PREPARING DATA FOR MICROSIMULATION OF SMART TRANSPORT SYSTEMS: A TOOL FOR EXPLOITING EXISTING TRAVEL MODEL RESULTS AND OPEN-SOURCE DATA

Jun Liu, Ph.D.
Postdoctoral Fellow
Center for Transportation Research
The University of Texas at Austin
jun.liu@utexas.edu

Kara M. Kockelman, Ph.D., P.E.
Professor, and E.P. Schoch Professor in Engineering
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin – 6.9 E. Cockrell Jr. Hall
kkockelm@mail.utexas.edu

Published in *Smart Transport for Cities & Nations: The Rise of Self-Driving & Connected Vehicles* (2018)

Abstract

Car-sharing offers travelers an alternative method of transport in or between cities with the transformative implementation of connected and autonomous vehicles (CAVs) likely further promote the sharing. To provide decision makers reasonable information about car-sharing strategies or shared CAVs, transportation planners and researchers are looking for advanced travel modeling approaches. Activity-based modeling (ABM) is one of the most promising approaches, modeling travel demand at person-level and offering great temporal and spatial details about individuals' travel patterns. Currently, the four-step travel demand modeling process is the most commonly used approach which is trip-based, modeling travel demand at an aggregated level of Traffic Analysis Zones (TAZs). However, this approach is unable to track individuals' travel patterns with great spatial details. ABM can be used to estimate the impacts of car-sharing in transportation systems and evaluate the policies/ strategies related to the CAV operations. ABM takes the individual's daily activities chained by a series of travel trips, namely a tour, as the travel demand input. The input can be simply summarized into "4Ws": Who this individual is, where this individual lives and works, what daily activities this individual person does, and when this individual plans to perform activities. This study delivers a methodological framework to prepare the "4W" inputs, taking advantage of existing travel model data (including the travel survey data) and open source data (e.g., OpenStreetMaps). This paper presents a programming-based tool composed of a series of algorithms that output synthetic population, synthetic locations for activities, travel tours (i.e., chained trips and activities), and travel schedules for performing activities, respectively. The tool is particularly useful for planning practitioners from state agencies and regional planning organizations who already have the data (e.g., regional travel models and travel survey data) and seek to convert their existing trip-based models to activity-based models that may be more suitable for simulating the individuals' travel patterns.

Keywords: Travel demand data, Activity-based modeling, Synthetic population, Travel tour, Open source

INTRODUCTION

Observation of travel patterns is evolving in many aspects including the new operational strategies using existing transportation tools (e.g., vehicle/ride sharing) and upcoming transportation innovations (e.g., connected and autonomous vehicles). The emerging travel patterns may require advanced modeling techniques for traffic forecasting and evaluations of transportation policies and projects.

Car-sharing is transforming the way people travel, live and socialize (*1*). Including Uber, Car2go, Lyft, Zipcar, Hertz and Enterprise, there were more than 35 major car-sharing industrial participants/competitors in North America that managed or operated more than 25 thousand shared vehicles in July 2015 (*2*). Car-sharing offers mobility to travelers without the burden of owning a vehicle and the car-sharing services are more flexible than transit (*3*). Emerging transportation tools such as connected and autonomous vehicles (CAVs) will further facilitate the growth of the car-sharing market. CAVs are expected to significantly improve surface transportation systems from three aspects: safety (*4*), mobility (*5-7*) and sustainability (*8, 9*).

Currently, most in-use state and regional travel models are "four-step" trip-based (*10*) and these models are aggregated at the level of traffic analysis zones (TAZs) with trip origins and destinations at TAZ centroids. New travel patterns require the modeling of individual trips (rather than aggregated trips between TAZs) at great spatial and temporal details. For example, the car-sharing system needs a model to capture how a service may connect two individual trips. If two trips are connected in the same TAZ, the four-step travel model is unable to capture such car-sharing patterns. Therefore, people are seeking advanced travel modeling approaches; and activity-based modeling (ABM) is considered one of the most promising approaches. ABM is based on the principle that travel demand/trips are derived from activities that people plan to perform daily (*11*). ABM approach is tour-based, ensuring spatial, temporal and modal consistency between trips made by the same person during the course of a day and within the same tour. A tour is a chain of trips made by the same person to conduct activities throughout the day. Individuals' daily travel and activities can be tracked in ABM, capturing car-sharing between individuals and answering questions regarding car-sharing operational strategies (e.g., evaluating car-sharing services or estimating the demand given one proposed car-sharing policies).

The properties of ABM present a challenge to transportation planning practitioners. ABM is a data-hungry approach that requires detailed input information about individuals instead of TAZs in trip-based model. In order to prepare the ABM travel demand input data, one may think of conducting a comprehensive travel survey that asks every person in a modeling region about their activity diary (key information should include the times, locations and types of activities). However, it sounds financially infeasible. This study delivers a methodology of preparing travel demand data for ABM, utilizing available data sources including the existing household travel surveys, well-established trip-based models, and open-source data. This study is particularly useful for transportation planners who develop and apply trip-based travel models in their jurisdiction since the input data used in this study are commonly available for transportation planning researchers and practitioners. The methodology offers insights in preparing the data for ABM that help simulate and understand the individuals' travel patterns, and evaluate the transportation policies/strategies under the environment of shared economy and new travel modes, e.g., shared connected and autonomous vehicles. This study presents an example of using data that are easily accessible by the public. Other data sources, such as transportation's Big Data platforms like Streetlights (www.streetlightdata.com) and AirSage (www.airsage.com), which may be private but provide great travel data can also be used in preparing activity-based model input data.

METHODOLOGICAL FRAMEWORK

This study proposes a methodology of preparing the input data for ABM. The input data may be summarized as "4Ws" for each traveler's choices, as shown in Figure 1. The core of the framework consists of a series of algorithms that output "4Ws" by inputting the aggregated data at zone level. The framework starts with generating synthetic population and households based on land use and socioeconomic data. The output at this step provides information of "Who," defining travelers individually based on age, gender, employment, car ownership, and household characteristics. The next step is locating of households and employments, the information of "Where", taking advantage of the Open Street Map (OSM) data that contains the layout of buildings in a region. These locations are designated areas for conducting activities. The following two steps together output the information of "What," a chain of daily activities that form a travel tour. Zone level travel demand is converted to person-level travel demand by chaining the trips between zones and assigning locations for trips' origins and destinations (that are also the activity locations). The last step is to prepare the information of "When," a tentative schedule for traveling or performing activities.

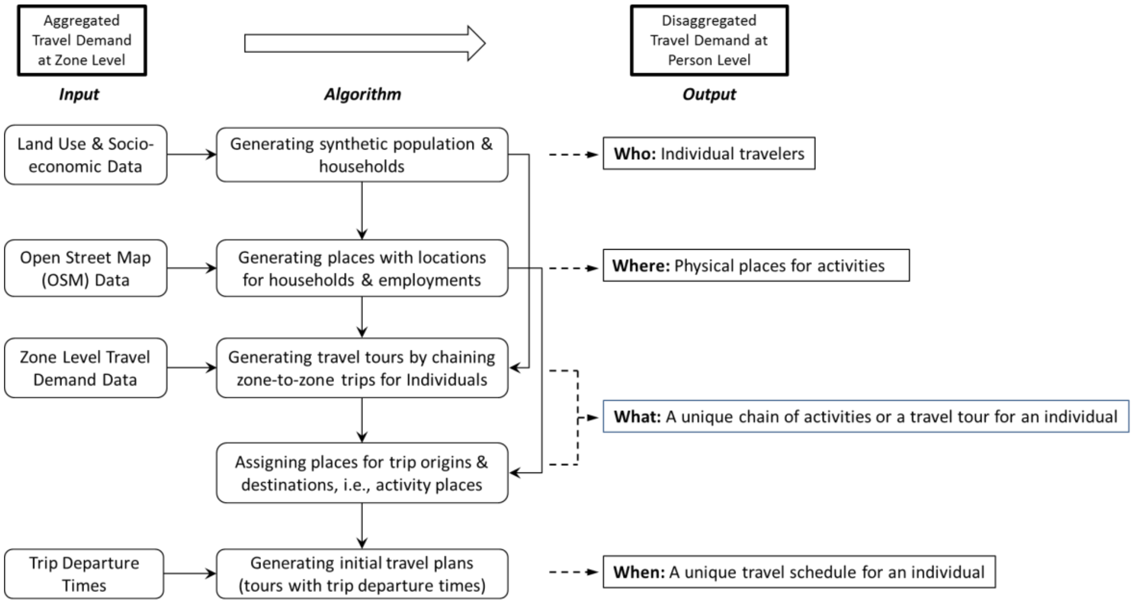


FIGURE 1 Methodological framework of outputting personal level travel demand at person level from zone-level travel demand

DATA PREPARATION

Three data types were suggested for synthesizing a region's population and generating their travel tours or itineraries: 1) travel demand data from trip-based or four-step travel models, 2) model equations' parameter values, and 3) open-source map data. Table 1 lists the specific data sets used here, for method illustration.

TABLE 1 Data Sources for Preparing AMB Inputs

Source	Data	Key variables	Data source
Travel demand data	TAZ land use data and its shape file	- Population - Household size - Employment - Car ownership - Income level	Regional travel demand models: https://www.campotexas.org/
	Trip distribution table (i.e., OD matrix)	- Trip purpose - Number of trips	
Parameter data	Population age distribution	- Age - Percent	Census: https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
	Trip departure time distribution	- Trip purpose - Time of day - Percent	NCHRP Report 716: http://www.trb.org/Publications/Blurbs/167055.aspx
	Trip patterns	- Number of trips in a daily tour - Percent	NHTS datasets: http://nhts.ornl.gov/download.shtml
Map data	OpenStreetMap data	- Road network - Building/housing footprint	OpenStreetMap data: http://www.openstreetmap.org/

Travel Demand Data

Travel demand data were extracted from Austin's (CAMPO's) regional travel demand model. The region covers over 5,000 square miles, including Bastrop, Burnet, Caldwell, Hays, Travis, and Williamson Counties in Texas. CAMPO's 2010 Planning Model is a largely traditional four-step macroscopic travel demand model (*12*). This study used data from the 2020 scenario.

Two model data sets were used: TAZ land use data (for jobs and population counts, by type, across zones) and trip distribution data (between zones for each trip purpose). The TAZ land use data is important for synthesizing population. The trip distribution data is also called the origin-destination matrix (OD matrix), offering a picture of possible trips between/within TAZs. Six types of trip purposes were considered in the tour generation process: Home-based work (HBW), Home-based school (HBSc), Home-based retail (HBR), Home-based other (HBO), Non-home-based work (NHBW), and Non-home-based other (NHBO) trips.

Parameter Data

The age distribution parameter is used to control population age structure in a model region. The trip pattern parameter tells the tour generation process how many trips a person may make in a day. According to the U.S.'s 2009 National Household Travel Survey (NHTS), the average number of daily trips for Texans is 3.76 trips (or 3.78 trips-per-day nationally). Figure 2 (a) presents the distribution of daily trips per person, with 15.7% of Texans making zero trips on any given day, and 22.6% making exactly two trips in one day.

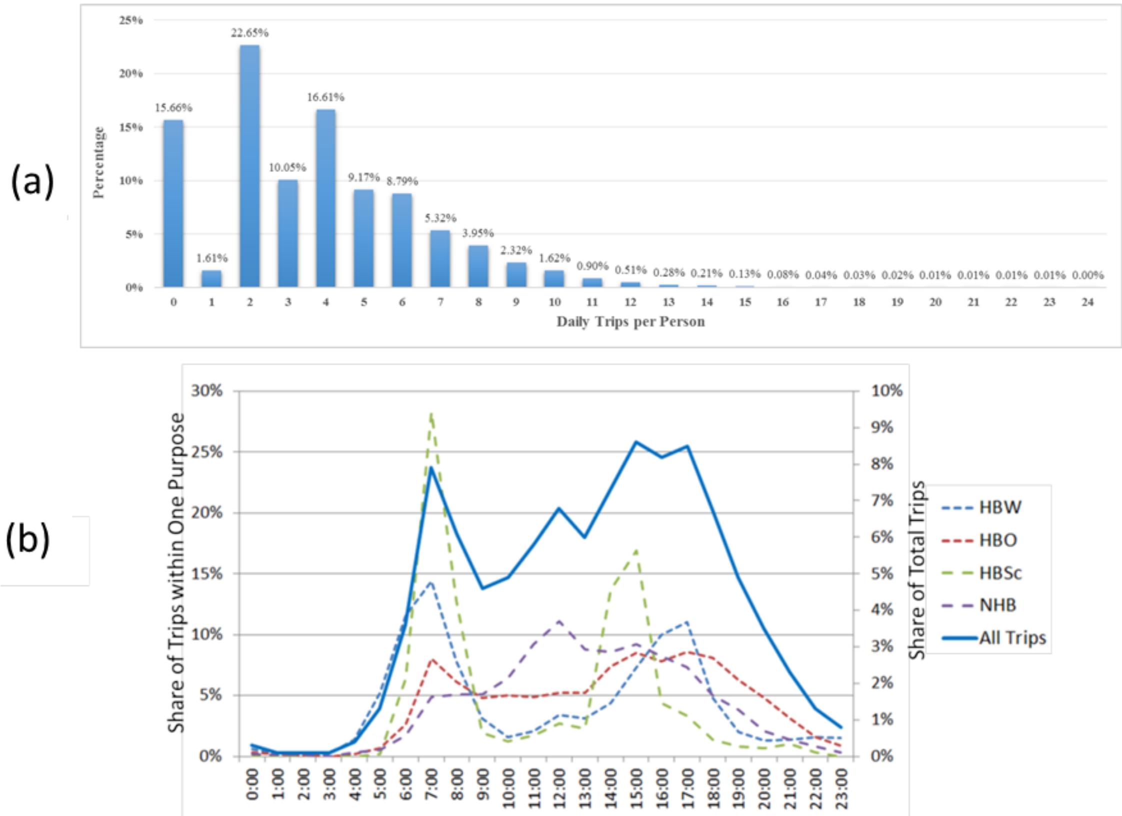


FIGURE 2 Parameter data: (a) trip count in daily travel tours and (b) time-of-day distributions by trip purpose

The trip departure time tells when a trip may start. This parameter is important for observing the time-of-day variation of travel patterns. NCHRP Report 716 provides a list of parameters for trip departure times by trip purpose (*10*). The parameters show percentages of trips starting within one hour in a day, as shown in Figure 2(b).

Map Data

Great spatial details come from the map data. This study extracted the road network data and the house/building footprints from an open source map data – OpenStreetMap (www.openstreetmap.org). The map data provide not only how people get place to place, i.e., routes, but also the information about the locations for homes and employments.

PROGRAMMING

This section presents key programming algorithms for preparing disaggregated travel data. The algorithm codes are available from authors, and will be released as open source whenever ready.

Generating Synthetic Population and Household

The publicly available survey data (e.g., household travel surveys and Census data) offer insights in the socio-economy or land use at an aggregated level, for instance, tracts for census data, and Traffic Analysis Zones in travel surveys. The socio-economy or land use information is closely related to the generation and attraction of travel trips. Four-step regional travel demand modeling often starts with the socio-economic data as the inputs in the models. A complete travel model is supposed to have a database containing TAZs' socio-economic or land use information which may be a synthesis from various data sources including household travel surveys and census data. For modeling of future years, the projected socio-economic data is also provided in travel models by coupling with experts' opinions, general population growth rates and regional land use plans . This study used the CAMPO Model's forecasted socio-economic data for year 2020 to generate information for individual travelers in the model region, including their personal information (age, gender, employment, car ownership) and their household's information (household size and income level). Note, car ownership is a household property in input data, and assigned to specific household members in this program according a person's age and employment status.

Allocating Locations for Households and Employments

Activities are expected to happen in either homes or employment locations, which are also the trip origins and destinations. Daily activities include home, work, school, shopping and other activities. Home activities are performed at homes, and work-related activities occur at employment locations. School, shopping and other activities (e.g., eating, exercise, etc.) are also likely to occur at certain locations that are associated with employments (e.g., teacher, salesman, chief or servant, or physical couch).

This study extracted the polygon information from the OSM's building layer. A centroid (longitude and latitude) was obtained as the physical location of each building polygon, and the site area of each polygon was also calculated. Besides the physical location and site area, other information (such as the floor area) is also valuable, but not available from data used in this study. According to the site area, the physical locations obtained from OSM data are grouped into: 1) small size, < 5000 sq ft, 2) medium size, 5000 ~ 10000 sq ft, and 3) large size, > 10000 sq ft. Small sized buildings are assumed to be single-family homes, medium sized ones are apartments, and large sized buildings are places of employment. The household income is regarded as a key factor in the building/location allocation. Single-family homes are likely to be medium and high income households; and apartments are for lower to medium household incomes levels. All assumptions are not strict but just represent the most likely situations. Randomness is involved in algorithms.

Since this study uses the future year's travel demand model data and the OSM data contains the information about existing houses and buildings, it is fairly reasonable to generate new employment locations to handle additional employments (which represent the future land use development). The type of employments is also considered, which tells what a location is for in terms of activity type or travel trip purpose, including basic work, shopping/retail, education and other. A location may be only for basic work, such as office buildings. A location can also have multiple functions, such as schools where faculty work and students attend for educational activities.

In general, a higher income level household is associated with a bigger house. The households with lower level incomes are likely to be limited to apartment buildings. Unlike houses, the apartment buildings can house multiple households. Assuming most apartment buildings are 2 ~ 3 floors and each unit is about 1000 sq ft, the apartment buildings are split into multiple pieces by dividing the site area over 500; and these units from the same building share the identical location. Projected households are included in the future year's demand model; therefore, additional houses may be generated in some TAZs.

Chaining Trips between TAZs

Trips made by one traveler in one day form a trip chain, called a tour. Based on the OD matrices from four-step travel models, this study develops algorithms to chain the trips to generate tours for individual travelers. The tour pattern, i.e., number of trips in a tour, is defined according to 2009 National Household Travel Survey (NHTS), as shown in Figure 2. Zero-trip makers are likely to be either too young or too old to make a travel on a daily basis, i.e., younger than 5 years old or older than 85 years old. In addition, the individuals who do not have a car and are unemployed have an increased likelihood of making zero trips daily than those who have a car and a job. The number of trips for travelers who own a vehicle is generally more than that for those who do not have a vehicle.

The OD matrices specify trips by purposes, including home-based work (HBW), home-based school (HBSc), home-based retail (HBR), home-based other (HBO), non-home-based work (NHBW), and non-home-based other (NHBO). The trip purposes tell the origin and destination characteristics. For example, a HBW trip links a home and an employment location; an HBSc trip connects a home and an educational facility; and a NHBW trip starts from non-home and non-work place to a work place. In general, a traveler is assumed to have one home, one place for work, one place for educational activities, and may have multiple places for other activities.

Allocating Locations for Trip Origins and Destinations

The previous step chains trips between TAZs, while the locations of trips' origin and destination are not specified yet. The activity-based models require specific locations for trips' origins and destinations. Based on the output from the previous step, this step allocates locations for trips within particular TAZs. The allocation process is according to the trip-associated activity type. For example, a HBW trip connects one home and one work location, requiring a search for the associated traveler's home location (which is pre-specified in the synthetic households after home location assignment) and his or her work place from all possible employment locations within a specific TAZ. All work-related trips for one person are linked with the same work place. For other types of trips including HBR or HBO, a location is needed for shopping or other activities within the target TAZ. Different trips for shopping or other activities may be connected with different places as long as the location's type is correct and it is in the corresponding TAZ.

With the exception of home locations, the number of trips received by a facility or building is proportional to the number of employments generated in an early step (it can also be obtained from land use surveys if available). For example, a shopping facility has ten salesmen; therefore it may receive more (not exactly twice) trips than one facility which has five salespersons.

The travel mode is also determined in this process, according to the vehicle ownership. Assumptions include: if a person owns a vehicle, he or she drives; if a person does not own a vehicle but his or her family owns at least one, he or she may carpool; and if the entire household owns no vehicle, he or she has to choose other modes.

Generating Initial Travel Plans

This step generates information about when a traveler may start a trip. NCHRP Report 716 provides the patterns of trip departure times, showing in general when a trip may begin. In addition, the activity durations and trip durations are also important, as they are major time consumers. Time may be regarded as the resource of making travel plans; and 24 hours is the total resource for an individual to make his or her travel plan in one day. Typical activity durations and start times are assumed in this study. For example, most work activities may start around 8 AM in the morning and last about 8 hours. The activity durations are also dependent upon the number of activities planned by a traveler. The more activities planned for one day, the shorter the average activity duration is. The trip durations are determined by the trip distance and average trip speed. In this program, the bee line distance was quickly calculated according to a trip's origin and destination, and 35 mph is assumed to be the average trip speed. The average speed may also be obtained from the skims in 4-step travel models. Initial travel plans only tell when a traveler is likely to make a trip. Travelers may modify their travel plans (like changing trip departure times, or re-scheduling the activities) in order to avoid the excessive time spent on roads, reaching the user-equilibrium situation, which is discussed in the section of Cast Study in this paper.

PROGRAM OUTPUTS

Synthetic Population

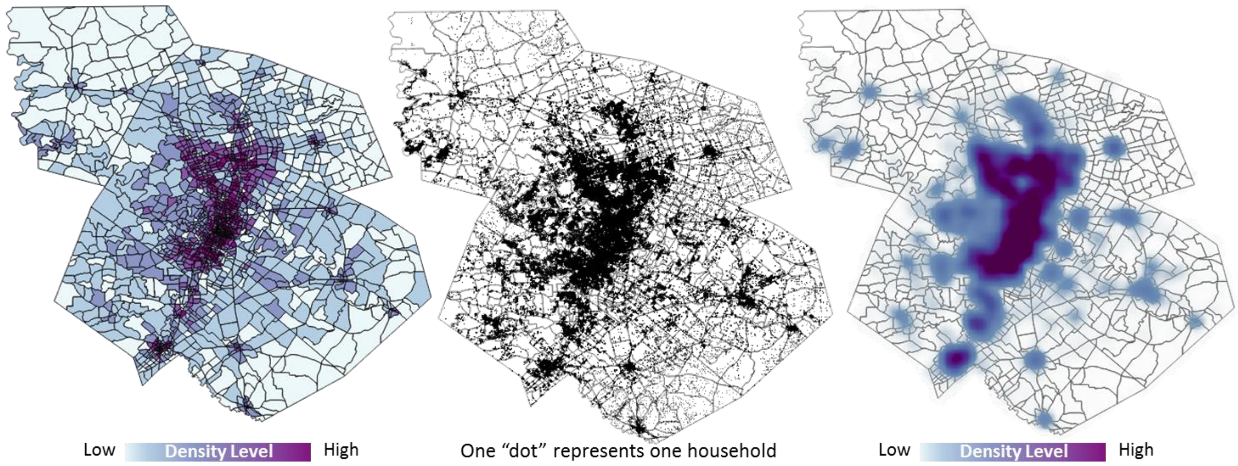
The program was designed to use the surveyed data and projected demographics used in travel models (summarized at TAZ level) to generate a synthetic population, though the randomness is included in the generation process. The data outputted from the program is supposed to match the statistics of input data at a large extent. Minor differences (<1%) are found between the outputted synthetic population and the inputs (socio-demographic data of CAMPO travel model). The differences are mainly due to the randomness and number rounding. Using the CAMPO's 2020 travel model inputs, the program generated a synthetic population of 2,325,116 individuals from 895,082 households in the model region. For each individual, generated information includes age, gender, employment and car ownership. In addition, individuals are also linked with their household characteristics including household size, household income level, number of employed members, number of vehicles and household locations (longitude and latitude). All these factors are important in activity-based travel modeling process. Figure 3 (a) and (b) presents the example data of synthetic population at household and person level. From the spatial perspective, the synthetic population is also expected to mirror the aggregated input data. Figure 3 also presents (c) the input data of population and households aggregated at TAZ level from the CAMPO's 2020 Travel Model, (d) the spatial distribution of synthetic households, and (e) the density map of synthetic population.

(a) Example household data

hh_id	hh_tazid	hh_size	hh_income	hh_worker	hh_auto	longitude	latitude
1	1	6	5	4	3	-97.876433	30.5517077
2	1	4	5	2	2	-97.87921	30.5536900
3	1	2	5	2	2	-97.874095	30.5503466
4	1	3	5	2	2	-97.878559	30.5530491
5	1	1	3	1	1	-97.86365	30.5586560
6	1	5	5	2	2	-97.877618	30.5526646
7	1	2	5	0	2	-97.87601	30.5516081
8	1	4	5	2	2	-97.877382	30.5532665
9	1	4	5	2	2	-97.874668	30.5500271
10	1	4	5	2	2	-97.864282	30.5478816

(b) Example person data

per_household_id	per_id	per_age	per_gender	per_working	per_car
1	1	33	0	1	1
1	2	47	1	1	1
1	3	26	0	1	1
1	4	26	1	1	0
1	5	25	0	0	0
1	6	5	0	0	0
2	1	36	0	1	1
2	2	42	1	1	1
2	3	7	0	0	0
2	4	62	0	0	0
3	1	29	1	1	1
3	2	31	1	1	1



(c) Population density at TAZ level (CAMPO 2020 Travel Model) **(d) Locations of synthetic households (sample size = 895,082)** **(e) Synthetic population density based on household locations + households sizes**

FIGURE 3 Synthetic population and households

Activities and Synthetic Locations

Synthetic locations are needed to house these activities in the model region. The household locations are for home activities. For the other types of activities, the program generated employment-based locations to house them, though people may not go there for work but for other purposes such as shopping or taking classes. Figure 4 presents the example data of generated facilities for activities and Figure 5 also shows the locations for four types of activities. Compared with the household locations (as shown in Figure 4), the school and shopping locations are more likely to concentrate to the urban centers; locations for other activities are close to how households are spatially distributed in space.

Example Data of Synthetic Facilities for Employments and Activities

facility_id	Total_ employment	basic_work	shop	other	school	TAZ_id	longitude	latitude
1	94	0	18	6	70	1	-97.86418	30.55943
2	65	0	14	2	49	1	-97.865997	30.55937
3	34	0	26	0	8	1	-97.864968	30.56016
4	11	0	8	0	3	1	-97.86087	30.55812
5	14	0	11	0	3	1	-97.858756	30.54917
6	14	0	11	1	2	1	-97.863279	30.55795
7	20	0	20	0	0	1	-97.859803	30.55556
8	29	0	29	0	0	1	-97.863641	30.54751
9	19	0	19	0	0	1	-97.861061	30.55083

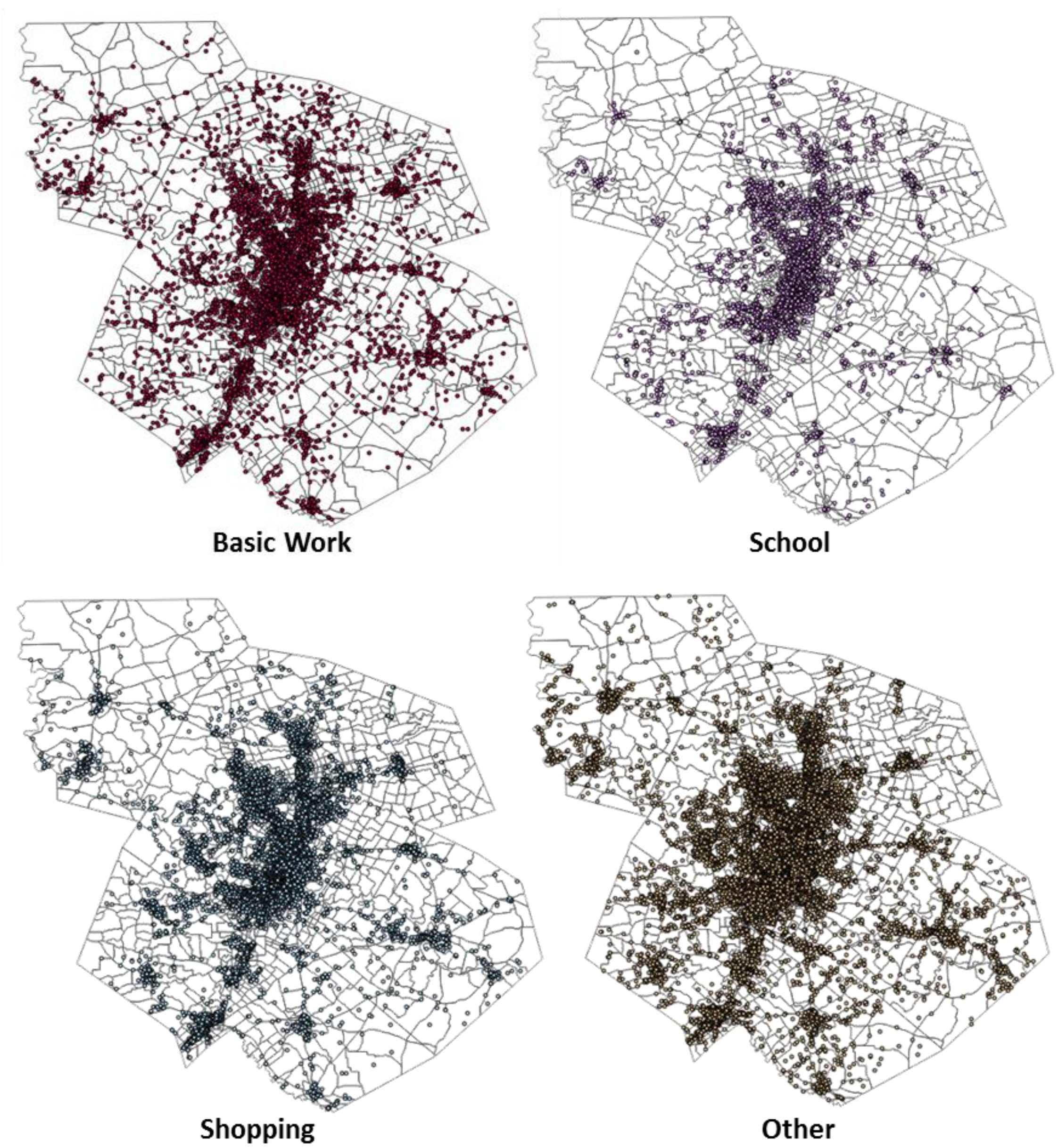


FIGURE 4 Example data of synthetic facilities and spatial distributions of facilities for different types of activities (except home activity)

Trip Chains

The core procedures of tour generation involved chaining the trips between TAZs (estimated in CAMPO's model) to form a tour for an individual, according to this traveler's demographics and NHTS's survey about the daily tour-making patterns (i.e., the number of trips made by a person, as shown in Figure 2). The program generated in total 1.96M tours that chain 8.7M trips for 1.96M individuals who actually travel on a daily basis (which leaves 0.36M persons who do not travel during 24 hours and are assumed staying at home for the whole day). The output resulted in about 3.9 trips per traveler in model region. Figure 5 presents the example data of synthetic trip chains, and two example tours in space: a four-trip tour with HBO ⇨ NHBO ⇨ NHBO ⇨ HBO trips, and a five-trip tour with HBW ⇨ NHBO ⇨ NHBW ⇨ NHBO ⇨ HBR trips.

Example Data of Synthetic Tours or Trip Chains

PerId	TripId	activityType	OriginFacilityTAZ	DestinFacilityTAZ	OriginFacilityID	DestinFacilityID
1	0	other	1	1866	1000001	56690
1	1	school	1866	1057	56690	35852
1	2	shop	1057	1595	35852	47825
1	3	home	1595	1	47825	1000001
2	0	work	1	162	1000002	6880
2	1	other	162	929	6880	34081
2	2	work	929	162	34081	6880
2	3	shop	162	118	6880	3937
2	4	home	118	1	3937	1000002

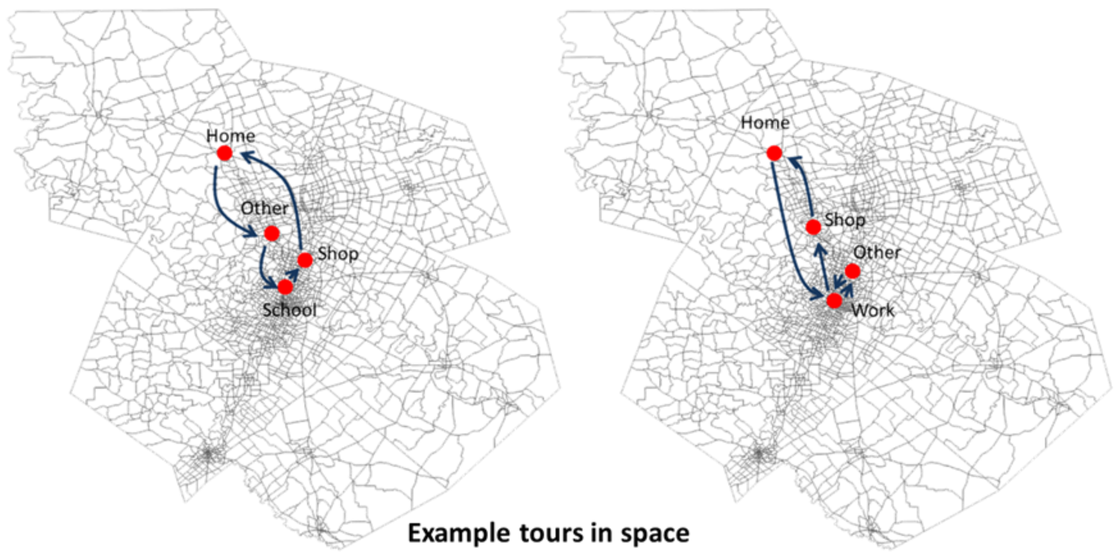


FIGURE 5 Example of Synthetic Tours or Trip Chains

Travel Plans

The outputted travel plan contains information about the person's age, employment status, and a chain of activities with a tentative schedule. Figure 6 shows two example travel plans. The travel plan is the core input of ABM. The travel plan reveals a typical schedule for travel and activities. During the modeling process, the travel plan may be modified given constraints of one-day time and space in roadway network. Late arrival, early departure, or cancelling an activity will cause loss of utility, while being stuck in traffic will also negate the production of values. More details are presented in the cast study in this paper.

```xml
<person id="1" age="33" employed="yes">
  <plan selected="yes">
    <act type="home" facility="1000001" x="-97.87921008" y="30.55369003" end_time="08:56:59"/>
    <leg mode="car" dep_time="08:56:59"/>
    <act type="other" facility="56690" x="-97.86081666" y="30.53089811" end_time="09:33:00"/>
    <leg mode="car" dep_time="09:33:00"/>
    <act type="school" facility="35852" x="-97.77755189" y="30.43983589" end_time="15:12:00"/>
    <leg mode="car" dep_time="15:12:00"/>
    <act type="shop" facility="47825" x="-97.79296572" y="30.5232782" end_time="17:53:59"/>
    <leg mode="car" dep_time="17:53:59"/>
    <act type="home" facility="1000001" x="-97.87921008" y="30.55369003"/>
  </plan>
</person>
<!--===================================================================-->
<person id="2" age="47" employed="yes">
  <plan selected="yes">
    <act type="home" facility="1000002" x="-97.87409538" y="30.55034662" end_time="08:11:59"/>
    <leg mode="car" dep_time="08:11:59"/>
    <act type="work" facility="6880" x="-97.7487424" y="30.34361178" end_time="12:05:00"/>
    <leg mode="car" dep_time="12:05:00"/>
    <act type="other" facility="34081" x="-97.80691386" y="30.47037602" end_time="12:45:02"/>
    <leg mode="car" dep_time="12:45:02"/>
    <act type="work" facility="6880" x="-97.7487424" y="30.34361178" end_time="16:32:00"/>
    <leg mode="car" dep_time="16:32:00"/>
    <act type="shop" facility="3937" x="-97.69170071" y="30.45757895" end_time="19:07:24"/>
    <leg mode="car" dep_time="19:07:24"/>
    <act type="home" facility="1000002" x="-97.87409538" y="30.55034662"/>
  </plan>
</person>
```


FIGURE 6 Example travel plans

Spatial Details

This program generates specific physical locations for individuals to perform activities and these locations are the origins and destinations of trips (rather than TAZ centroids in 4-step travel models). These locations are scattered in TAZs, as shown in Figure 7 (a). There are two types of scatter patterns. One type has quite clear patterns, shown in Figure 7 (b), along the road links, as these locations are known places for households and employments according to used open-source data. The other type seems to be irregular patterns, shown in Figure 7 (c). These locations were generated according to the road link/node locations and the number of households and employments in a TAZ. The irregularity is due to the limitations in open source data (e.g., incomplete records) and the need for understanding future travel patterns.

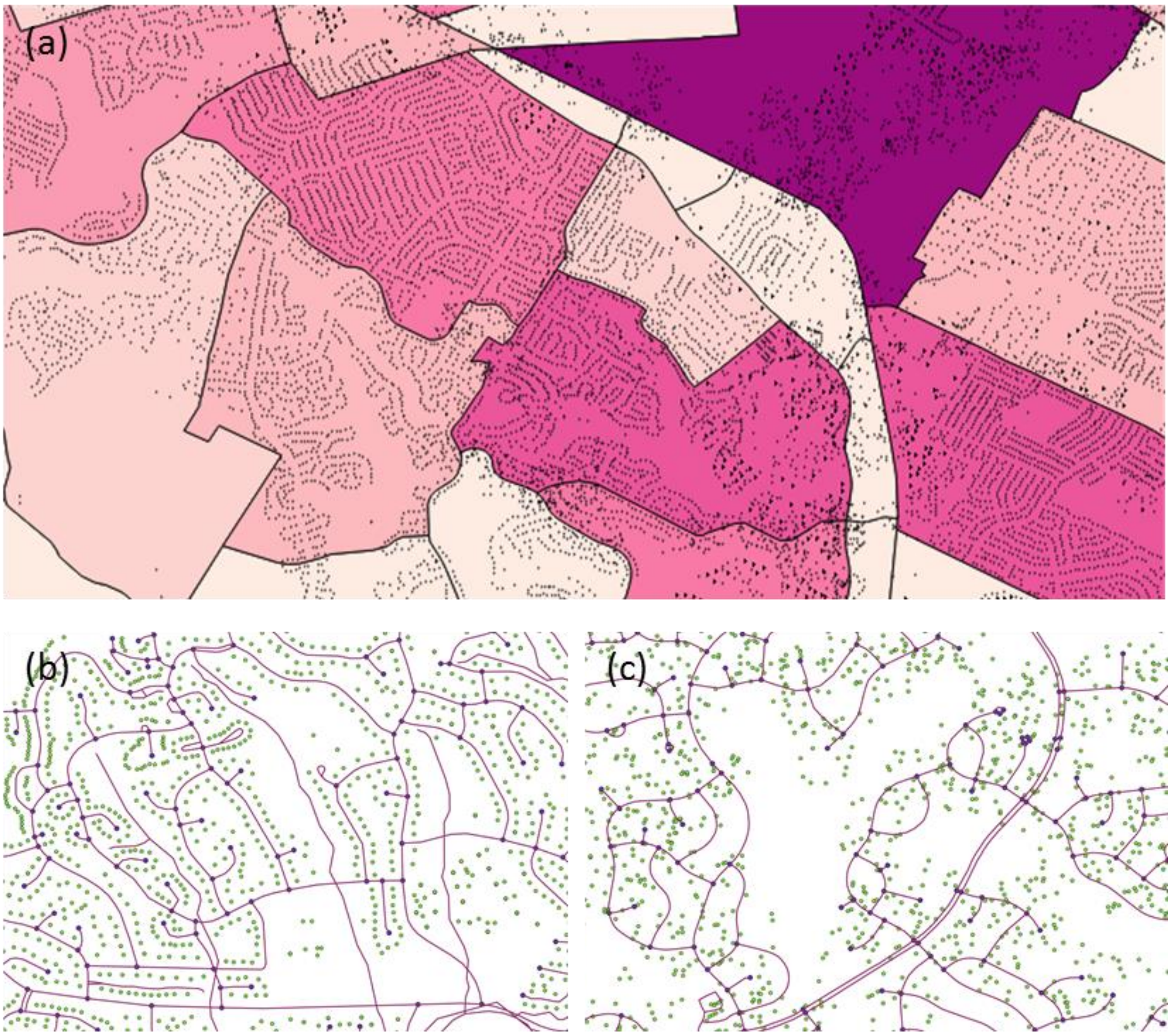


FIGURE 7 Spatial Details for Activity Locations

CASE STUDY

This section briefly presents a case study, to construct an activity-based model using the synthetic activity data generated by the program in this study. The model was built on the platform MATSim, an open-source agent-based simulation tool for large-scale activity-based microsimulations. MATSim is based on the co-evolutionary principle. Every agent (i.e., traveler) repeatedly optimizes his or her travel solutions based on their initial travel plans while competing for limited space-time slots with all other agents in the transportation network (citation, MATSim book). A MATSim run starts with initial travel plans, i.e., the chains of trips or activities a person plan to make on a daily basis. During iterations, the initial travel plans are then optimized individually. Every agent possesses a memory containing a number of day travel solutions, where each solution is composed of a daily trip chain and an associated score. The MATSim scoring function is based on the econometric utility of time. Unlike studies or programs where the utility is calculated for travel only (the mode or route choice), the utility function in MATSim accounts for both the travel and the activities an agent performs one a daily basis:

$$U = \sum_{i=1}^{q} U_{travel,i} t_{travel,i} + \sum_{j=1}^{q+1} U_{activity,j} t_{activity,j}$$

where U = Total utility of a travel solution composed of a daily trip chain; $U_{travel,i}$ = Utility of travel for i^{th} trip in a day; i = 1, 2, 3, …, q trips; $t_{travel,i}$ = Travel time for i^{th} trip; $U_{activity,j}$ = Utility of performing the j^{th} activity in a day; j = 1, 2, 3, …, q+1 activities; and $t_{travel,i}$ = Duration of

j^{th} activity. Moreover, $\sum_{i=1}^{q} t_{travel,i} + \sum_{j=1}^{q+1} t_{activity,j}$ = 24 hours.

Monetary payments (e.g., tolls and fares) and the value of travel time (VOTT) are included in the $U_{travel,i}$. The utility of performing an activity, $U_{activity,j}$ is related to value of activity time (e.g., hourly wage). The travel utility is generally negative while the activity utility is positive. The travel solution optimization is to maximize the total utility of a chain of trips an agent may take to perform his or her planned activities. More details about the MATSim scoring function are available in the MATSim Book. The MATSim's iterative process is to improve the utility by re-planning the travel trips i.e., modifying time choice, mode choice, or destination choice, and finally to reach dynamic user equilibrium (DUE). After reaching DUE, the MATsim outputs a most executable travel solution for each agent. As MATSim simulations are to mimic the process of travelers looking for the best travel solutions for their daily activities in real-world, the MATSim simulation results have been revealed to match the real-world travel patterns very well (*13, 14*).

The outputs of MATSim simulations include an optimal travel plan for each agent. Through a closer look at the plan, researchers or modelers can track each agent in the network. In other words, at any time of a day, where an agent is and what this agent is doing can be presented. The animation of simulated activities and travel trips is available at https://www.youtube.com/watch?v=kqHI3xc3nC0.

LIMITATIONS

The accuracy of synthetic data generated in this study is heavily dependent upon the accuracy of inputs including the travel demand data, parameter data and map data. In addition, the program presented in this study generates synthetic activity and travel data according to limited data sources with a number of assumptions. The validity of these assumptions remains unknown, and surveys are needed to validate these assumptions in the future. If using a desktop level computer or laptop, the generation of synthetic data using the current program may be a computational burden for large-scale travel model regions (population > 1 M), due to the massive searching cases (e.g., assigning a location for an activity), and matching requirements (the disaggregated synthetic data are required to match the aggregated data at TAZ level from various prospects, e.g., the total population, household, vehicle ownership, employments, etc.). The use of workstation level computers may facilitate the run of the program.

SUMMARY

New travel patterns, e.g., car-sharing behavior, present an opportunity and also a challenge for transportation planners and researchers to explore the disaggregated travel demand at person level, in addition to the aggregated demand at zone level which has been well modeled using trip-based approach. This opportunity allows planners and researchers to confront the new questions regarding the new travel patterns and emerging transportation modes (e.g., autonomous vehicles), while the challenge may hold them back due to difficulty of obtaining disaggregated input data for advanced travel demand modeling at person level. This study offers a methodological framework to prepare input data for activity-based modeling (ABM), one of the most promising modeling approaches for person-level travel demand. The core of this framework is composed of a series of algorithms that take advantage of publicly available data sources (that are often aggregated at zone level) and produce the disaggregated data at person-level for ABM. The data sources used in this study include land use and socio-economic data, household travel surveys, Open Street Maps and regional trip-based travel models.

This study summarized ABM data into "4Ws" regarding an individual's daily travel: who this person is, where this person lives and works, what daily activities this person does, and when this person plans to perform activities. A program, consisting of a series of algorithms, was designed to generate the data that provide information about the "4Ws". First, the program generated synthetic population based on the zone-level land use and socioeconomic data. Every individual in the modeling region is included in synthetic population; generated attributes include age, gender, employment, car ownership, and household characteristics. Second, places for households and employments were generated to answer where a person lives and works. Open Street Map (OSM) data provide the information about possible locations/places for households and employments. Then the program converted the zone-level travel demand (i.e., trips between zones) to person-level demand (i.e., a unique chain of activities, forming a travel tour which connects specific physical locations instead of zone centroids in trip-based models). The program gave answers to what activities a person does. Last but not least, a schedule for traveling or performing activities was generated by the program to tentatively answer when a person plans to perform activities. Example outputs are showed in this paper. The outputs present great temporal and spatial details about the individuals' travel patterns.

This study offers both methodological and practical contributions. The framework proposed in this study offers theoretical insights about the "4Ws" as the input components for constructing activity-based travel models and from what public data sources can be used to prepare the "4W" information. This study delivers a practical tool that can help transportation planners and researchers to prepare the "4W" information for ABM. The tool is a computer-based program developed in R environment, composed of a series of algorithms that take advantages of the publicly available data sources and produce person-level information for ABM.

ACKNOWLEDGEMENTS

The authors thank the Texas Department of Transportation (TxDOT) for financially supporting this research (under project 0-6838), and the Capital Area Metropolitan Planning Organization (CAMPO) for sharing its regional travel model. The authors are grateful to MATSim's developers for their efforts in improving this unique and open-source toolkit (at http://www.matsim.org/).

REFERENCES

1. Cohen, Boyd, and Jan Kietzmann. 2014. Ride on! Mobility business models for the sharing economy. *Organization & Environment* 27.3: 279-296.

2. Martin, E., and S. Shaheen. 2016. Impacts of car2go on Vehicle Ownership, Modal Shift, Vehicle Miles Traveled, and Greenhouse Gas Emissions: An Analysis of Five North American Cities. Transportation Sustainability Research Center (TSRC), UC Berkeley. http://innovativemobility.org/wp-content/uploads/2016/07/Impactsofcar2go_FiveCities_2016.pdf. Accessed March 15, 2017.

3. Liu, J., K. Kockelman, P. Boesch and F. Ciari. 2017. Tracking a System of Shared Autonomous Vehicles across the Austin, Texas Network using Agent-Based Simulation. Forthcoming in *Transportation*. In: *Transportation Research Board 96th Annual Meeting*. http://www.caee.utexas.edu/prof/kockelman/public_html/TRB17SAVs_acrossAustin.pdf

4. Kockelman, K. M., and T. Li. 2016. Valuing the Safety Benefits of Connected and Automated Vehicle Technologies. In: *Transportation Research Board 95th Annual Meeting*. https://www.caee.utexas.edu/prof/kockelman/public_html/TRB16CAVSafety.pdf

5. Fagnant, D. J., and K. Kockelman. 2015. Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations. *Transportation Research Part A: Policy and Practice* 77: 167-181.

6. Fagnant, D. J., and K. Kockelman. 2016. Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in Austin, Texas. *Transportation* (2016): 1-16.

7. Chen, D., and K. M. Kockelman. 2016a. Management of a Shared Autonomous Electric Vehicle Fleet: Implications of Pricing Schemes. *Transportation Research Record: Journal of the Transportation Research Board* 2572: 37-46.

8. Reiter, M. S., and K. M. Kockelman. 2017. Emissions and Exposure Costs of Electric Versus Conventional Vehicles: A Case Study for Texas. *International Journal of Sustainable Transportation* 11.7: 486-492.

9. Bansal, P., K. M. Kockelman, and Y. Wang. 2015. Hybrid Electric Vehicle Ownership and Fuel Economy Across Texas: An Application of Spatial Models. *Transportation Research Record* No. 2495: 53-64.
10. NCHRP. Travel Demand Forecasting: Parameters and Techniques. VOL. 716. Transportation Research Board, 2012. https://www.nap.edu/catalog/14665/travel-demand-forecasting-parameters-and-techniques. Accessed April 12, 2016.
11. Castiglione, J., M. Bradley, and J. Gliebe. 2015. *Activity-based travel demand models: a primer*. No. SHRP 2 Report S2-C46-RR-1. https://www.nap.edu/catalog/22357/activity-based-travel-demand-models-a-primer. Accessed March 12, 2017.
12. CAMPO. 2015. CAMPO 2010 Planning Model Guide, Capital Area Metropolitan Planning Organization (CAMPO), Austin, TX. http://www.campotexas.org/plans-programs/. Accessed April 12, 2016.
13. Ciari, F., M. Balac, and K. W. Axhausen. 2016. Modeling carsharing with the agent-based simulation MATSim: state of the art, applications and future developments. In: *Transportation Research Board 95th Annual Meeting*. http://e-citations.ethbib.ethz.ch/view/pub:161004.
14. Bösch, P. M., F. Ciari, and K. W. Axhausen. 2016. Required Autonomous Vehicle Fleet Sizes to Serve Different Levels of Demand. In: *Transportation Research Board 95th Annual Meeting*, https://www.ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/vpl/reports/ab1089.pdf