

1  
2 **Simulator of Activities, Greenhouse Emissions, Networks, and**  
3 **Travel (SimAGENT) in Southern California**  
4  
5

6  
7 Konstadinos G. Goulias<sup>1</sup> (corresponding author)

8 Chandra R. Bhat<sup>2</sup>

9 Ram M. Pendyala<sup>3</sup>

10 Yali Chen<sup>1</sup>

11 Rajesh Paleti<sup>2</sup>

12 Karthik C. Konduri<sup>3</sup>

13 Ting Lei<sup>1</sup>

14 Daimin Tang<sup>1</sup>

15 Seo Youn Yoon<sup>1</sup>

16 Guoxiong Huang<sup>4</sup>

17 And

18 Hsi-Hwa Hu<sup>4</sup>  
19  
20  
21  
22  
23  
24  
25  
26

27 <sup>1</sup>GeoTrans & Department of Geography, University of California, Santa Barbara, California, 93106-  
28 4060. Email: goulias@geog.ucsb.edu (corresponding author).

29 <sup>2</sup>Department of Civil, Architectural and Environmental Engineering, The University of Texas, Austin,  
30 Texas 78712.

31 <sup>3</sup>School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe,  
32 Arizona, 85287-5306.

33 <sup>4</sup>Southern California Association of Governments, Los Angeles, California, 90017.  
34  
35  
36  
37  
38  
39

40 **Manuscript submitted for Presentation at the 2012 Annual Transportation Research Board**  
41 **Meeting and Publication in the Transportation Research Record**  
42 **July 28, 2011.**

# 1 **Simulator of Activities, Greenhouse Emissions, Networks, and** 2 **Travel (SimAGENT) in Southern California** 3 4 5

6 **Abstract**—In this paper we describe the recently developed large scale spatio-temporal simulator of  
7 activities and travel for Southern California. The simulator includes population synthesis that recreates  
8 the entire resident population of this region, provides locations for residences, workplaces, and schools  
9 for each person, estimates car ownership and type as well as main driver for each vehicle, and provides  
10 other key personal and household characteristics. Then, a synthetic schedule generator recreates for  
11 each resident person in the simulated region a schedule of activities and travel that reflects intra-  
12 household activity coordination for a day. These synthetic activity and travel daily schedules are then  
13 converted to multiple Origin Destination (OD) matrices at different times in a day. These are in turn  
14 combined with other OD matrices (representing truck travel, travel from and to ports and airports, and  
15 travel generated outside the region) and assigned to the network in multiple periods in a day. The  
16 assignment output is then used in the software EMFAC to produce estimates of fuel consumed and  
17 pollutants emitted (including CO<sub>2</sub>) by different classes of vehicles. The overall model system also  
18 includes provision for finer spatial and temporal resolutions that is pilot tested using TRANSIMS.  
19 Numerical examples from each major modeling group are also provided together with an outline of next  
20 steps.

21  
22  
23 **Number of Words = 5372**  
24 **Figures = 6**  
25 **Tables = 4**  
26 **Total Word Equivalentents = 7872**  
27  
28  
29  
30

31  
32 **Manuscript submitted for Presentation at the 2012 Annual Transportation Research Board**  
33 **Meeting and Publication in the Transportation Research Record**  
34 **July 28, 2011.**  
35

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

## 1. Introduction

The State of California has recently embarked on an aggressive movement towards reducing greenhouse gas emissions that contribute to global climate change, promoting sustainability, and better managing vehicular travel demand. The recent California State Senate Bill 375 explicitly calls for major metropolitan areas in California to meet ambitious greenhouse gas (GHG) emission reduction targets within the next several years. Metro areas are considering a range of policies to meet the emission reduction targets including land use strategies, pricing mechanisms, managed lanes, telecommuting and flexible work hours, enhancement of transit and pedestrian/bicycle modes, and use of technology to better utilize existing capacity. The analysis of these policies, and responding to the mandates of legislative actions such as Senate Bill 375 in California, calls for the adoption of model systems that are able to accurately represent activity-travel patterns in a fine-resolution time-space continuum. Moreover, these model systems are expected to provide a platform for simulating integrated land use and transportation plans that are better able to represent gains in emission control in the medium (5-10 years) and the longer term (10-25 years) horizons.

The Southern California Association of Governments (SCAG), the metropolitan planning agency for the Southern California region (includes the counties of Imperial, Los Angeles, Orange, Riverside, San Bernardino, and Ventura), is moving forward with the development of a comprehensive activity-based microsimulation model system of travel demand to enhance its ability to estimate the impacts of a range of policy measures in response to Senate Bill 375 (<http://www.scag.ca.gov/sb375/index.htm>). SCAG is also required to develop a “Sustainable Community Strategy” through integration of land use and transportation planning and demonstrate its ability to meet the GHG emissions reduction targets by 2020 (8% GHG per capita per day reduction) and 2035 (13% GHG per capita per day tentatively). These are challenging targets for such a vast region, which includes a population of approximately 18.6 million people in 2008 (expected to grow to 23 million by 2035) and offers an extremely complex multimodal and diverse planning context with multiple actors in different jurisdictions. The new activity-based microsimulation model system is developed to address exactly this diversity among persons and contexts, it is expected to be used in the 2016 Regional Transportation Plan (RTP), and is described in this paper.

This model system is the outcome of the second phase of research and development as well as

1 application of the Simulator of Activities, Greenhouse Emissions, Networks, and Travel (SimAGENT),  
2 which is tailored to the Southern California region and is compared to the four step model system used  
3 in the SCAG 2008 Regional Transportation Plan.  
4

## 5 **2. SimAGENT**

6 The overall model structure is presented in Figure 1 in a schematic cascading form. The set of blocks on  
7 the left hand side represents groups of models that are designed for the first year (baseline) of the  
8 simulation that for this application is 2003 to align with the four-step model of SCAG developed for the  
9 2008 Regional Transportation Plan. Each block of the figure represents a group of techniques and  
10 statistical models many of which are developed to address policy actions aiming at replicating the  
11 resident population activity and travel decision making. In essence this first set of models on the left  
12 side of Figure 1 recreates the resident population and attributes to each person a daily schedule and  
13 ultimately assigns traffic to the network and computes emissions. The middle four blocks evolve the  
14 region's economic and demographic landscape over time. This computerized evolution is done using a  
15 land use model based on the spatial input-output growth forecast model called PECAS (Production,  
16 Exchange and Consumption Allocation System - <http://www.scag.ca.gov/modeling/mtf/index.htm>).  
17 This is paralleled by a set of algorithms and negotiated-with-local-jurisdictions forecasts of residential  
18 development, industrial location, and demographic land use-regional economy evolution at aggregate  
19 levels (cities and subareas within cities). In the middle of the growth forecast and land-use regional  
20 economy components is the household evolution module, which microscopically considers every  
21 resident household (and its members) and gives it transitions and changes in time and space. It also  
22 converts travel times and the spatial distribution of economic activity and residential locations into  
23 accessibility indicators that are used to also drive travel behavior. The right side set of blocks is a  
24 repetition of the daily activity and travel patterns models but at the next and all subsequent years of the  
25 simulation.  
26

27 In this way land use policies of increased density and land use mix can be reflected in shifts in spatial  
28 distribution of economic activity, location decisions, car ownership and use, and activity participation  
29 and destination choices (including decisions to participate in activities and to travel alone or with  
30 others). In building policy scenarios, we start with an assembly of data on the entire roadway network  
31 and its characteristics (roadway types, intersections and ramps, number of lanes, and speed limits), and  
32 the public transportation network (type of service, routes, and schedules). In parallel, we assemble data

1 on the resident population at different levels of geographical aggregation with core data at the zonal  
2 level of the pre-existing four-step model to enable use of existing forecasts and comparisons with  
3 simpler model systems. We also assemble external information about demographics, social and  
4 economic conditions of the study area, as well as any forecasts available. In this paper we focus on the  
5 first set of modeling blocks for which a finalized version was completed during the second phase of a  
6 project for SCAG and offer a few schematic and numerical examples of output that help us explain some  
7 of the findings, issues, and next steps.

8

### 9 *2.1 PopGen (Population Synthesis)*

10 In SimAGENT, the process starts with PopGen in which the entire resident population is synthetically  
11 generated/recreated person-by-person and household-by-household based on the method described in Ye  
12 et al., 2009, and further enhanced and improved for more recent applications (see  
13 <http://urbanmodel.asu.edu/popgen.html>). The input to this software and block of methods is the spatial  
14 organization of the simulated area in the form of zone-specific univariate distributions of resident person  
15 and household characteristics provided by the US Census and SCAG for the baseline year (in this case  
16 2003). As the population is recreated on a person-by-person and household-by-household basis, these  
17 distributions are used as the control totals for each spatial unit of analysis (approximately 4,000 Traffic  
18 Analysis Zones in this version of the model system) in an iterative algorithm that starts from a  
19 multivariate set of relationships (in essence a cross-tabulation) among the person and household  
20 variables used as seed information. For future years, these distributions are SCAG forecasts based on  
21 procedures of the growth forecast block in Figure 1 for 2020 and 2035 corresponding to the GHG target  
22 years. The multivariate set of relationships can be kept constant (assuming a steady state of  
23 demographic relationships) or can be changed to capture the impact of changing population composition  
24 and associated relationships including but not limited to age, birth rates, and household size.

25

### 26 *2.2 Accessibility by Time of Day*

27 To represent employment opportunities and the spatio-temporal distribution of activity participation  
28 opportunities, we also developed opportunity-based accessibility indicators at the level of the US Census  
29 block (203,191 US Census blocks cover the entire study area). In this way, we represent the ease (or  
30 difficulty) of reaching 15 different types of industries (representing the opportunities for activity  
31 participation) from each of these blocks within 10, 20, and 50 minutes of roadway travel buffers from  
32 each of the 203,000 pegs (Chen et al, 2011). The types of industries included in this version are: (a)

1 Agriculture, forestry, fishing and hunting and mining; (b) Construction; (c) Manufacturing; (d)  
2 Wholesale trade; (e) Retail trade; (f) Transportation and warehousing and utilities; (g) Information; (h)  
3 Finance, insurance, real estate and rental and leasing; (i) Professional, scientific, management,  
4 administrative, and waste management services; (j) Educational; (k) Health; (l) Arts, entertainment,  
5 recreation, accommodation and food services; (m) Armed forces; (n) Public administration; and (p)  
6 Other services (except public administration). Different accessibility values are obtained for the morning  
7 peak period (6 to 9 AM), midday (9 AM to 3 PM), evening peak period (3 to 7 PM), and at night (7 PM  
8 to 6 AM) capturing not only the different roadway conditions, but also the patterns of opening and  
9 closing of businesses during the day by allowing within each period above to also have different opening  
10 and closing hours of each industry type. Figure 2 provides an example of this spatial distribution by  
11 time of day. The top left hand quadrant also shows the percent of persons arriving and staying at the  
12 workplace (in this example banks and related institutions). As expected after 7:00 pm accessibility to  
13 these services is dramatically lower because the striking majority of these businesses are closed. The  
14 resident population with its detailed characteristics and a selection of indicators of the accessibility they  
15 enjoy are the inputs for the next block. Accessibility indicators are used in many of the behavioral  
16 models of the baseline year. Then, for subsequent simulation years they are modified based on the  
17 middle blocks of Figure 1 when the spatial distribution of economic activities change, and they are also  
18 modified based on travel times that may change based on network flow.

19

### 20 2.3. CEMSELTS (Long Term Choices)

21 This block of models was first developed as part of a larger model system (Eluru et al., 2008) and was  
22 modified and tailored to the SCAG region using local data. In CEMSELTS (Comprehensive  
23 Econometric Microsimulator of Socio-Economics, Land use and Transportation Systems), each person  
24 and household created in PopGen, and located in each zone of the study region, is given additional  
25 characteristics.

26

27 For example, when we examine persons in college a model is used to assign a college location, which is  
28 also a hierarchical function of accessibility. Workers are identified using a labor force participation  
29 model that is a function of age, gender, education, and presence of children in the household. Employed  
30 persons are then assigned (in a probabilistic way) to their type of industry, work location (which is also a  
31 function of accessibility), weekly work duration, and work flexibility. Each individual is also assigned a  
32 driver's license depending on age, gender, and race. Using these characteristics, household income is

1 computed as a function of race, presence of elderly individuals, education level of members of  
2 households, and employment industry of workers in the household. This is followed by a residential  
3 tenure model (own or rent) and a housing type model to assign each household to a single-family  
4 detached, single-family attached, apartment, and mobile home or trailer type of residence.

5  
6 An important model in this simulation system is car ownership and type. This type of model in essence  
7 determines the predicted non-commercial regional vehicle fleet mix that is used as input to the emission  
8 estimation software. This is also particularly important for California because of the expected market  
9 penetration of electric cars and the incentive programs created at the state and federal levels in the US to  
10 promote this type of technology. A model system like this can be used to assess different incentive  
11 structures promoting environmentally friendly technologies in cars. One of the inhibitors in building  
12 car ownership, car type and make models is the existence of many possible alternatives in this choice  
13 setting that includes many combinations of available alternatives. The solution here is to use Bhat's  
14 (2005) random utility model Multiple Discrete-Continues Extreme Value (MDCEV) model, which is  
15 capable of modeling multiple vehicle holdings, body types, fuel types, age, and use (miles)  
16 simultaneously. This model includes 55 alternatives for body type/vintage (9 body types – Subcompact  
17 Sedan, Compact Sedan, Mid-size Sedan, Large Sedan, Coupe, Cross-utility car, SUV, Van, Pickup) and  
18 5 vintage categories (New to 1 year, 2-3 years, 4-5 years, 6-9 years, 10-12 years, >12 years), 47  
19 alternatives for vehicle make (Ford, Chevrolet, Toyota, Honda, etc.), and several hundreds of models  
20 across the many body type/make combination categories. Both groups of models include as explanatory  
21 variables household composition indicators and residential accessibility reviewed above (Vyas et al.,  
22 2011). The model is currently being enhanced to incorporate the additional dimension of fuel type used.  
23 It also has the potential of expanding the pool of options to include commercial vehicle fleets. After this  
24 step, each of the household vehicles is allocated to a driver in each household based on a probabilistic  
25 model using as predictors variables such as gender, education, and employment.

26  
27 At this point of the simulation cascade on the left side of Figure 1, the model system produces the spatial  
28 distribution of all the residents by different social and demographic levels (including race) as well as  
29 employment and school locations assigned to each person. In addition, each household is assigned to a  
30 housing type. This resembles a complete Census of the resident population and can be done at any level  
31 of spatial aggregation. One could also draw samples from this population or proceed to the next step  
32 using 100% of the simulated residents. This is particularly convenient and useful in testing different  
33 policy scenarios to select a few to study in more detail. It is also possible to focus on a specific subarea

1 (e.g., a city) and perform more detailed analysis and modeling while keeping the rest of the region as an  
2 evolving background. The next set of model simulates the life of persons in a day.

#### 3 4 *2.4 CEMDAP (Daily Schedules and Choices)*

5 For each synthetically generated household and person within each household, daily activity and travel  
6 patterns are created in this block of models. To do so, a new modified version of the Comprehensive  
7 Econometric Microsimulator of Daily Activity-travel Patterns (CEMDAP  
8 <http://www.ce.utexas.edu/prof/bhat/CEMDAP.htm>) is used as the modeling engine that simulates  
9 activity-travel patterns of all individuals in the region for a 24 hour period along a continuous time axis.  
10 This model block creates synthetic schedules in two steps: (a) **the generation step** in which work and  
11 school activity participation and timing decisions are created, children's travel needs are estimated and  
12 an allocation of escort responsibilities to parents takes place, and independent and joint activity  
13 participation decisions are modeled; and (b) application of **the scheduling of activities** that produces the  
14 sequence of activities, with the departure and arrival times, activity duration(s), mode for each trip, and  
15 determination of the location of each activity. The models in this way create a complete description of  
16 the movement of each individual over space and time that is congruent with the movements of the rest of  
17 the household in which each person belongs. In this way, for each person, we have information about  
18 the type of activity, when, where, how long, with whom, in what sequence, and interrelationships with  
19 other persons and locations in the engagement pattern.

20  
21 In the generation step, working and student adults are first passed through models that predict if they  
22 will work or go to school in the simulated day. Then, they are given start and end times for their work  
23 and school activity. Conditional on this a household level MDCEV model is then used to simulate  
24 combinations of joint and solo activities for all persons. This provides an intra-household consistent  
25 schedule of activities and makes the entire simulation feasible because its formulation decreases the  
26 number of alternatives to simulate. Joint and solo activity durations are predicted for shopping,  
27 maintenance, social, entertainment, visit, active recreation, eat out, and other. In addition, durations for  
28 work-related and other serving passengers are also modeled. Activities are then arranged in tours  
29 (complete sequence of stops and trips starting and ending at the same location) and the modes used are  
30 predicted accordingly. The majority of these behavioral facets are modeled using econometric models  
31 (discrete choice models, hazard models, regression equation, etc.) that use as explanatory variables  
32 household composition and characteristics (e.g., household size, number of children, number of vehicles



1 owned), individual characteristics (e.g., gender, age, race/ethnicity, job type), location characteristics  
2 (e.g., opportunity-based accessibility indicators, infrastructure available, transit availability, population  
3 density) , and tour/mode/stops alternative attributes (e.g., in-vehicle travel time, distances to  
4 destinations. In addition to the typical tour regression models in the tour-based activity models (Vovsha  
5 et al., 2005, Bradley et al. 2010), SimAGENT also includes activity types (e.g., entertainment, eat out,  
6 recreation, serve passengers, etc) and duration of activities at each location (including at home, work,  
7 and school). The model system also includes tour modes that are drive alone, drive with passenger,  
8 shared ride, transit, and walk. The end result resembles Figure 3 in which we have two adults that go to  
9 work and a child going to school independently. In the evening they all go out for dinner. The trips,  
10 stops, activity types, activity start and ends times, modes are all determined by the simulation that  
11 includes 17,317,284 Persons in 5,721,914 Households. Simulation of a policy causes changes in  
12 destinations, activities alone and with others, durations, and modes creating “realistic” behavioral  
13 changes. For additional detail of this SimAGENT aspect see Bhat et al., 2011.

14  
15 Table 1, provides a brief description of a few key behavioral indicators from the entire population  
16 synthesized by SimAGENT (with heading SimAGENT In table 1), the survey based on which  
17 behavioral equations were estimated and included in the simulation code, and weighted survey averages  
18 using the sampling weights provided by NUSTAS, 2003. Table 1 also reports the number of  
19 observations that had at least one trip (or one minute of activity for the durations) with activities  
20 measured in hours per day. The worktrips include trips to work and back from work. Schooltrips  
21 include trips to school and back from school. The shopping trips include trips to shopping (or  
22 equivalently the stops with shopping as the activity at the stop). Also, “Activity Duration” in Table 1 is  
23 the time between the departure time of the first trip from home and the arrival time of the last trip at  
24 home. Overall there is a general agreement between survey data and simulated population by  
25 SimAGENT. The standard deviations reported in Table 1 also show a general agreement with similar  
26 magnitude between simulated population and survey data with exception the school trips for which all  
27 school age children have exactly two trips per day. The output of the model system contains more  
28 complete data than an activity survey diary database because it recreates the entire population of the  
29 SCAG region and does not have any missing data for each person within each household. This is an  
30 important consideration for verification and validation because the simulation produces more  
31 information than is currently available from other sources to use as the gold standard or benchmark. At  
32 this point, the output from CEMDAP can be used in many different ways. For example, we have  
33 developed and pilot tested policy scenarios and studied their impact in timing decisions of individuals

1 (e.g., advancing or postponing the starting of trips). We also coupled this output with the more  
2 traditional four-step model routines to perform traffic assignment and emission estimation. Moreover,  
3 we are also advancing along the path of using detailed routing algorithms that can track the simulation  
4 of individual vehicles and eventually compute emissions at finer spatial and temporal resolutions  
5 moving us along the model development path of TASHA in Toronto (Hao et al., 2010).

### 7 *2.5 Routes, Assignment to Networks, and Emissions*

8 The output of CEMDAP in this block is aggregated (i.e., converting person trips to zonal sums of trips)  
9 for each of the 4192 zones to create a trip interchange matrix of Origins-Destinations (OD) among the  
10 traffic analysis zones at different time periods in a day using the same four periods presented previously  
11 for the accessibility indicators. These are combined with similar ODs for heavy vehicles (trucks), and  
12 vehicular OD from and to ports and airports, as well as ODs of traffic generated outside the region.  
13 These additional OD matrices are the same as the four-step model used by SCAG in its 2008 RTP.  
14 Using these trip interchange matrices, traffic assignment (i.e., vehicular origins and destinations are  
15 converted into traffic volumes on the highway network based on behavioral and mathematical  
16 principles) produces estimates of traffic volumes on the links of the network allowing comparisons  
17 among different methods. Figure 4 is the output of traffic assignment during the AM peak, midday, and  
18 midnight periods showing vehicle flow per hour. This figure also shows the added output from  
19 SimAGENT, which is the number of persons at each location (in this case traffic analysis zone  
20 centroids) by each activity type engaged in for that specific hour (7:00 to 8:00 AM and 11:00 AM to  
21 12:00 Noon in Figure 4). During the day the number of persons at work (light green in Figure 4) and  
22 school (gold color in Figure 4) are substantial. In the evening hours the maps are dominated by blue  
23 color, which is for stay at home activity. In addition, using the typical way of verifying outputs of  
24 assignment we examine how close to the observed traffic are traffic volumes estimated by different  
25 models. To this end, agencies identify strategic locations forming a ring around major attractors and use  
26 them as benchmarks (called screenline and contains 23 locations in this case). Figure 5 shows each of  
27 these locations and the relative ‘closeness’ reproduced by the trip based aggregated four-step model  
28 (which is iterated to match these daily traffic counts) and the SimAGENT, which achieves this closeness  
29 in one step with no additional adjustments or iterations.

30  
31 Some additional tasks of feasibility testing are also complete during this phase of model development.  
32 In fact, at the time this paper was written, experiments with alternatives to the multi-period static traffic

1 assignment were also completed. One of these experiments uses TRANSIMS along similar lines of  
2 development as in Lawe et al., 2010, with an application that aims at recreating the detailed stops along  
3 routes that vehicles experience in the simulation using the highway and transit networks. In this  
4 application, approximately 64.5 million activity records and 156.5 million travel plans are routed in the  
5 network. Table 2 is an example of two households (with household identification numbers 74, and 468).  
6 The Table also shows the person identification in the household and each activity simulated. Two  
7 persons from household 74 did not have out of home activities. Table 2 also shows the duration of each  
8 activity, location where it was completed, mode used to arrive to each activity location, and other data  
9 about interaction in household of activity and travel, and the vehicle used. The bottom half of the table  
10 also shows tracking of vehicle information and the assignment to the main driver within the household.  
11 In this way activities are sorted by person in each household and vehicles are connected to each  
12 household by the vehicle ID. This output from SimAGENT is restructured to become input for  
13 TRANSIMS. This requires adding a home-based activity from the start of the day to the time when the  
14 first actual activity begins and other simple conversion of data shown in Table 3. One of the challenges  
15 during the process of converting SimAGENT activity records to TRANSIMS activities is the activity  
16 location assignment. The original activities are defined as zone-to-zone daily activities. However  
17 TRANSIMS needs the activities and the movements from one activity location to another at finer spatial  
18 resolution. For each group of activity locations in each of the 4109 zones we randomly assign to each an  
19 activity location within the zone using as seed information observed activity locations. In addition, after  
20 all the activities of one household are assigned, we adjust the first and the last activity location of  
21 household members to the same location if they have the same zone ID and assume this is the home  
22 location of this household (in essence forcing all recorded persons to start and end their day at home and  
23 giving them an exact location of home).

24  
25 Figure 6 shows the daily paths of these two households. Household “74” has 5 persons (with ids 7401,  
26 7402, 7403, 7404, 7405) but the 3<sup>rd</sup> and 4<sup>th</sup> persons are not included because they have no travel. The  
27 second household “468” has all its four persons (46801, 46802, 46803, and 46804) represented in the  
28 tables and paths of Figure 6. The TRANSIMS router uses the SimAGENT input data and generates a  
29 series of activity path records which are composed of travel mode, time period of travel, origin-  
30 destination locations, turning points on the path, and so forth. This information is stored in an  
31 Arcview® polyline file to show the travel path on the network. The entirety of paths for each household  
32 is a travel plan and the two travel plans of the households are shown in Figure 6. The household  
33 members of the right hand side sketch with ids “46802” and “46804” have trips that did not get routed

1 on the TRANSIMS network and they directly move from an origin to the destination with a travel mode  
2 of “Magic Move”. This is because the persons traveled by “School Bus” (Mode = “6”) and “Driven by  
3 Parent (for child)” (Mode = “4”) and VEHID = “-1” which could not be routed on the network. This is  
4 an example of the type of details that need to be post-processed in addition to a variety of other  
5 comparisons among different assignment algorithms and sensitivity analysis that are planned for the  
6 next phase in model development.

7  
8 Testing of emissions estimation was also performed at this stage of model development using the more  
9 traditional approach that is also used in conformity studies with the four step model, but this time  
10 employing SimAGENT multi-period static traffic assignment. The output of the traffic assignment by  
11 different time periods in a day is then used as input to the software EMFAC (a California region-tailored  
12 emissions calculating software) that produces estimates of fuel consumed and GHG emissions. Table 1  
13 compares the EMFAC output between the two model systems (4-step and simAGENT). Both systems  
14 use the same truck traffic, special generators (ports and airports), and external to the study area traffic.  
15 The model year for regional fleet composition is 2003. We also experimented with different year-by-  
16 year fleet compositions in EMFAC2007 and years of simulation showing a dramatic decrease in  
17 pollutant emissions but very low sensitivity to CO2 emissions due to the fairly constant fuel efficiency  
18 of the technology groups in the software.

### 19 20 **3. Future Research and Analysis**

21 Although the SimAGENT version described here appears to work well for the policy analysis purpose  
22 that was designed, it is currently being tested and strategically modified using external data for  
23 verification and validation. In doing this we identified a variety of methodologically thorny issues about  
24 verification and validation that would be appropriate for microsimulated environments and the daily life  
25 of their residents that do not have a readily available solution or benchmark data. At the time of writing  
26 this paper, we completed an extensive accessibility computation exercise using a combined network of  
27 walking and transit for comparison with highway-base accessibility indicators (Lei, 2011). This new set  
28 of indicators can be used for mode/destination type of models. However, this requires extensive testing  
29 of the algorithms used and re-estimation of the different behavioral facets in the current version of  
30 SimAGENT. We also envision the development of a next version of model system that may be used in  
31 the next cycle of regional model development in SCAG that will include a much richer and larger zonal  
32 system of approximately 12,000 zones and finer network for all travel components.

1 From a behavioral viewpoint, we also performed sensitivity to a variety of explanatory variables used in  
2 the regression models that show high sensitivity to land use (e.g., density, mixed land use, and distance  
3 to major centers), high sensitivity to travel times, and low sensitivity to costs. This is particularly  
4 important when interfacing and integrating SimAGENT with other simulation software and we include  
5 in the next development phase a task to study the relationships depicted at different scales between  
6 travel costs and the spatial distribution of activities. In the next few months we are also planning to  
7 include a MATSIM (Gao et al., 2010, Bekhor et al., 2011) application that has important similarities and  
8 differences with TRANSIMS to explore additional next steps. Moreover, we are developing the  
9 component labeled “household evolution” in Figure 1 that provides the demographic microsimulation  
10 needed to evolve the synthetic population in multiple years. In 2010 California started the design and  
11 implementation of the California Household Travel Survey, which is currently in pilot testing and  
12 expected completion by early 2013. This survey will provide many of the data required for  
13 SimAGENT. One final comment is about computational speed and efficiency. In the experiments we  
14 performed with this model system, computational time can be reduced dramatically employing multiple  
15 core computers reducing computational time substantially as one would expect. However, the most time  
16 consuming task is the exchange of data among different modules with many of them requiring data  
17 manipulation by researchers. This calls for the design of interfaces among modules that are coded in  
18 different languages and platforms, an issue that we leave for the future.

## 21 **Acknowledgments**

22  
23 Funding for this project was provided by the Southern California Association of Governments, the  
24 University of California (UC) Lab Fees program through a grant to UCSB on Next Generation Agent-  
25 based Simulation , and the UC Multicampus Research Program Initiative on Sustainable Transportation.  
26 Past research grants to UCSB from the University of California Transportation Center (funded by US  
27 DOT RITA and Caltrans), to UT Austin from the Texas Department of Transportation, and to Arizona  
28 State University by the Federal Highway Administration have also supported the development of  
29 methods used in this effort. This paper does not constitute a policy or regulation of any public agency.

## References

- Bekhor S., C. Dobler, and K. Axhausen (2011) Integration of Activity-Based with Agent-Based Models: an Example from the Tel Aviv Model and MATSim. Paper presented at the 90<sup>th</sup> Annual Transportation Research Board Meeting, Washington D.C., January 23-27, 2011
- Bhat, C.R. (2005), "A Multiple Discrete-Continuous Extreme Value Model: Formulation and Application to Discretionary Time-Use Decisions," *Transportation Research Part B*, Vol. 39, No. 8, pp. 679-707.
- Bhat, C.R., Paleti R., Pendyala R., K.G. Goulias, and H. Hu (2011) Intrahousehold Interaction with MDCEV for SimAGENT. Paper submitted for presentation at the 2012 Annual Transportation Research Board Meeting.
- Bradley, M., J. L. Bowman and B. Griesenbeck (2010) SACSIM: An applied activity-based model system with fine-level spatial and temporal resolution, *Journal of Choice Modeling* Vol, 3(1) 2010.
- Chen, Y., S. Ravulaparthi, K. Deutsch, P. Dalal, S.Y. Yoon, T. Lei, K.G. Goulias, R.M. Pendyala, C.R. Bhat, and H-H. Hu (2011) Development of Opportunity-based Accessibility Indicators, Paper presented at the 90<sup>th</sup> Annual Transportation Research Board Meeting, Washington D.C., January 23-27, 2011 and *Transportation Research Record (in press)*.
- Eluru, N., A.R. Pinjari, J.Y. Guo, I.N. Sener, S. Srinivasan, R.B. Copperman, and C.R. Bhat (2008), "Population Updating System Structures and Models Embedded in the Comprehensive Econometric Microsimulator for Urban Systems," *Transportation Research Record*, Vol. 2076, pp. 171-182..
- Gao W., M. Balmer, and E. Miller (2010) Comparison of MATSim and EMME/2 on Greater Toronto and Hamilton Area Network, Canada. *Transportation Research Record*, Vol. 2176, pp. 118-128.
- Hao J. Y., Hatzopoulou M., and E. Miller (2010) Integrating an Activity-Based Travel Demand Model with Dynamic Traffic Assignment and Emission Models: Implementation in the Greater Toronto, Canada, Area. *Transportation Research Record*, Vol. 2176, pp. 1-13.
- Nustats. (2003). Southern California Association of Governments Post Census Regional Household Travel Survey: Data User's Manual, June 2003.
- Lawe S., J. Lobb, A.W. Sadek, Huang S. and C. Xie (2010) TRANSIMS Implementation in Chittenden County, Vermont: Development, Calibration, and Preliminary Sensitivity Analysis. *Transportation Research Record*, Vol. 2132, pp. 113-121.
- Lei T., Chen Y. and K. G. Goulias (2011) Opportunity-based Dynamic Transit Accessibility in Southern California: Measurement, Findings, and a Comparison with Automobile Accessibility. GeoTrans working paper, 07-01-11, Santa Barbara, CA.

1 Vovsha, P., Bradley M, Bowman J. (2005) Activity-Based Travel Forecasting Models in the United  
2 States: Progress since 1995 and Prospects for the Future. In Progress in Activity-Based Analysis, H.  
3 Timmermans (ed.), Elsevier Science Ltd, Oxford, UK, 389-414.

4  
5 Vyas G., Bhat C.R., Pendyala R., K.G. Goulias, and H. Hu (2011) MDCEV vehicle holding model  
6 implemented in SimAGENT. Paper submitted for presentation at the 2012 Annual Transportation  
7 Research Board Meeting.

8  
9 Ye X., Konduri K., Pendyala R.M., Sana B., and P. Wadell (2009), “A Methodology to Match  
10 Distributions of both Household and Person Attributes in the Generation of Synthetic Populations.”  
11 Paper presented at the 2009 Transportation Research Board Meeting, Washington D.C. (also  
12 available at [http://urbanmodel.asu.edu/popgen/papers/PopulationSynthesizerPaper\\_TRB.pdf](http://urbanmodel.asu.edu/popgen/papers/PopulationSynthesizerPaper_TRB.pdf)).

13  
14

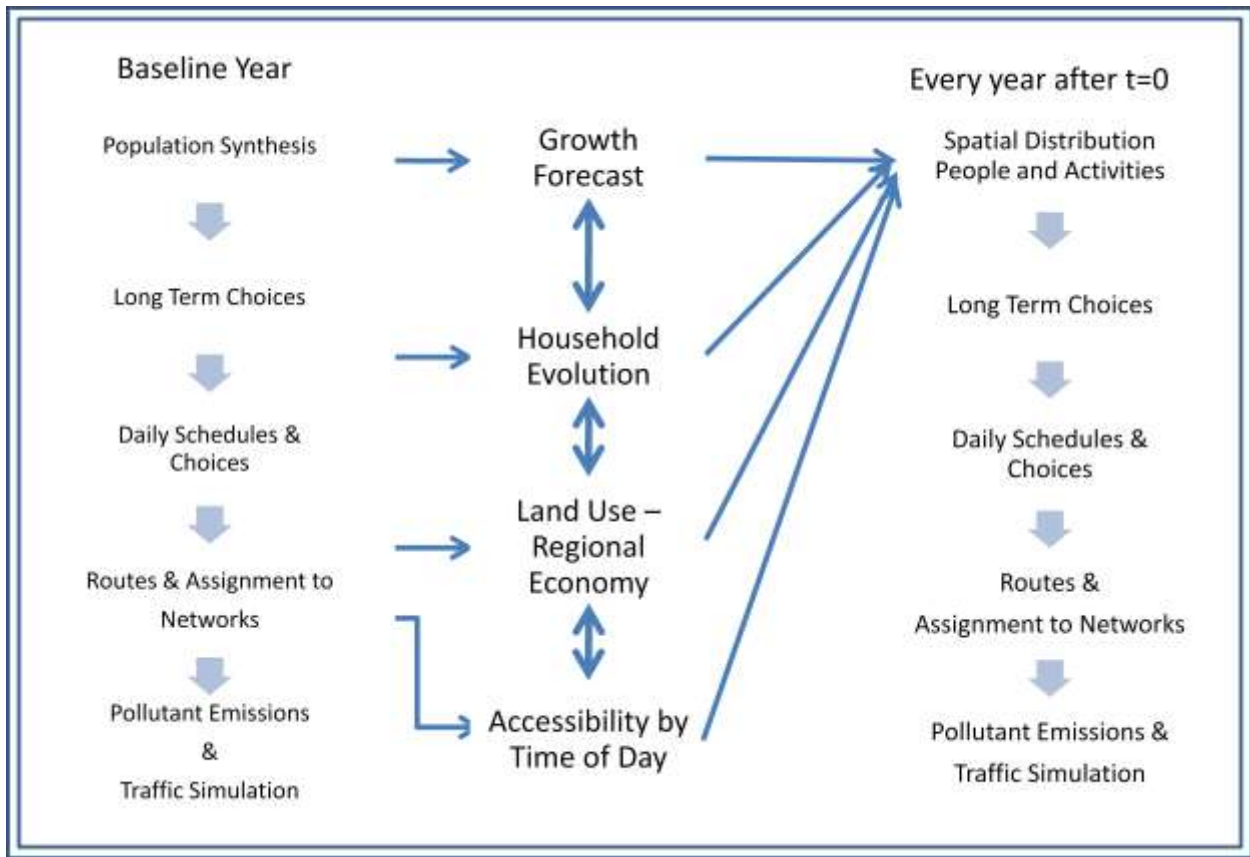


Figure 1 Schematic Representation of SimAGENT Blocks

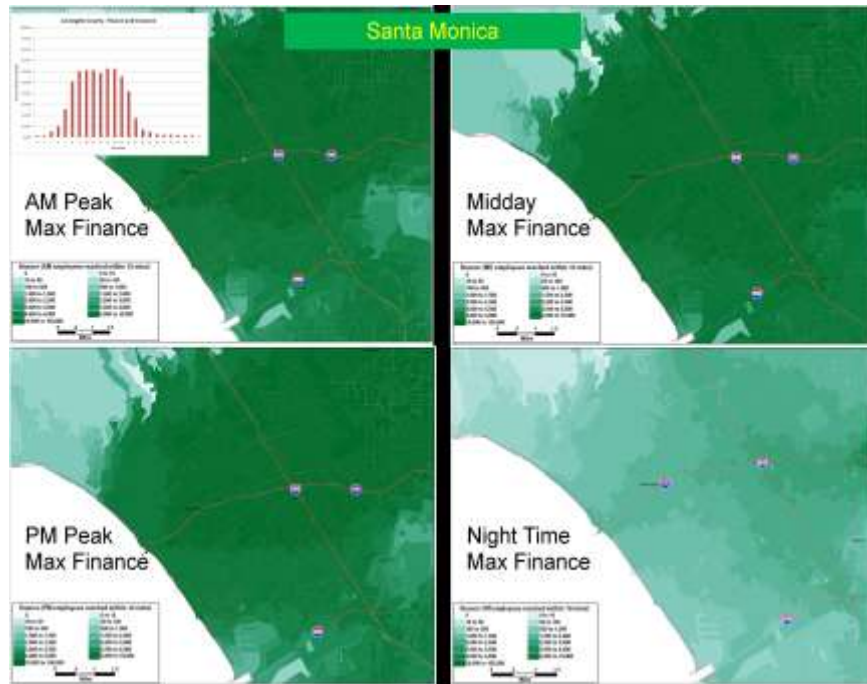


Figure 2 Time of Day Accessibility Map Example in Santa Monica, CA

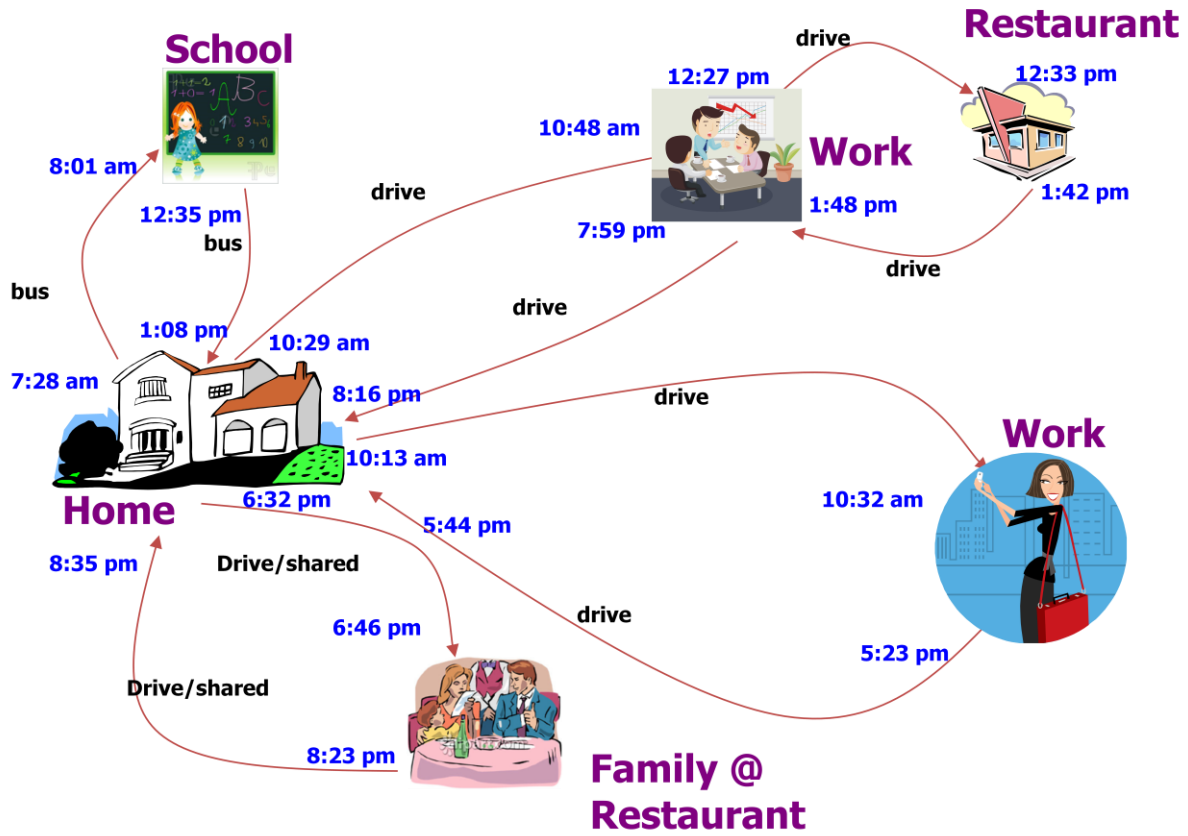
492  
493  
494  
495

496  
497  
498



1

2



3

4

5

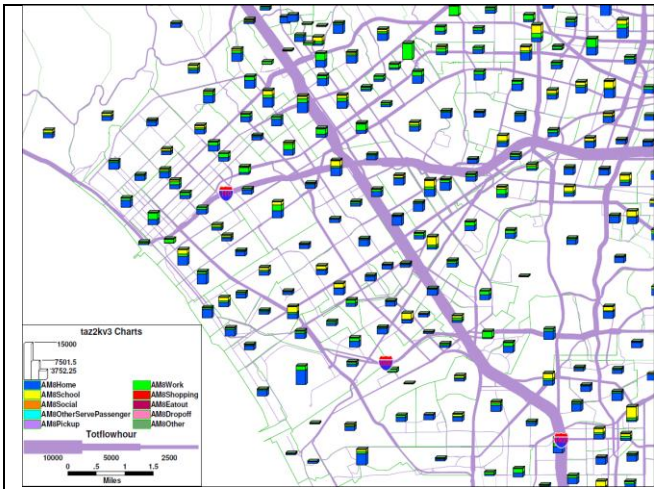
6

7

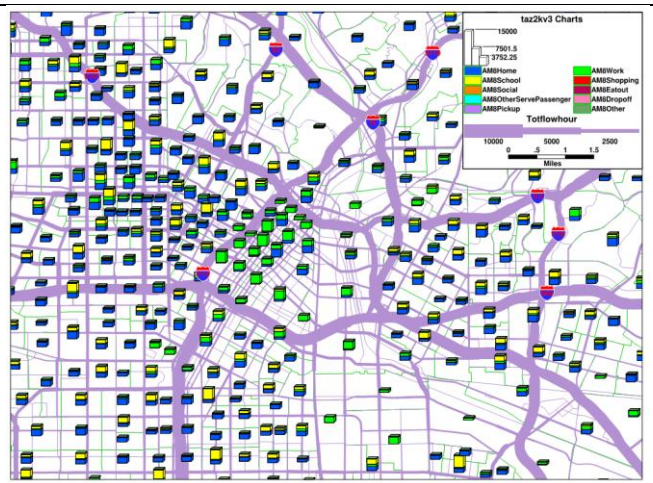
8

9

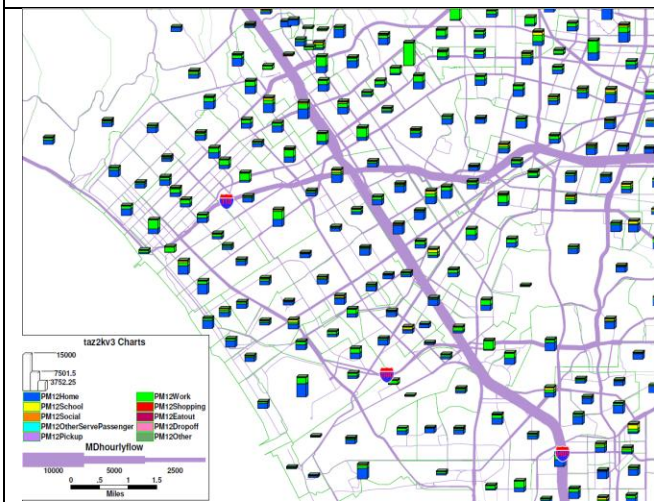
Figure 3 Example of a Daily Household Schedule



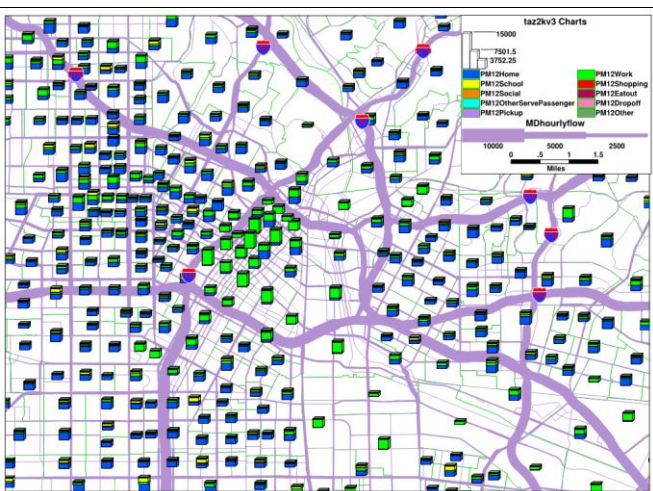
Santa Monica From 7:00 AM to 8:00 AM



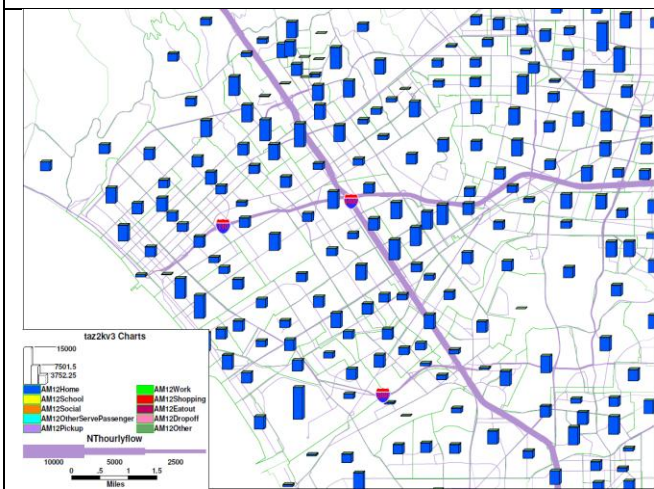
LA Downtown From 11:00 AM to 12:00 Noon



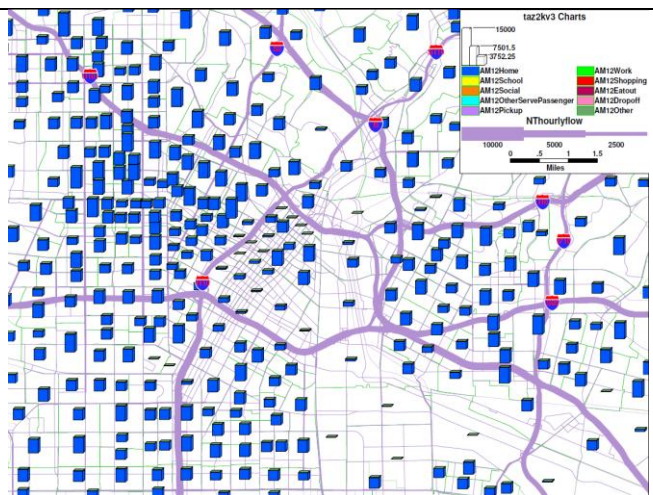
Santa Monica From 11:00 AM to 12:00 Noon



LA Downtown From 11:00 AM to 12:00 Noon

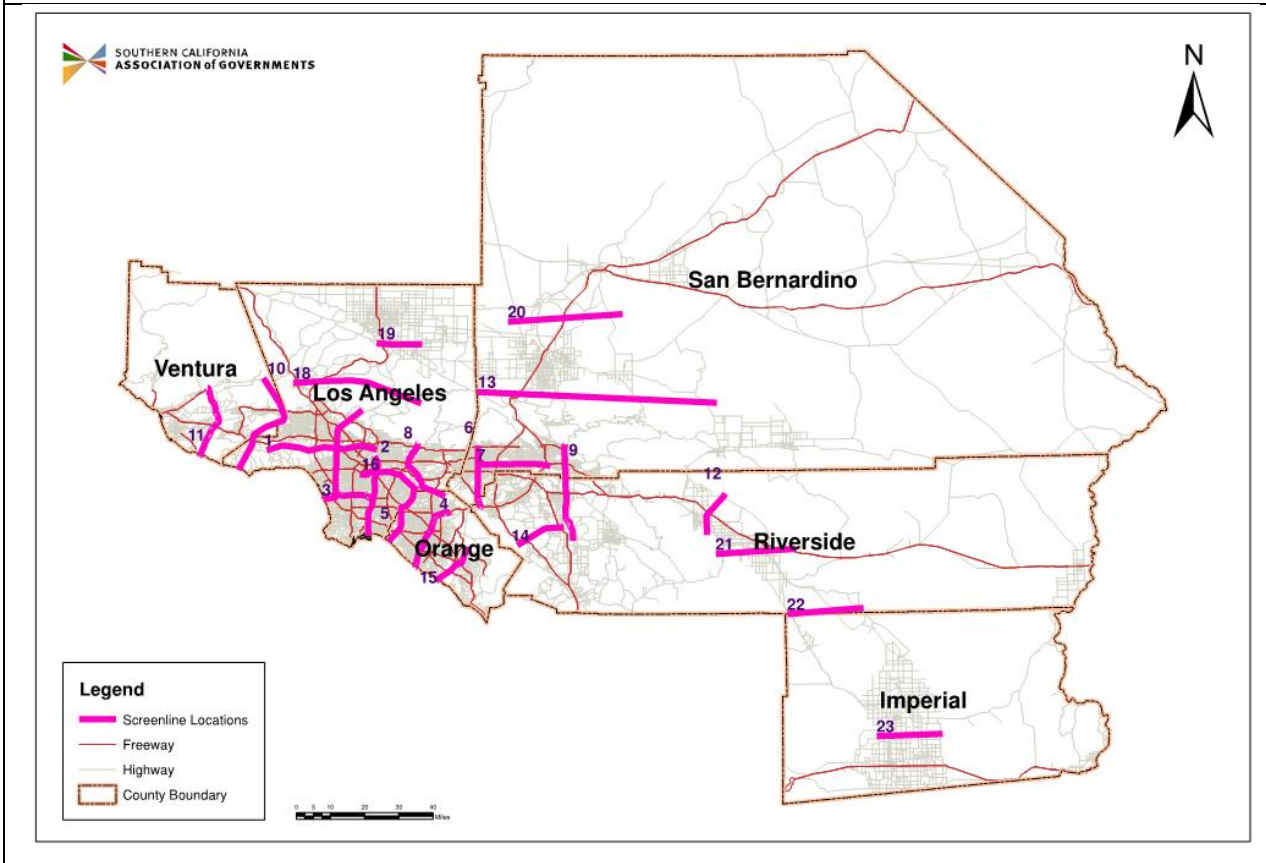
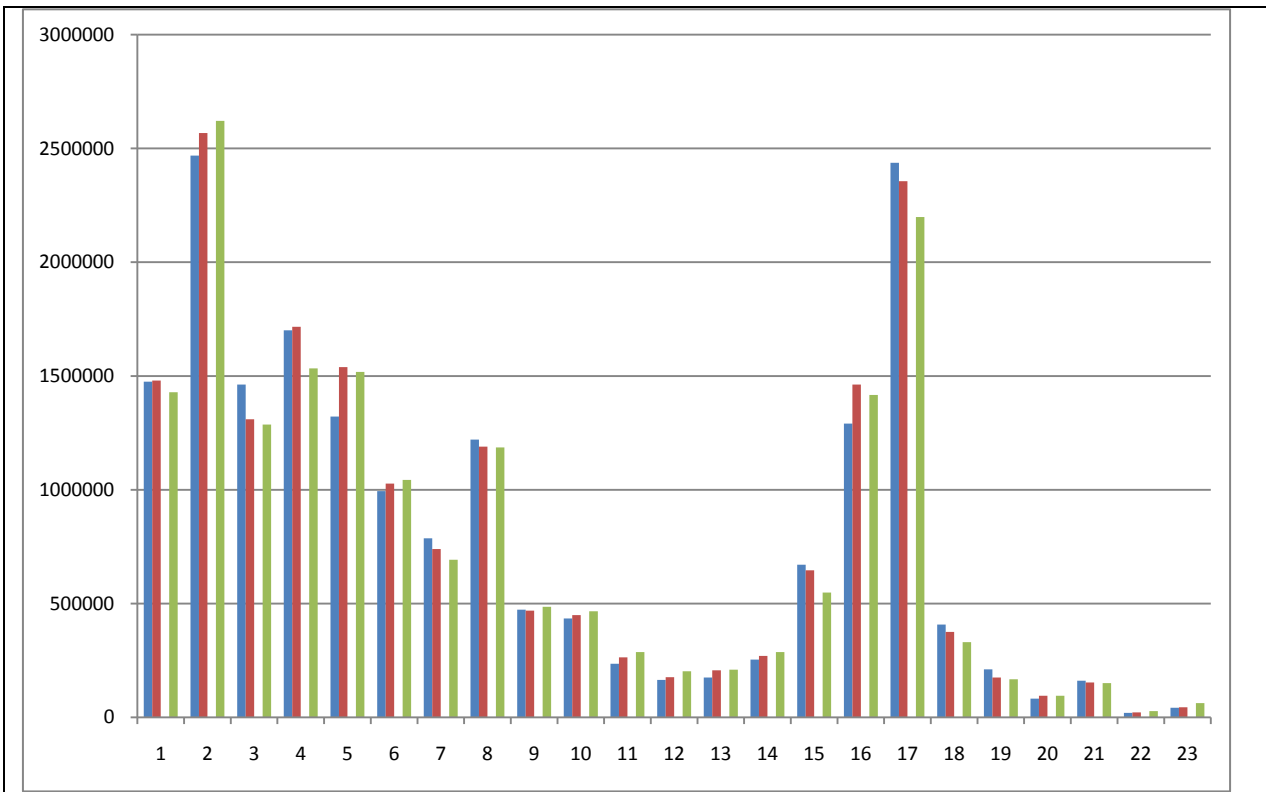


Santa Monica From 11:00 PM to 12:00 AM

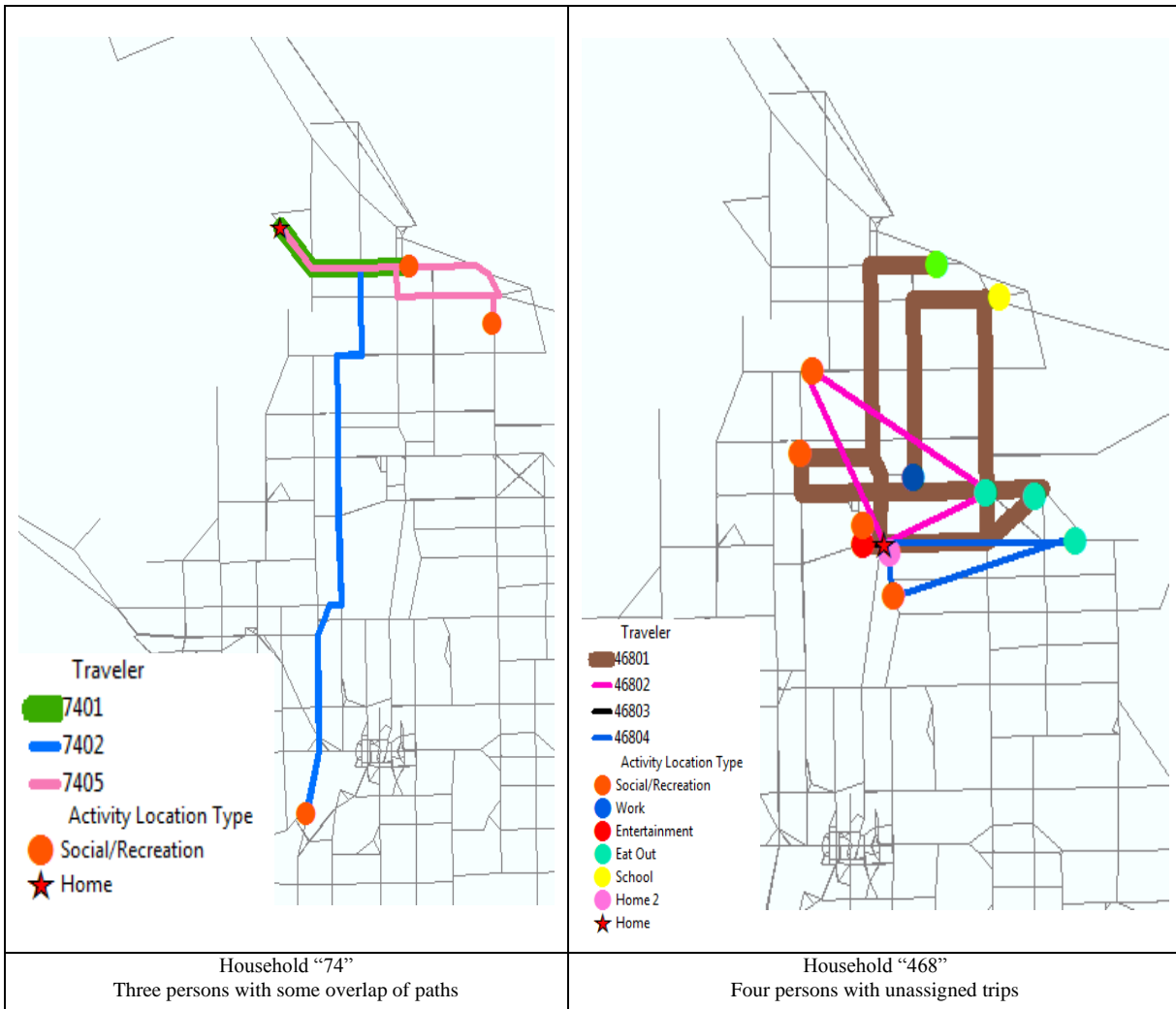


LA Downtown From 11:00 PM to 12:00 AM

1 **Figure 4 Assigned Traffic Output in SimAGENT and Number of Persons at Each Location by**  
 2 **Each Activity Type.**  
 3



1 **Figure 5 Screenline Daily Traffic Counts Comparisons at 23 different network locations (red is**  
 2 **trip based model, green is SimAGENT, and blue is from observed counts)**  
 3



2  
3  
4  
5  
6

**Figure 6: Travel Plans for Households 74 and 468**

Table 1 SimAGENT Average per Person per Day Characteristics Compared to Survey Data

	Survey			Survey Weighted*			SimAgent		
	Mean	Std.	N**	Mean	Std.	N**	Mean	Std.	N**
Home Based Work	1.77	0.638	8,815	1.76	0.604	9,011	1.71	0.456	4,226,380
Home Based Other	2.81	1.653	22,135	2.83	1.669	26,165	2.70	1.632	11,301,831
Non-Home Based	2.45	1.858	13,164	2.37	1.811	14,204	2.46	1.760	5,304,019
All trips	4.11	2.576	26,754	4.01	2.550	30,962	3.70	2.522	13,728,666
# of work trip	2.37	0.948	9,980	2.32	0.889	10,347	2.58	1.124	4,544,427
# of school trip	2.05	0.343	4,712	2.03	0.280	7,100	2.00	0.001	2,790,932
# of shopping trip	1.47	0.861	6,990	1.42	0.815	7,116	1.84	1.198	2,745,813
# of other trip	3.40	2.388	19,841	3.42	2.425	22,329	3.21	2.411	9,008,573
work duration	8.14	2.675	9,837	8.15	2.633	10,182	8.06	2.184	4,544,427
school duration	6.32	2.092	4,707	6.48	1.983	7,089	5.53	1.563	2,790,932
shopping duration	0.97	0.978	6,990	0.98	1.016	7,116	1.71	1.850	2,745,813
Other activity duration	12.51	3.903	26,690	12.58	3.908	30,898	12.69	5.320	9,008,573
# of go-home trips	1.48	0.784	26,295	1.48	0.798	30,482	1.39	0.742	13,728,666
Activity Duration	8.50	4.454	26,754	8.49	4.339	30,962	7.77	4.681	13,728,666

\*sample weights are provided by NUSTATS (2003) to adjust the sample and match known population characteristics. \*\*Number of observations (persons) with at least one trip for trip averages here or one minute of activity for activity durations reported here.

1  
2  
3  
4  
5

**Table 2 Example of SimAGENT Activity and Vehicle Data**

SimAGENT Activity										
HID	PID	TID	ActType	Duration	ZoneID	ArriveT	Mode	JNTNUM	JNTRVL	VEHID
74	1	1	8	30	101010004	228	7	0	0	1
74	1	1	12	1170	101010000	270	7	0	0	1
74	2	1	2	40	101030002	507	0	0	0	2
74	2	1	12	853	101010000	587	0	0	0	2
74	5	1	7	31	101010003	752	0	0	0	0
74	5	1	12	640	101010000	800	0	0	0	0
468	1	1	11	5	101010001	538	0	0	0	0
468	1	1	3	14	101010002	552	0	0	0	0
468	1	1	12	375	101010001	573	0	0	0	0
468	1	2	3	18	101010001	951	0	0	0	0
468	1	2	12	27	101010001	972	0	0	0	0
468	1	3	9	2	101010003	1008	0	0	0	0
468	1	3	12	121	101010001	1023	0	0	0	0
468	1	4	0	3	101010004	1151	0	0	0	0
468	1	4	12	277	101010001	1163	0	0	0	0
468	2	1	15	243	101010001	295	6	0	0	-1
468	2	0	3	14	101010002	552	4	0	0	-1
468	2	0	12	867	101010001	573	4	0	0	-1
468	3	1	15	290	101010001	282	2	0	0	-1
468	3	0	14	843	101010001	596	2	0	0	-1
468	4	1	15	258	101010001	280	2	0	0	-1
468	4	0	3	14	101010002	552	4	0	0	-1
468	4	0	12	867	101010001	573	4	0	0	-1

SimAGENT Vehicle							
HID	VID	BODTYP	AGECAT	MAKE	PRIMDRV	ANNMIL	THREAD
74	0	2	2	13	5	45.4288	1
74	1	3	6	2	1	45.4288	1
74	2	1	3	13	2	45.4288	1
468	0	9	6	5	1	13.61	1

1

Table 3 Example of TRANSIMS Activity and Vehicle Data

HHOLD	PERSON	ACTIVITY	PURPOSE	START	END	DURATION	MODE	VEHICLE	LOCATION	ZONE	PASSENGER
74	1	1	12	0	3.8	3.8	2	741	172838	4013	0
74	1	2	8	3.8	4.3	0.5	2	741	129412	4062	1
74	1	3	12	4.5	24	19.5	2	741	172838	4013	1
74	2	1	12	0	8.45	8.45	2	742	172838	4013	0
74	2	2	2	8.45	9.117	0.667	2	742	172421	4018	0
74	2	3	12	9.783	24	14.217	2	742	172838	4013	0
74	5	1	12	0	12.533	12.533	2	740	172838	4013	0
74	5	2	7	12.533	13.05	0.517	2	740	131714	4086	0
74	5	3	12	13.333	24	10.667	2	740	172838	4013	0
468	1	3	3	9.2	9.433	0.233	2	4680	172959	4088	0
468	1	4	12	9.55	15.8	6.25	2	4680	131311	4064	0
468	1	5	3	15.85	16.15	0.3	2	4680	131304	4064	0
468	1	6	12	16.2	16.65	0.45	2	4680	131304	4064	0
468	1	7	9	16.8	16.833	0.033	2	4680	172967	4086	0
468	1	8	12	17.05	19.067	2.017	2	4680	131079	4064	0
468	1	9	0	19.183	19.233	0.05	2	4680	129347	4062	0
468	1	10	12	19.383	24	4.617	2	4680	131379	4064	0
468	2	1	12	0	4.917	4.917	2	-1	131379	4064	0
468	2	2	15	4.917	8.967	4.05	9	-1	131671	4064	0
468	2	3	3	9.2	9.433	0.233	10	-1	131491	4088	0
468	2	4	12	9.55	24	14.45	10	-1	131379	4064	0
468	3	1	12	0	4.7	4.7	2	-1	131379	4064	0
468	3	2	15	4.7	9.533	4.833	1	-1	131372	4064	0
468	3	3	14	9.933	23.983	14.05	1	-1	131379	4064	0
468	4	1	12	0	4.667	4.667	2	-1	131379	4064	0
468	4	2	15	4.667	8.967	4.3	1	-1	131185	4064	0
468	4	3	3	9.2	9.433	0.233	10	-1	131569	4088	0
468	4	4	12	9.55	24	14.45	10	-1	131379	4064	0

2

TRANSIMS Vehicle										
VEHICLE	HHOLD	LOCATION	TYPE	SUBTYPE	BODTYP	AGECAT	MAKE	PRIMDRV	ANNMIL	THREAD
740	74	172838	1	0	2	2	13	5	45.4288	1
741	74	172838	1	0	3	6	2	1	45.4288	1
742	74	172838	1	0	1	3	13	2	45.4288	1
4680	468	131379	1	0	9	6	5	1	13.6082	1

3

4

5

1

**Table 4 Aggregate-Level Comparison Between 4-step and SimAGENT(for the Year 2003)**

Emissions and Fuel Consumption	4-Step Model*	SimAGENT Baseline*
Organic Gases (g/mile)	0.943	0.926
CO (g/mile)	9.498	9.348
NOx (g/mile)	1.929	1.955
CO2 (g/mile)	561.340	543.545
Gasoline (gallons/mile)	0.051	0.050
Gasoline (mile/ gallons)	19.377	20.203
Diesel (gallons/mile)	0.102	0.101
Diesel (mile /gallons)	9.833	9.893
Organic Gases (g/person-day)	22.291	21.333
CO (g/person-day)	224.553	215.388
NOx (g/person-day)	45.606	45.050
CO2 (g/person-day)	13271.790	12524.452
Vehicle Miles Travel/person-day	23.643	23.042

2

\*Both models use the same regional fleet distribution for SCAG, which are classified by technology groups for exhaust and other emissions. In the passenger cars fleet, these include approximately 70% in the technology group 23 (model years 1997 to 2005 that are LEV), 20% in the technology group 24 (1997 to 2005 that are ULEV), and 8.4% in the technology group 31 that are Partial ZEV), and the rest are a mix of a variety of groups. Additional details can be found at EMFAC2007 version 2.3 User's Guide ([http://www.arb.ca.gov/msei/onroad/downloads/docs/user\\_guide\\_emfac2007.pdf](http://www.arb.ca.gov/msei/onroad/downloads/docs/user_guide_emfac2007.pdf)).

7