the list in the nested dictionary
        else:
                CSTDM_TAZ_dict[CSTDM_ID] = [(x + y) for x,y in zip(MPO_Nest_dict[CSTDM_ID],
                        MPO_Split_Sum_dict[CSTDM_ID])]      # otherwise, sum the values for both the nested and
                                                            # non-nested TAZs

# use the CSTDM TAZ dictionary to update the fields in the CSTDM file
rows = arcpy.UpdateCursor(CSTDM_TAZ_working)                # create cursor to update all rows in shapefile
for row in rows:
        CSTDM_ID = row.getValue(CSTDM_TAZ_F)                           # get current CSTDM TAZ
        n = 0
        for field in field_list:
                row.setValue(field, CSTDM_TAZ_dict[CSTDM_ID][n])       # set the value in the current row and field to
                                                                       # the value from the CSTDM TAZ dictionary
                n += 1
        rows.updateRow(row)                                            # commit the change

del row                                                                # delete the row and...
del rows                                                               # cursor objects to free any locks on the table
```

Appendix D: Python Scrip for Aggregating Primary Forecast Data to CSTDM TAZ

```
import arcpy, os
##### Set local variables #####
# geoprocessor variables
arcpy.env.workspace = 'E:/Projects/HSRA_Forecasting/FinishedSplitting/Shasta/working_08252011/input'
arcpy.env.overwriteOutput = True

# other workspaces operations...
os.chdir(arcpy.env.workspace)
output_workspace = os.path.abspath('../output')          # the output directory must already exist

# --------------- Input Data ---------------
Subclass_Table_base = 'SH08_LU.dbf'                      # Table containing the base subclass totals to be allocated
Subclass_Table_future = 'SH50_LU.dbf'                    # Table containing the future subclass totals to be allocated
MPO_Split = 'Shasta.shp'                                 # Split MPO TAZs
MPO_Nest = 'Shasta_TAZ_nest_table.dbf'                   # Nesting MPO TAZs
CSTDM_TAZ = 'CSTDM_Shasta.shp'                           # CSTDM TAZs covering the MPO area

### input fields ###
# **********   'Base Subclass' table   **********
SubCl_MPO_TAZ_F_base = 'TAZ'                             # field containing the MPO TAZ number

# list of fields containing data to be subdivided
field_list_base = [   'EMP_INDSTL',
                      'EMP_WHOLSL',
                      'EMP_SRVCOM',
                      'EMP_RETAIL',
                      'EMP_RETLHI',
                      'EMP_RETWAR',
                      'EMP_OFFICE',
                      'EMP_SCHOOL',
                      'EMP_COLLGE',
                      'EMP_MEDOFF',
                      'EMP_HOSPTL',
                      'EMP_RESCAR',
                      'EMP_CHLDCR',
                      'EMP_DEVREC',
                      'EMP_CASINO',
                      'EMP_HOTEL',
                      'EMP_RESTNT',
                      'EMP_RESTHI',
                      'EMP_INST',
                      'RMP_GOVT',
                      'EMP_GOVTHI'
                      ]
```

```
# **********  'Future Subclass' table  **********
SubCl_MPO_TAZ_F = 'TAZ'                    # field containing the MPO TAZ number

# list of fields containing data to be subdivided
field_list = [          'EMP_INDSTL',
                        'EMP_WHOLSL',
                        'EMP_SRVCOM',
                        'EMP_RETAIL',
                        'EMP_RETLHI',
                        'EMP_RETWAR',
                        'EMP_OFFICE',
                        'EMP_SCHOOL',
                        'EMP_COLLGE',
                        'EMP_MEDOFF',
                        'EMP_HOSPTL',
                        'EMP_RESCAR',
                        'EMP_CHLDCR',
                        'EMP_DEVREC',
                        'EMP_CASINO',
                        'EMP_HOTEL',
                        'EMP_RESTNT',
                        'EMP_RESTHI',
                        'EMP_INST',
                        'RMP_GOVT',
                        'EMP_GOVTHI'
                        ]
field_type = 'LONG'             # the script generally assumes that it is processing integer values (people, households, etc.)
                                # the field type entered here is actually only used for the output table, so as long as the input
                                # can be interpreted as an integer, it doesn't matter what it actually is


# ********** 'Split MPO TAZ' shapefile **********
Spl_CSTDM_TAZ_F = 'TAZ'                    # field containing the CSTDM TAZ number
Spl_MPO_TAZ_F = 'ZONE'                     # field containing the MPO TAZ number
fraction_field_base = 'BaseEmpPrc'         # proportion that each fragment should receive of the base number
fraction_field_growth = 'SplitPercE'       # proportion that each fragment should receive of the new growth


# **********  'Nesting MPO TAZ' table  **********
Nest_CSTDM_TAZ_F = 'TAZ'                   # field containing the CSTDM TAZ number
Nest_MPO_TAZ_F = 'ZONE'                    # field containing the MPO TAZ number


# **********   'CSTDM TAZ' shapefile   **********
CSTDM_TAZ_F = 'TAZ'                        # field containing the CSTDM TAZ number
```

```
# --------------- Output Data ---------------
outfile = "CSTDM_Shasta_2050_EmpByCategory.shp"        # name of output file, to be written to output directory


################################
# Process data:

# build a dictionary containing the base MPO values to be subdivided, keyed by MPO TAZ
Subclass_Base_Table_dict = {}                          # initialize empty dictionary
                                                       # create cursor to read all rows in table
rows = arcpy.SearchCursor(Subclass_Table_base, "", "", SubCl_MPO_TAZ_F_base + ';' + ';'.join(field_list_base))
for row in rows:                                       # iterate through the rows in the cursor
        Subclass_Base_Table_dict[row.getValue(SubCl_MPO_TAZ_F_base)] = []        # create keys from the MPO TAZ
                                                                                 # numbers, pointing to empty lists
        for field in field_list_base:
                                                       # fill the lists with the (base) values to be subdivided
                Subclass_Base_Table_dict[int(row.getValue(SubCl_MPO_TAZ_F_base))].append(int(row.getValue(field)))

# build a dictionary containing the future MPO values to be subdivided, keyed by MPO TAZ
Subclass_Table_dict = {}
                                                       # initialize empty dictionary
                                                       # create cursor to read all rows in table
rows = arcpy.SearchCursor(Subclass_Table_future, "", "", SubCl_MPO_TAZ_F + ';' + ';'.join(field_list))
for row in rows:                                       # iterate through the rows in the cursor
        Subclass_Table_dict[row.getValue(SubCl_MPO_TAZ_F)] = []          # create keys from the MPO TAZ numbers,
                                                                         # pointing to empty lists
        for field in field_list:
                                                       # fill the lists with the values to be subdivided
                Subclass_Table_dict[int(row.getValue(SubCl_MPO_TAZ_F))].append(int(row.getValue(field)))

# build a dictionary containing the split MPO values, keyed by (CSTDM TAZ, MPO TAZ)
MPO_Split_dict = {}                                    # initialize empty dictionary to contain split MPO values
rows = arcpy.SearchCursor(MPO_Split, "", "", Spl_CSTDM_TAZ_F + ';' + Spl_MPO_TAZ_F + ';' +
                        fraction_field_base + ';' + fraction_field_growth)   # create cursor to read all rows in shapefile
for row in rows:                                                             # iterate through the rows in the cursor
        CSTDM_ID = row.getValue(Spl_CSTDM_TAZ_F)                             # get current CSTDM TAZ
        MPO_ID = row.getValue(Spl_MPO_TAZ_F)                                 # get current MPO TAZ
        split_value_base = 0.01 * row.getValue(fraction_field_base)          # get current base split value
        split_value_growth = 0.01 * row.getValue(fraction_field_growth)      # get current growth split value
        base_value = Subclass_Base_Table_dict[MPO_ID]
        future_value = Subclass_Table_dict[MPO_ID]
        MPO_Split_dict[(CSTDM_ID, MPO_ID)] = []
        for n in range(len(base_value)):
                base_proportion = base_value[n] * split_value_base
                growth_proportion = (future_value[n] - base_value[n]) * split_value_growth
```

```
                if growth_proportion >= 0:
                        future_proportion = base_proportion + growth_proportion
                elif growth_proportion < 0:
                        future_proportion = split_value_base * future_value
                # multiply MPO values by the split value for each subclass, then round and convert them to integers
                MPO_Split_dict[(CSTDM_ID, MPO_ID)].append(int(round(future_proportion)))

# re-aggregate the the split MPO values for use in the next section
MPO_SumCheck_dict = {}                  # initialize empty dictionary to contain split MPO values, which have
                                        # been re-aggregated by MPO TAZ
for key in MPO_Split_dict:
        MPO_ID = key[1]                 # get current MPO TAZ
        if MPO_ID in MPO_SumCheck_dict: # if there is already a record for the MPO TAZ then...
                                        # add the MPO value to the existing record
                MPO_SumCheck_dict[MPO_ID] = [(x + y) for x,y in zip(MPO_SumCheck_dict[MPO_ID], MPO_Split_dict[key])]
        else:                           # if not, then ...
                MPO_SumCheck_dict[MPO_ID] = MPO_Split_dict[key]     # create a record for that MPO TAZ

##### The following section of code compares the original values to the values that have been split and     #####
##### then re-aggregated, to make sure that people or households have not been lost or added during the    #####
##### splitting process. If there are differences found, then they are corrected by iteratively adding or  #####
##### subtracting 1 unit from the largest subsection.                                                      #####

run = True                              # initialize run to True
while run == True:                      # this loop will run until no differences are found
        run = False                     # after entering the first loop, set run to False so that the loop will end after
                                        # the first time through if no differences are found
        for key in MPO_SumCheck_dict:   # loop through all of the re-aggregated records
                for n in range(len(MPO_SumCheck_dict[key])):    # loop through each value in the re-aggregated records
                        # if there is a (greater than) difference between the re-aggregated and the original value, then...
                        if MPO_SumCheck_dict[key][n] > Subclass_Table_dict[key][n]:
                                high_value = -9999
                                for k,v in MPO_Split_dict.items():          # find the largest split area...
                                        if k[1] == key and v[n] >= high_value:
                                                high_value = v[n]
                                                high_value_key = k[0]
                                MPO_Split_dict[(high_value_key, key)][n] -= 1     # and subtract 1 unit from it...
                                MPO_SumCheck_dict[key][n] -= 1      # also, subtract 1 unit from the same area in
                                                                    # the re-aggregated dictionary, so that the
                                                                    # next time it is checked, the correction will
                                                                    # be represented
                                run = True          # since a difference was found, set run to True, so
                                                    # that every value will be checked again
                        # if there is a (less than) difference between the re-aggregated and the original value, then
                        # do the same as above, but add a unit...
                        elif MPO_SumCheck_dict[key][n] < Subclass_Table_dict[key][n]:
```

```
                    high_value = -9999
                    for k,v in MPO_Split_dict.items():
                        if k[1] == key and v[n] >= high_value:
                            high_value = v[n]
                            high_value_key = k[0]
                    MPO_Split_dict[(high_value_key ,key)][n] += 1
                    MPO_SumCheck_dict[key][n] += 1
                    run = True

##############################
##############################

# now that all of the erroneous changes have been corrected, aggregate the split MPO values by CSTDM TAZ
MPO_Split_Sum_dict = {}                     # initialize empty dictionary
for key in MPO_Split_dict:
    CSTDM_ID = key[0]                       # get current CSTDM TAZ
    if CSTDM_ID in MPO_Split_Sum_dict:      # if there is already a record for the current CSTDM TAZ then...
        MPO_Split_Sum_dict[CSTDM_ID] = [(x + y) for x,y in zip(MPO_Split_Sum_dict[CSTDM_ID],
            MPO_Split_dict[key])]           # add the split MPO values to the existing values
    else:                                   # if there is no record for the current CSTDM TAZ then...
        MPO_Split_Sum_dict[CSTDM_ID] = MPO_Split_dict[key]     # create a new record for the split MPO values

# build a dictionary containing the summed values of the nesting MPO TAZs, keyed by CSTDM TAZ
MPO_Nest_dict = {}                          # initialize empty dictionary
rows = arcpy.SearchCursor(MPO_Nest, "", "", Nest_CSTDM_TAZ_F + ';' + Nest_MPO_TAZ_F)     # create cursor to read all
                                                                                        # rows in shapefile
for row in rows:                            # iterate through the rows in the cursor
    CSTDM_ID = row.getValue(Nest_CSTDM_TAZ_F)       # get current CSTDM TAZ
    MPO_ID = row.getValue(Nest_MPO_TAZ_F)           # get current MPO TAZ
    if CSTDM_ID in MPO_Nest_dict:           # if there is already a record for the current CSTDM TAZ then...
        MPO_Nest_dict[CSTDM_ID] = [(x + y) for x,y in zip(MPO_Nest_dict[CSTDM_ID],
Subclass_Table_dict[MPO_ID])]               # add the MPO values to the existing values
    else:                                   # if there is no record for the current CSTDM TAZ then...
        MPO_Nest_dict[CSTDM_ID] = Subclass_Table_dict[MPO_ID]     # create a new record for the
                                                                  # MPO values


# copy the CSTDM TAZs to the output workspace
result = arcpy.Copy_management(os.path.abspath(CSTDM_TAZ), os.path.join(output_workspace, outfile))
CSTDM_TAZ_working = result.getOutput(0)

# add the fields from the input field list to the working CSTDM TAZ shapefile
for field in field_list:
    arcpy.AddField_management(CSTDM_TAZ_working, field, field_type)

# build a dictionary containing the summed values of the nested and non-nested TAZs, keyed by CSTDM TAZ
```

```
CSTDM_TAZ_dict = {}                                     # initialize empty dictionary
rows = arcpy.SearchCursor(CSTDM_TAZ_working, "", "", CSTDM_TAZ_F)        # create cursor to read all rows in shapefile
for row in rows:                                        # iterate through the rows in the cursor
        CSTDM_ID = row.getValue(CSTDM_TAZ_F)            # get current CSTDM TAZ
        if CSTDM_ID not in MPO_Nest_dict:               # if there is no record in the nested dictionary, then...
                CSTDM_TAZ_dict[CSTDM_ID] = MPO_Split_Sum_dict[CSTDM_ID]    # just use the list in the
                                                                           # split dictionary
        elif CSTDM_ID not in MPO_Split_Sum_dict:        # if there is no record in the split dictionary, then...
                CSTDM_TAZ_dict[CSTDM_ID] = MPO_Nest_dict[CSTDM_ID]         # just use the list in the
                                                                           # nested dictionary
        else:
                CSTDM_TAZ_dict[CSTDM_ID] = [(x + y) for x,y in zip(MPO_Nest_dict[CSTDM_ID],
                        MPO_Split_Sum_dict[CSTDM_ID])]          # otherwise, sum the values for both the
                                                                # nested and non-nested TAZs

# use the CSTDM TAZ dictionary to update the fields in the CSTDM shapefile
rows = arcpy.UpdateCursor(CSTDM_TAZ_working)            # create cursor to update all rows in shapefile
for row in rows:
        CSTDM_ID = row.getValue(CSTDM_TAZ_F)            # get current CSTDM TAZ
        n = 0
        for field in field_list:
                row.setValue(field, CSTDM_TAZ_dict[CSTDM_ID][n])        # set the value in the current row and field to
                                                                        # the value from the CSTDM TAZ dictionary
                n += 1
        rows.updateRow(row)                             # commit the change

del row                                                 # delete the row and...
del rows                                                # cursor objects to free any locks on the table
```

Appendix E: Employment Industry Category Cross-Classification

1. Amador

MPO Category	Description	CSTDM Category
MED_KSF	Medical	EdMed
ES_ENR	Elementary School	
HS_ENR	High School	
LI_KSF	Light Industry	Prim_Sec, Whole, TranU
HI_KSF	Heavy industry	
CBD_KSF	CBD	Office, Retail, LeisHosp, OthServ
SC_KSF	Shopping center	
GC_KSF	General Commercial	
OFF_KSF	Office	
SKI	Ski	
CASINO	casino	
WINE	wine	
REC	recreation	

2. Tulare

MPO Category	Description	CSTDM Category
RETEMP	Retail	Retail
OFFEMP	Office	Office, EdMed, LeisHosp, OthServ
SEREMP	Service	
EDUEMP	Education	
GOVTEMP	Government	
AGREMP	Agriculture	Prim_Sec, Whole, TranU
INDEMP	Industry	

3. Madera

MPO Category	Description	CSTDM Category
RET_EMP	Retail	Retail, LeisHosp
OFF_EMP	Office	Prim_Sec,Whole, TranU, Office, EdMed, OtherSer
MAN_EMP	manufacture	
OTH_EMP	Other	
GOV_EMP	Government	
EDU_EMP	Education	

4. Shasta

MPO Category	Description	CSTDM Category
EMP_SCHOOL	School	EdMed
EMP_COLLGE	College	
EMP_MEDOFF	Medical Office	
EMP_HOSPTL	Hospital	
EMP_DEVREC	Developed Recreation	LeisHosp
EMP_CASINO	Casino	
EMP_HOTEL	Hotel	
EMP_RESTNT	Restaurant	
EMP_RESTHI	Restaurant	
EMP_OFFICE	Office	Office
RMP_GOVT	Governement	
EMP_GOVTHI	Governement	
EMP_INDSTL	Industrial	Prim_Sec, TranU, OthServ
EMP_SRVCOM	Service Commercial	
EMP_RESCAR	Residential Care	
EMP_CHLDCR	Child Care	
EMP_INST	Institutional (religious, clubs)	
EMP_RETAIL	Retail	Retail
EMP_RETLHI	Retail High	
EMP_RETWAR	Retail Wharehouse (big box)	
EMP_WHOLSL	Wholesale	Whole

5. Stanislaus

MPO Category	Description	CSTDM Category
RET_EMP	Retail	Retail, LeisHosp
SER_EMP	Service	Office, EdMed
EDU_EMP	Education	
GOV_EMP	Government	
OTH_EMP	Other	Prim_Sec, Whole, TranU, OthServ

6. SCAG

MPO Category	Description	CSTDM Category
Ag_emp	Agriculture and Mining	Prim_Sec
Const_emp	Construction	
Manu_emp	Manufacturing	
Whole_emp	Wholesale trade	Whole
Ret_emp	Retail	Retail
Trans_emp	Trans, warehousing, util	TranU
Infor_emp	Information	Office
FIRE_emp	financial	
Prof_emp	professional/business	
Educ_emp	educatino and health	EdMed
ArtEnt_emp	leisure and hospitality	LeisHosp
OthSer_emp	other services	OthServ
PubAdm_emp	public administration	Office

7. San Joaquin County

MPO Category	Description	CSTDM Category
Ret	Retail	Retail
Ser	Service	Prim_Sec,Whole,TranU,Office,EdMed,LeisHosp, OthServ
Oth	Other	

8. SANDAG

MPO Category	Description	CSTDM Category
emp_edhs	education and health services jobs	EdMed
emp_lh	leisure & hospitality jobs	LeisHosp
emp_mil	military jobs	Military
emp_agmin	agricultural jobs	Prim_Sec, Office,OthServ
emp_cons	construction jobs	
emp_mfg	manufacturing jobs	
emp_fre	finance and real estate jobs	
emp_info	information jobs	
emp_pbs	professional and business services jobs	
emp_os	other services jobs	
emp_gov	government jobs	
emp_sedw	self-employment	
emp_retrade	retail jobs	Retail
emp_twu	transportation and warehousing jobs	TranU
emp_whtrade	wholesale jobs	Whole

9. SACOG

MPO Category	Description	CSTDM Category
Retail	Retail	Retail, LeisHosp
Office	Office	Office
Medical	Medical	EdMed
Education	Education	
Manuf_Other	Manufacturing and Other	Prim_Sec, Whole, TranU

10. MTC

MPO Category	Description	CSTDM Category
FPSEMPN	Financial and professional services employment	Prim_Sec, Whole,TranU,Office
OTHEMPN	Other employment	
AGREMPN	Agricultural and natural resources employment	
MWTEMPN	Manufacturing, wholesale trade and transportation employment	
HEREMPN	Health, educational and recreational service employment	EdMed, LeisHosp,OthServ
RETEMPN	Retail trade employment	Retail

11. Merced

MPO Category	Description	CSTDM Category
AGR	Agriculture	Prim_Sec, TranU, Office, EdMed, OthServ
IND	Industrial	
OFF	Office	
SCH	School	
RET	Retail	Retail, Whole, LeisHosp

12. Kings

MPO Category	Description	CSTDM Category
Industrial	Industrial	Prim_Sec, Whole, TranU
Agriculture	Agriculture	
Retail	Retail	Retail, LeisHosp
Office	Office	Office, EdMed, OthServ
Government	Government	
Education	Education	
Other	Other	

13. Fresno

MPO Category	Description	CSTDM Category
Retail	Retail	Retail
Services	Services	Office, EdMed, OthServ
Government	Government	
Education	Education	
Other	Other	Prim_Sec, Whole, TranU, LeisHosp

14. BCAG

MPO Category	Description	CSTDM Category
RET_KSF	Retail	Retail, LeisHosp
RRET_KSF	Regional retail	Prim_Sec, Whole, TranU, Office, EdMed, OthSer
IND_KSF	Industry	
OFF_KSF	Office	
MED_KSF	Medical	
HOSP_KSF	Hospital	
PQP_KSF	Public-Quasi Private	

15. Kern

MPO Category	Description	CSTDM Category
BASIC_EMP	Basic	Prim_Sec, Whole
RHRET_EMP	Retail High	Retail, LeisHosp
RMRET_EMP	Retail	
SCSER_EMP	Service	EdMed, OthServ
SOSER_EMP	Other Service	Office
BWOTH_EMP	Warehousing/Trans	TranU

16. AMBAG

MPO Category	Description	CSTDM Category
FARM	Farm	Prim_Sec, Whole, TranU
INDU	Industrial	
CONS	Construction	
RETL	Retail	Retail
SERV	Service	Office, EduMed,LeisHosp, OthServ
GOVT	Government	

17. SBCAG

MPO Category	Description	CSTDM Category
Finance, Ins, RE	finance	Office
Govt*	Government	
Services	Services	EdMed, LeisHosp, OthServ
Ag	Agriculture	Prim_Sec
Mining	Mining	
Const	Construction	
Manuf	Manufacturing	
Retail	Retail	Retail
Tran	Transportation	TranU
Whole	Whole	Whole

18. SLOCOG

MPO Category	Description	CSTDM Category
RETAIL_JOB	Retail	Retail, LeisHosp
NONRET_JOB	Non-retail	Prim_Sec, Whole, TranU, Office, EdMed, OtherSer

19. Wine Country

MPO Category	Description	CSTDM Category
HEALTH_EMP	Health care and Social assistance	EdMed
EDU_EMP	Education	
LEI_EMP	Leisure and Recreation	LeisHosp
IND_EMP	Industrial	Prim_Sec, Whole, TranU
AGR_EMP	Agriculture	
RET_EMP	Retail	Retail
OTH_EMP	Other services	Office, OthSer

20. Tuolumne

MPO Category	Description	CSTDM Category
IND	Industrial	Prim_Sec, Whole, TranU
CHURCH	Church	Office, Retail, LeisHosp, OthServ
MICOM	Minor Commercial	
MACOM	Major Commercial	
MOTEL	Motel	
SS	Service Station	
MUCOM	Multiple Use/Other Commercial	
CASINO	Casino	

Appendix F: California Department of Finance Forecasts

Interim Projections

The interim data projections updated DOF previous projections[1] (developed in June 2007**)** using 2010 Census data. The interim projections latest time horizon year was 2020, while the previous forecasts had a 2050 horizon year. The June 2011 DOF interim projections are reproduced in the table below:

June 2011

INTERIM COUNTY PROJECTIONS
Projections for 2010, 2015, and 2020
Benchmarked on Census 2010

COUNTY	July 2010	July 2015	July 2020
ALAMEDA	1,514,614	1,553,847	1,604,285
ALPINE	1,164	1,192	1,224
AMADOR	38,207	41,316	44,647
BUTTE	220,965	242,868	267,544
CALAVERAS	45,741	49,442	53,384
COLUSA	21,450	23,859	26,583
CONTRA COSTA	1,053,198	1,117,679	1,196,467
DEL NORTE	28,648	30,753	33,083
EL DORADO	181,922	195,179	210,202
FRESNO	935,253	1,027,838	1,132,389
GLENN	28,155	31,128	34,431
HUMBOLDT	135,030	137,348	140,459
IMPERIAL	175,449	196,866	219,950
INYO	18,589	18,978	19,607
KERN	845,087	937,644	1,041,725
KINGS	153,702	171,160	190,830
LAKE	64,911	69,456	74,059
LASSEN	34,911	36,647	38,664
LOS ANGELES	9,826,646	10,025,486	10,358,135
MADERA	151,690	174,200	198,027
MARIN	252,853	252,203	255,889
MARIPOSA	18,304	19,427	20,596
MENDOCINO	88,001	91,327	95,273
MERCED	257,176	289,153	325,075
MODOC	9,669	10,631	11,703

[1] State of California, Department of Finance, *Race/Ethnic Population with Age and Sex Detail, 2000–2050.* Sacramento, CA, July 2007.

MONO	14,273	15,622	17,219
MONTEREY	416,056	431,786	452,240
NAPA	137,054	145,852	157,417
NEVADA	99,057	103,697	109,131
ORANGE	3,016,973	3,121,120	3,254,203
PLACER	350,701	385,169	426,161
PLUMAS	19,977	20,299	20,749
RIVERSIDE	2,207,875	2,500,421	2,829,908
SACRAMENTO	1,424,367	1,491,005	1,571,320
SAN BENITO	55,116	63,039	72,190
SAN BERNARDINO	2,044,160	2,212,736	2,401,846
SAN DIEGO	3,106,307	3,241,325	3,404,737
SAN FRANCISCO	807,418	814,160	822,064
SAN JOAQUIN	688,655	782,486	891,753
SAN LUIS OBISPO	270,515	279,059	290,312
SAN MATEO	719,991	724,322	734,240
SANTA BARBARA	425,391	432,660	443,934
SANTA CLARA	1,787,090	1,840,995	1,913,716
SANTA CRUZ	263,057	269,655	278,457
SHASTA	177,595	191,085	206,148
SIERRA	3,218	3,127	3,062
SISKIYOU	44,983	46,523	48,387
SOLANO	414,321	438,936	468,060
SONOMA	485,530	503,138	528,403
STANISLAUS	515,978	573,521	640,876
SUTTER	95,179	111,954	130,803
TRINITY	13,782	15,058	16,471
TULARE	444,845	501,827	566,045
TUOLUMNE	55,426	57,586	59,875
VENTURA	826,313	864,802	912,370
YOLO	201,972	218,658	237,240
YUBA	72,313	84,589	98,236
CALIFORNIA	37,370,595	39,301,563	41,678,215

For additional information, see www.dof.ca.gov/research/demographic/reports/projections/interim/

The interim 2011 projections (new) and the earlier 2007 projections (old) were used together to provide adjusted projections for 2035 and 2050 as described in the following formula:

$$Pop_{2035} = (Pop_old_{2035} - Pop_old_{2020}) + Pop_new_{2020}$$

$$Pop_{2050} = (Pop_old_{2050} - Pop_old_{2030}) + Pop_{2035}$$

Projection Distributions

The earlier DOF projections to 2050 included distributions by county by age and by ethnicity; however, the interim projections did not include these distributions (only total population by county). A sample of the downloaded data is shown:

County	Year	Sex	Age	White	Hispanic	Asian	Pacific Islande	Black	American Indi	Multirace
Alameda	2020	F	0	2135	3971	2507	112	998	43	337
Alameda	2020	F	1	2150	3919	2514	114	1018	45	318
Alameda	2020	F	2	2157	3858	2525	114	1040	46	301
Alameda	2020	F	3	2161	3800	2536	114	1063	47	285
Alameda	2020	F	4	2164	3738	2551	115	1082	48	270
Alameda	2020	F	5	2170	3687	2576	116	1094	49	256
Alameda	2020	F	6	2179	3664	2612	115	1101	51	243
Alameda	2020	F	7	2181	3614	2648	115	1101	52	231
Alameda	2020	F	8	2176	3564	2683	115	1091	53	220
Alameda	2020	F	9	2166	3510	2716	115	1072	54	209
Alameda	2020	F	10	2155	3492	2755	114	1058	55	192
Alameda	2020	F	11	2168	3500	2798	114	1052	56	177
Alameda	2020	F	12	2185	3509	2840	114	1048	58	163

Target files used in the population synthesis were obtained first from MPO/RTPA and the reformatted and split into the model zonal system (see Section 2: MPO and RTPA Forecasts for 2020 and 2035) although not all MPOs had this data available, in which case the DOF projections were used instead (as described in Section 3). Most of the control totals come directly from the MPO. To get the age breakdown the rates per *county, sex, age* combination from DOF are used and then applied to the existing control totals by county, which illustrates the age distribution based on the total numbers that we have for the county.

	County	Year	Sex	Age	total	ctyTotal	ratio
1	Alameda	2010	F	0	9918	1530697	0.006479
2	Alameda	2010	F	1	9954	1530697	0.006503
3	Alameda	2010	F	2	9966	1530697	0.006511
4	Alameda	2010	F	3	9984	1530697	0.006523
5	Alameda	2010	F	4	9986	1530697	0.006524
6	Alameda	2010	F	5	9889	1530697	0.00646
7	Alameda	2010	F	6	10108	1530697	0.006604
8	Alameda	2010	F	7	10289	1530697	0.006722
9	Alameda	2010	F	8	10251	1530697	0.006697
10	Alameda	2010	F	9	10422	1530697	0.006809
11	Alameda	2010	F	10	9426	1530697	0.006158

For example, to get the number of 11 year old females in Alameda County for 2010 one would take the ratio (0.006158) and multiply it by the population control total for Alameda, which is provided by MTC, the regional MPO. This is done for all years to get the age distribution for the growing population.

Appendix G: Moody's and TREDIS Forecasts

TREDIS is a transportation economic impact tool that pivots economic growth (from transportation investment) off of a baseline economic forecast, and also links commodity trade to economic drivers. It has a series of modules that calculate different forms of impacts and benefits. One module is an economic adjustment (EA) module which calculates wider economic development benefits, including impacts of business productivity, economic development and multiplier effects from input-output analysis.

Employment data was based on the EA module and computed in the TREDIS methodology using input-output analysis ratios and multipliers from the Minnesota IMPLAN Group. The TREDIS forecasting system incorporates elements from Moodys'. Moody's publishes industry forecast at the national, state and county level, with a 30 year horizon (currently forecasting to 2041). The forecasts are at a highly detailed industry level. Moody's 196 industry sectors are NAICS bases and are therefore easily related to the data elements of TREDIS methodology. The TREDIS forecasts employment by industry, occupational detail was not available. The Bureau of Labor Statistics (BLS) was used to build a crosswalk between NAICS and SOC codes.

Moody's provides state level demographic changes in total population, population by age, number of households and average household income for the time period between 1981 and 2041. The changes in total population, household, population by age and income are also available at county level and can be aggregated into LUZ, TAZ, or state level for the same time period.

TREDIS provided us with county-level annual data for the years 2007 to 2041. The number of employees for each combination of industry and occupation were reported by county for each forecast year. The industry and occupation categories that TREDIS used were:

Employment by Industry:

1: AMUSEMENT SERVICES AND MEDIA PRESENTATION

2: Cattle ranching and farming

3: CONSTRUCTION New nonresidential space production

4: CONSTRUCTION New residential space production

5: CONSTRUCTION Nonresidential Space Renovation production

6: CONSTRUCTION Non-Space production

7: CONSTRUCTION Residential Space Renovation production

8: Dairy cattle and milk production

9: ELECTRIC UTILITIES production

10: FEDERAL GOVERNMENT accounts

11: Field and seed crops farming

12: FINANCE INSURANCE LEGAL

13: Fishing

14: Forestry and allied activities

15: Fruit and nut farming

16: GOVERNMENT ENTERPRISES

17: Greenhouse, nursery, and floriculture production

18: HEALTH SERVICES

19: HOTELS

20: HOUSEHOLDS

21: INDUSTRIAL BASED SERVICES

22: MANUFACTURING Chemicals, Plastic, Rubber, Glass, Cement (NAICS:325-327) production

23: MANUFACTURING Computers, Electronics and Electrical (NAICS: 334 and 335) production

24: MANUFACTURING Food production

25: MANUFACTURING Machinery and Transportation Equipment (NAICS: 333 and 336) production

26: MANUFACTURING Metal, Steel (NAICS: 331-332) production

27: MANUFACTURING Petro-Chemicals (NAICS:324) production

28: MANUFACTURING Textiles (NAICS:313-316) production

29: MANUFACTURING Wood Products, Printing, Furniture, Misc (NAICS:321,322,323,337,339) production

30: MEDIA PRODUCTION

31: MINING AND EXTRACTION production

32: Noncattle animal production

33: NON-ELECTRIC UTILITIES AND COMMUNICATIONS production

34: OFFICE BASED SERVICES

35: ONSITE BUSINESS SERVICES

36: ONSITE PERSONAL SERVICES

37: PRIMARY K-12 EDUCATION

38: PROFESSIONAL SERVICES

39: REAL ESTATE

40: RELIGIOUS SERVICES

41: RESTAURANTS (SIC:58)

42: RETAIL BASED SERVICES

43: RETAIL TRADE automotive (SIC:55)

44: RETAIL TRADE general merchandise (SIC:52-53,56,57,59)

45: RETAIL TRADE unprepared food (SIC:54)

46: SCRAP AND ADJUSTMENTS

47: SECONDARY EDUCATION

48: STATE AND LOCAL GOVT accounts

49: Support activities for agriculture and forestry

50: TRANSPORTATION SERVICES

51: Vegetable and melon farming

52: WHOLESALE TRADE

<u>Employment by Occupation:</u>

1: managerial labor (SOC: 11-)

2: business and financial operation labor (SOC: 23-)

3: professional labor (SOC: 15-23)

4: post-secondary education labor (SOC: 25-1000)

5: gradeschool education labor (SOC: 25-2000, 25-3000, 25-4000, 25-9000)

6: entertainer and media labor (SOC: 27-)

7: health labor (SOC: 29-31)

8: non-food service labor (SOC: 33-, 37-, 39-)

9: food service labor (SOC: 35-)

10: retail sales labor (SOC: 41-1000, 41-2000)

11: non-retail sales labor (SOC: 41-3000, 41-4000, 41-9000)

12: clerical labor (SOC: 43-)

13: agricultural labor (SOC: 45-)

14: construction labor (SOC: 47-)

15: maintenance and repair labor (SOC: 49-)

16: assembly and fabrication labor (SOC: 51-)

17: transport labor (SOC: 53-)

18: military labor (SOC: 55-, 99-)

Appendix H: Employment Floorspace by Industry

The inventory of floorspace by type by zone was generated from the Population Synthesizer (for residential space), the Employment Synthesizer (for non-residential space) and inventories of resource space (including recreational) and protected space. This total was allocated to individual parcels in the zone by the Floorspace Synthesizer, guided primarily by the "existing land use" layer. The "existing land use" was derived primarily from general plans supplemented by several other datasets, including the National Land Cover Dataset 2001, California Protected Areas Database and the Farmland Mapping and Monitoring Program.

The Floorspace Synthesizer (FS) for 2000 integrates the results from the Population Synthesizer (PS) and Employment Synthesizer (ES), as well as other datasets to form the base parcels for the Space Development (SD) module of the California PECAS model. It takes the Traffic Analysis Zone (TAZ) results from the PS and ES and distributes the amount of space for each space type to individual "parcels" in California on a TAZ by TAZ basis. The parcels are generated by dividing all the land into appropriately sized (2500 square meter) elemental units suitable for microsimulation. All remaining non-residential, non-employment-based space types, such forest space and park space, were added to the parcels using Geographic Information Systems (GIS) overlay techniques.

The data evolved through multiple steps to the finished product, which is a table in the SQL Server on the Chaos server in the ProSynthesizer database called "synthParcels". This table represents each 2500 square meter portion of California, what PECAS space type exists on it and what amount of floorspace exists on that parcel.

The purpose of the FS is to provide the base year parcel database for the California PECAS model. This the detailed spatial accounting system for the California PECAS model. As the model runs and more economic activity is created, a demand for more floorspace is created. It is the function of the SD module to provide for more floorspace

by developing parcels. This can be done by developing on previously vacant parcels or converting parcels with existing space. Development actions that can occur include no change, demolition, dereliction, renovation, addition and transition. Each parcel has a probability of one of these events for each year that SD runs, which is determined by various attributes of the parcel in conjunction with the price pressure coming from the PECAS Activity Allocation (AA) Module. In this way, SD creates new floorspace for AA in out years from the base year.

The FS zonal targets are derived from the PS and ES, so these targets inherit all the errors and assumptions of those data products. The base parcels' spatial data framework is based on a 50-meter GIS overlay on the California state boundary. The precision of this spatial data is quite high. The Synthetic Existing Land Use data that are used to inform the FS are based on county and city general plans, as well as a mixture of several other datasets. The accuracy and precision of these data vary depending on the source.

Though the end product of the FS is eventually a table in SQL Server, this table was created from GIS datasets. Principally, the FS is formed from the output of the Arc/Info GRID program, which rasterized a polygon dataset that represented the boundary for California. Each of the approximately 165 million parcels were then attributed with a unique id number so that the tabular data generated later in the process could be relinked to the spatial location of the cell (see sample data for an example). Other attributes in the FS end table come from other GIS processes that use the same Arc/Info GRID reference system and then the attributes are joined together using the unique cell id number.

The FS's final table has the following columns:
pecas_parcel_num (big int) - unique cell id
taz (int) - traffic analysis zone that cell resides in
space_type_id (int) - PECAS spacetype code assigned to cell
space_type_name (varchar 100) - string description of PECAS spacetype code

space_quantity (float) - amount of floorspace in square feet on cell

land_area (float) - amount of real land area on cell; in the case of the California model with fixed 50 meter cells, this is 26909.776 square feet

zoning_rules_code (int) - zoning code for future land use development (general plan derived)

year_effective (int) - the year that the information in the other columns is associated with

The classification of space type used by the Floorspace Synthesizer targets and final space designations at the grid cell level is shown in Table 1.

Table 1 - Space type classification used in floorspace synthesizer

Space Type ID	Space Type Name
101	Park space
102	Forest space
103	Mining and extraction space
104	Grazing space
105	Non-irrigated agricultural space
106	Irrigated agricultural space
107	Light industrial space
108	Heavy industrial space
109	Warehouse space
110	Highway retail space
111	Downtown retail space
112	Mall and big-box retail space
113	Neighborhood retail space
114	Low-density office space
115	High-density office space

116	Developed amusement parks space
117	Hospital space
118	Secondary education space
119	Primary K-12 education space
120	Religious space
121	Government operations space
122	Military space
123	Fishing dock space
124	Depot space
125	Rural luxury residential
126	Rural economy residential
127	Acreage luxury residential
128	Acreage economy residential
129	SFD luxury residential
130	SFD economy residential
131	Joined luxury residential
132	Joined economy residential
133	Low-rise luxury residential
134	Low-rise economy residential
135	High-rise luxury residential
136	High-rise economy residential
137	Urban mobile home residential
138	General quarters (GQ) residential

The Floorspace Synthesizer was built up from several different components. The zonal control totals come from the PS and ES, while the base parcels come from GIS

datasets. See the documentation on the PS and ES for the location of these datasets. The FS GIS processing files and input files are stored on the CHAOS server of the University of California, Davis, in folders E:\Production\gis directory and the SQL Server database named "ProSyntehsizer", respectively. The results of the synthesizer were then used to build the base parcels for the Space Development (SD) module in the ProSD database on the Chaos server and exported to FloorspaceI.csv for the 2000 floorspace input for the Activity Allocation (AA) module.

The FS control totals come from three sources. For residential and employment space types, control totals come from the Population and Employment Synthesizers, respectively. For non-residential, non-employment space types, the amount of space in each TAZ is obtained from GIS data assuming a floor area ratio (FAR) of 1 on each parcel (no wasted land in the parcel). The TAZ were split into three groups (low, medium and high FAR) in order to better simulate the floorspace in those zones; based on these groupings, the following three floorspace input tables were created in the database:

Floorspace_HighFAR_6thrun
Floorspace_MedFAR_6thrun
Floorspace_LowFAR_6thrun

Floorspace is assigned to grid cells based on a weighting scheme; as with floorspace total, separate schemes were created for low, medium, and high density FAR zones. For example, high FAR zones had different weights for space types that may be better represented in urban areas, such as downtown retail and downtown office. The match coefficient tables used in these schemes are as follows:

MatchCoeff_HighFAR_6thrun
MatchCoeff_MedFAR_6thrun
MatchCoeff_LowFAR_6thrun

The set of parcels used in the FS are derived from the 2000 Synthetic Existing Land Use dataset, which is comprised of 165 million 2500 square meter elements and uses

the existing land use attribute as the synthesizer's primary weighting attribute. Other attributes in the weighting scheme, such as distance to freeways and highways, comes from the local effects datasets, which is developed in GIS grid format, transformed into tabular form and loaded into the SQL Server as a table with attributes matching those in the match coefficient weighting table. The process of transformation from GIS dataset to tabular attributes in SQL Server began with rasterizing the attribute into a code if it was not already in this format. This statewide Arc/Info GRID was then clipped into 58 county datasets (the whole state could not processed as one because of software limitations). These county GRIDS were then converted to point shapefiles and their tabular attribute tables imported into SQL Server and joined back into a statewide table. The attributes were all joined together in the final table by using the unique id code associated with each parcel.

Next, the FS is run on a TAZ-by-TAZ basis in the three separate FAR groups with the Floorspace table as the control total and the Match Coefficient table as the weighting table. The synthesizer weights all parcel grid cells in each TAZ into ordered lists with equally-weighted cells randomly ordered relative to one another. Floorspace is then assigned to parcels with the highest weights until the parcel approached its target FAR; at that point, a negative weight is assigned and the parcel moves down the list in order for other parcels to be assigned floorspace. This process continues until all of the floorspace in the control table is assigned to individual parcel grid cells.
The completed FS provides the floorspace per parcel for residential- and employment-based floorspace. The non-residential, non-employment floorspace types were inserted using GIS processing. Various datasets were queried in order to find which parcels were occupied by park space, forest space, grazing space and irrigated and non-irrigated agricultural space.

The following list shows the sources for the various spacetypes:
101 (Park space) - California Protected Areas Database - http://calands.org
102 (Forest space) - National Land Cover Dataset 2001 - http://landcover.usgs.gov/faq.php#eight

103 (Mining and Extraction space) - Not yet completed

104 (Grazing space) - National Land Cover Dataset 2001 - http://landcover.usgs.gov/faq.php#eight

105 (Non-irrigated agricultural land) - Farmland Mapping and Monitoring Program - http://www.conservation.ca.gov/dlrp/fmmp/Pages/Index.aspx

106 (Irrigated agricultural land) - http://www.conservation.ca.gov/dlrp/fmmp/Pages/Index.aspx

The attributes in the finalized version of the FS are then transposed into the ProSD SQL Server database to be used as the base parcels for the SD module of the PECAS model. The Floorspace Synthesizer results are also summarized by TAZ and space type into the table FloorspaceI to be used as input floorspace for the AA module of the PECAS model.

The possible values for the space type and zoning rule attributes are shown in Tables 2 and 3.

Table 2 - Space_type_id domain.

Space_type_id	Space_type_name	PECAS Space Type Description
101	Park space	National, state and county/city parks, as well as several national forests
102	Forest space	Public and private forests, derived from the National Land Cover Database
103	Mining and extraction space	Derived from the mine spots shapefile (See the Metadata for mine spots shapefile for more details)
104	Grazing space	Range land with sparse trees, derived from Farmland 2000 shapefile
105	Non-irrigated	Derived from Farmland 2000 shapefile

	agriculture space	
106	Irrigated agriculture space	Derived from Farmland 2000 shapefile
107	Light industrial space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in several activities
108	Heavy industrial space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in several activities
109	Warehouse space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in wholesale trade industry
110	Highway retail space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in several activities
111	Downtown retail space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in several activities
112	Mall and big-box retail space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in several activities
113	Neighborhood retail space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in several activities
114	Low-density office space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in several activities
115	High-density office	Derived from design diagram of the PECAS

	space	production model as the sum of the space used by workers in several activities
116	Developed amusement parks space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in several activities
117	Hospital space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in hospitals
118	Secondary education space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in secondary education
119	Primary K-12 education space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in primary education
120	Religious space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in religious service
121	Government operations space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in several activities
122	Military space	Derived from design diagram of the PECAS production model as the military boundary shapefile
123	Fishing dock space	Derived from design diagraph of the PECAS production model
124	Depot space	Derived from design diagram of the PECAS production model as the sum of the space used by workers in wholesale trade

125	Rural luxury residential	Derived from design diagram of the PECAS production model as the sum of the rural luxury residential space used by some households
126	Rural economy residential	Derived from design diagram of the PECAS production model as the sum of the rural economy residential space used by some households
127	Acreage luxury residential	Derived from design diagram of the PECAS production model as the sum of the acreage luxury residential space used by some households
128	Acreage economy residential	Derived from design diagram of the PECAS production model as the sum of the acreage economy residential space used by some households
129	SFD luxury residential	Derived from design diagram of the PECAS production model as the sum of the SFD luxury residential space used by some households
130	SFD economy residential	Derived from design diagram of the PECAS production model as the sum of the SFD economy residential space used by some households
131	Joined luxury residential	Derived from design diagram of the PECAS production model as the sum of the joined luxury residential space used by some households
132	Joined economy	Derived from design diagram of the PECAS

	residential	production model as the sum of the joined economy residential space used by some households
133	Low-rise luxury residential	Derived from design diagram of the PECAS production model as the sum of the low-rise luxury residential space used by some households
134	Low-rise economy residential	Derived from design diagram of the PECAS production model as the sum of the low-rise economy residential space used by some households
135	High-rise luxury residential	Derived from design diagram of the PECAS production model as the sum of the high-rise luxury residential space used by some households
136	High-rise economy residential	Derived from design diagram of the PECAS production model as the sum of the high-rise economy residential space used by some households
137	Urban mobile home residential	Derived from design diagram of the PECAS production model as the sum of the urban mobile home residential space used by some households
138	General quarters residential	Derived from design diagram of the PECAS production model as the sum of the general quarters residential space used by some households

Table 3 - Existing land use domain

Zoning_rules_code	Description
1	Government parks
2	Multiple use public
3	Agricultural exclusive
4	Agricultural with 2 acres plus residential
5	Agricultural with 10 acres plus residential
6	Light industrial
7	Heavy industrial
8	Residential single-family dwellings maximum 10 units per acre
9	Residential single-family dwellings maximum 1 unit per acre
10	Residential single-family dwellings maximum 0.1 units per acre
11	Residential multi-family maximum 20 units per acre
12	Residential multi-family maximum 100 units per acre
13	Residential multi-family maximum 1000 units per acre
14	Urban mobile home maximum 20 units per acre
15	Grade School
16	College campus
17	Public quasi-public institutional
18	Business park maximum floor-area ratio of 1
19	Highway commercial
20	Downtown commercial
21	Downtown office
22	Mall and big-box retail
23	Neighborhood retail
24	Indian reservation

25	Urban reserve
26	Planned development
27	Fishing docks
28	Military
29	General commercial
30	Mixed suburban
31	Mixed urban
32	Mixed city center
33	Mixed urban live/work
98	Water
99	Other

The FS forms the back bone of the detailed spatial part of the ULTRANS spatial economic model, which is based on the PECAS framework. The outputs of the FS make up the base parcels for the base year of the SD Module of the PECAS model and also the base year floorspace total (FloorspaceI) for the AA Module of the PECAS model.

The FS results become the parcel table in the SD database where the spacetype and space quantity attributes are stored. The following query is run in SQL Server to generate the FloorspaceI table that is used as input into the AA module:
SELECT taz,space_type_id,SUM(space_quantity) FROM synthParcels GROUP BY taz,space_type_id ORDER BY taz,space_type_id

This table is then exported as a comma separated value (CSV) file and placed in the 2000 directory as input into AA in the base year.

Appendix I: Projections Household by Size, Income, Auto Ownership, and Worker Categories

In the Bay Area/California High-Speed Rail Ridership and Revenue Forecasting Study developed by Cambridge Systematics, Inc. they developed socio-demographic data "by combining urban area sociodemographic data by traffic analysis zone from the urban models with the Caltrans statewide model socioeconomic data and the U.S. Census Bureau data" (Final Report, section 2.3).

The sociodemographic data classifies the population by four categories. These are household size (1,2,3, or 4+ persons), household income group (low, medium or high), number of workers in the household (0,1, or 2+ workers), and household car ownership and worker category (0 car, 0<cars<workers, or cars>=workers). The income groups were split at: low = <$35,000, medium = $35,000 to $75,000, and high = >$75,000. These income categories are reported in 2005 dollar equivalents so they were converted to 2000 dollar equivalents because that is the standard CSTDM uses. The combination of these categories resulted in a table with 99 unique combinations.

We were able to get a copy of this data and used it as targets in the population synthesizer. The data was available for 2000 and 2030 since these were the 2 years that their model used. We needed targets for 2020, 2035, and 2050 so a methodology was created for each of these 3 future years.

For 2020, gross estimates were developed as interpolation of 2000 and 2030 values. If there was 0 population in a 2000 category we assumed linear growth and made the 2020 estimate equal 2/3 times the 2030 value. If there was 1 or more population in a 2000 category we assumed exponential growth. The annualized growth rate (G) was calculated using the following formula:

G = (Pop2030/Pop2000)^(1/30)

This growth rate was then applied for 20 years to the 2000 population using the following formula:

Pop2020 = Pop2000 * G^(20)

For 2035, gross estimates were developed as extrapolation from 2000 and 2030 values. If there was 0 population in a 2000 category we assumed linear growth and made the 2035 estimate equal 7/6 (35/30) times the 2030 value. If there was 1 or more population in a 2000 category we assumed exponential growth. The growth rate (G) was applied for 5 years to the 2030 population using the following formula:

Pop2035 = Pop2030 * G^(5)

For 2050, gross estimates were developed as extrapolation from 2000 and 2030 values. If there was 0 population in a 2000 category we assumed linear growth and made the 2050 estimate equal 5/3 (50/30) times the 2030 value. If there was 1 or more population in a 2000 category we assumed exponential growth. The growth rate (G) was applied for 20 years to the 2030 population using the following formula:

Pop2050 = Pop2030 * G^(20).