Total Design Data Needs for the New Generation Large Scale Activity Microsimulation Models

Konstadinos G. Goulias

Goulias@geog.ucsb.edu, University of California Santa Barbara

Ram M. Pendyala

Ram.Pendyala@asu.edu, Arizona State University,

and

Chandra R. Bhat

bhat@mail.utexas.edu University of Texas at Austin

Abstract

In this paper we describe a total design data collection method (expanding the definition of the usual "total design" terminology used in typical household travel surveys) to emphasize the need to describe individual and group behaviors embedded within their spatial, temporal, and social contexts. We first offer an overview of recently developed modeling and simulation applications predominantly in North America followed by a summary of the data needs in typical modeling and simulation modules for statewide and regional travel demand forecasting. We then proceed to describe an ideal data collection scheme with core and satellite survey components that can inform current and future model building. Mention is also made to the currently implemented California Household Travel Survey that brings together multiple agencies, modeling goals, and data collection component surveys.

Introduction and Background

Our recent experience testing a new generation of models for large scale policy analysis in a substantially heterogeneous region in the US motivates this paper that enumerates and explains the type of data needed for travel demand modeling and simulation. The fundamental motivation for our travel model system emerged from the policies examined, which include the following:

1. Land use changes to increase density and diversity of development (e.g., the California Senate Bill 375 that requires coordination of policies on land use with transportation and the creation of sustainability plans – see Appendix for an excerpt of the requirements);
2. Impacts of changing demographics due to aging of population, in- and out-migration, and changes in fertility, mortality, as well as changes in educational attainment, family formation and dissolution, and changes in employment achievement and prospects;
3. Impacts of market introduction and penetration of new types of vehicles including flexible fuel, electric, and hybrid vehicles, as well as the resulting changes in the composition and use of household vehicle fleets;
4. New development and the addition of roadway and transit infrastructure components;
5. Pricing of services (including parking and toll roads) and access restrictions to part of a city; and
6. Spatio-temporal changes through policy interventions in the housing, school, and employment patterns.

A modular simulation model system that we have developed to support the above policy needs, includes the following components: (1) synthetic population generation to recreate the resident population of a study area; (2) household evolution and spatial location choice models (e.g., for residence, work, and school) to represent socio-demographic changes at the most elementary level of a person and a household, and concomitant changes in the ability to participate in activities; (3) highway and transit accessibility to capture changes in the activity opportunities and the transportation systems that serve them; (4) activity scheduling and daily simulation to represent individual and household activity and travel; (5) vehicle utilization and allocation within the households; (6) interfaces with other travel demand models, including network

assignment; and (7) energy consumption and emissions estimation. These components are strung together either in a sequential manner or with important "feedbacks" to ensure coherency in the overall simulation. It should be noted that these components are also found in other modeling and simulation applications (Donnelly et al., 2010). In addition, a variety of data exchange and model interfaces are also designed to convert data from one form to another, and provide databases and maps for verification, validation, and visualization of policy scenarios.

The core data used by many of these components are from household activity-travel surveys that collect information on: (1) household-level characteristics (e.g., household location, household structure, life cycle stage, income, vehicle ownership, bicycle ownership, housing characteristics, and house ownership); (2) individual-level characteristics (e.g., age, gender, race, ethnicity, education, student status and school location, employment status, employment location, work hours, and the ownership of driving license); and (3) information of the travel undertaken by individuals, which includes how, why, when, and where they traveled. Time use and activity surveys (and their variants of place-based surveys) may also collect additional information on the activities participated by individuals such as timing, and duration of in-home, at work, and at other place activities. Many household travel surveys also include information about the fleet of household vehicles owned, including the make, model, year of manufacture, and additional information about each vehicle and related transactions.

Surrounding these core data elements are land use and infrastructure data that include information on the spatial residential characteristics of households, employment locations, school and activity opportunity locations (often aggregated at the level of traffic analysis zones), and transportation network data that includes highway network (roadway functional class, distance, direction, number of lanes, hourly capacity, posted speed limit, and so forth) and transit network data (routes followed by buses and trains, the frequency of service, travel speeds, distance, and travel time among nodes in the network). These data elements are used by major public agencies to develop and test simulation models. However, such data and information are not always sufficient to estimate models that assess policies at the level of individual decision makers. For this reason, supplemental data are obtained from secondary sources (e.g., the US census and other ad-hoc social and demographic surveys) or from specifically designed surveys to answer

policy questions (e.g., stated preference surveys). The joint use of different surveys for the same model development task requires data adjustments and courageous assumptions to make the surveys coherent in time, space, and the social milieu.

We describe in this paper a total design data collection method (expanding the definition of the usual "total design" terminology used in typical household travel surveys) to emphasize the need to describe individual and group behaviors embedded within their spatial, temporal, and social contexts. We first offer a summary of the data needs in a typical modeling and simulation application for regional travel demand forecasting in the US (see also Bhat et al., 2011, Goulias et al., 2011, Pendyala et al., 2011, Vyas et al., 2011), and point out important data gaps in key descriptors of behavior as well as other external data we need for model development and verifications tasks. We then proceed to create a staged development approach with core and satellite surveys to sketch out an ideal data collection program that can inform current and future model building. We also describe briefly a new generation of data collection efforts that contain additional core questions and survey modules to inform the immediate next generation of activity-travel simulation models. Many of the ideas discussed in this paper are extracted by merging our experience with modeling and simulation reviews by developers of activity-based models in the US (see Bowman, 2009, Rossi et al., 2010, http://urbanmodel.asu.edu/intmod/reports.html, Ferdous et al., 2009, Europe (www.sustaincity.org), and less developed countries in the world (Yagi, S., and A. Mohammadian, 2010). There are important variants to these modeling and simulation approaches (some of these currently developed by the authors of this paper) that are not included here to make the presentation tractable and to focus on developments that have a stronger connection to applications and practice that we see emerging in regional travel models in the US.

Travel Demand Forecasting Model Design Framework

Figure 1 shows a schema of the cascading form of model components for a typical activity-based microsimulation travel demand forecasting system. Although the sketch is from a recently developed model for a mega-region, it is representative of many models developed in the US to

address "green" policies that aim at motivating people to adopt strategies that decrease fuel consumption and mobile source emissions. The set of blocks on the left side represents groups of models that are designed for the first year (baseline) of the simulation that usually corresponds to a baseline year used in a regional or statewide transportation plan. Each block of the figure represents a set of techniques and statistical models developed to replicate the resident population activity and activity-travel decision making in a region.

Figure 1 goes here.

In particular, the entire first set of models on the left side of Figure 1 recreates the resident population, and attributes a daily activity-travel schedule to each individual in the population. The individual activity-travel schedules are then translated into vehicular travel patterns and assigned to the network to compute energy consumption and emissions for a baseline year. The middle block evolves the region's economic and demographic landscape over time to a pre-specified time increment. This computerized evolution is done using land use simulation models, an evolutionary engine of households and persons, and data fusion of information from cities and subareas within cities to build scenarios of macro changes. The right hand side block of models is a repetition of the daily activity and travel patterns models but at the next and all subsequent years of the simulation. Key difference, however, is the synthetic population that does not need to be repeated in the same fashion as the baseline year but it is used to guide the demographic microsimulation of the middle block.

The model system above can represent the activity-travel impacts of a whole host of land-use and transportation policies. For instance, land use policies of increased density and land use mix can be reflected in shifts in spatial distribution of economic activity (middle block), location decisions, car ownership and use, and activity participation and destination choices (including decisions to participate in activities and to travel alone or with others) (right block). But before being able to do so, we need to estimate and validate the model components represented in Figure 1, which leads to the issue of data needs, as described below.

Population Synthesis

The process starts with synthetically generating/recreating the entire resident population person-

by-person and household-by-household (see http://urbanmodel.asu.edu/popgen.html). The input to this software and block of methods is the spatial organization of the simulated area in the form of zone-specific target univariate distributions of resident person and household characteristics, as provided by agencies that track the demographic characteristics of a region. As the intent of population synthesis is to recreate the population on a person-by-person and household-by-household basis, the target univariate distributions are used as the control totals for each spatial unit of analysis in an iterative algorithm that starts from a multivariate set of relationships (in essence a cross-tabulation) among the person and household variables used as seed information More specifically, there are two types of data needed (Pendyala et al., 2011): (a) one or more contingency tables (cross-classification) from microdata to capture the relationships among variables at the household and person levels (these data provide the seed information); and (b) univariate distributions that the synthetic population should satisfy to represent the resident population in each geographical subdivision, such as a block, a block group, a traffic analysis zone, or a tract (this data serves as the target).

Accessibility by Time of Day

To represent employment opportunities and the spatio-temporal distribution of activity participation opportunities and the desirability of activity locations, opportunity-based accessibility indicators are developed at fine spatial resolutions. In this way, we represent the ease (or difficulty) of reaching different types of industries (representing the opportunities for activity participation) from each geographical location within a pre-specified travel time (or generalized cost-time buffer). In a recent application called SimAGENT we used 10, 20, and 50 minutes of roadway travel buffers from each of the 203,000 micro-zones and computed the number of industry-specific employees that can be reached (Chen et al., 2011). To reflect the time-of-day variation in accessibility, different values are obtained for the morning peak period (6 to 9 AM), midday (9 AM to 3 PM), evening peak period (3 to 7 PM), and at night (7 PM to 6 AM) capturing not only the different roadway conditions, but also the patterns of opening and closing of businesses during the day. Data needed for this block include: (a) indicators of industry-specific presence (e.g., number of firms by industry and number of employees, number of firms by industry type and floor space, number of employees in a geographical area by industry type and associated activity); (b) travel time and travel cost among all pairs of origins

and destinations by time-of-day; and (c) availability of firms or groups of firms by time-of-day (or proxies of this availability).

Long Term Choices

In a schema such as the one in Figure 1, the residential location of a household is generally established by the population synthesis process for the base year (the left side of the figure). However, for subsequent model years and to evolve the population over time or for locating households in individual parcels or building units, a residential location/relocation model is also needed. Conditional on home locations, other long term decisions are then modeled in this "long-term choices" block. In particular, every student (whether full time or part time) needs a school location and every worker needs a work location. These locations often serve as anchors (outside home) around which people engage in discretionary activities and travel. School location choice models are often difficult to simulate. Many children study at neighborhood schools, but a substantial proportion also go to private or charter schools that are located farther away from home. Adults may choose to attend different universities and colleges in the metropolitan area. For this reason additional models are developed at the person level. For example, when we examine persons in college, a model is used to assign a college location that is a hierarchical function of accessibility. In addition to standard demographic characteristics and measures of accessibility, information about school enrollment and school quality would be useful for modeling school location choice. It is generally possible to obtain information about school locations and enrollment. However, it is more difficult to get information about school quality, one of the key predictors of school location choice.

Workers are identified using a labor force participation model that is a function of age, gender, education, and presence of children in the household. Employed persons are then assigned using probabilistic choice models to their type of industry, work location (which is also a function of accessibility), weekly work duration, and work flexibility. Each individual is also assigned a driver's license depending on age, gender, and race. Using these characteristics, household income is computed as a function of race, presence of elderly individuals, education level of members of households, and employment industry of workers in the household. Then, a residential tenure model (own or rent) follows with a housing type model to assign each household to a single-family detached, single-family attached, apartment, and mobile home or trailer type of residence.

Car Ownership and Type

Policy analysis provides estimates of greenhouse gas emissions and energy consumption. This motivates the inclusion of model components that are capable of explicitly modeling vehicle fleet composition and usage, and the allocation of vehicles to primary drivers in the household. This set of models may also be viewed as longer- or medium-term choices that households make as opposed to daily short-term choices that are made in the context of daily activity travel engagement. This type of model in essence determines the predicted non-commercial regional vehicle fleet mix that is used as input to the emission estimation software. This is also particularly important because of the expected market penetration of "new" technology fuel-efficient vehicles (such as electric cars) and the incentive programs created at the state and federal levels in the US to promote such vehicles. A model system like this can be used to assess different incentive structures promoting environmentally friendly technologies in cars. One of the inhibitors in building car ownership, car body type, car vintage, and make models is the existence of many possible alternatives based on the combinations of the different body types, vintages, and makes. Further, a household may own more than one vehicle.

A very good option is to use a multiple discrete continuous extreme value (MDCEV) model capable of simulating the entire fleet of vehicles in a household and the annual mileage that each vehicle is driven or used. The multiple discrete continuous extreme value (MDCEV) model extends the classic multinomial logit model in innovative ways to accommodate the multiple discreteness in the choice process and simultaneously predict the total annual mileage that a vehicle would be driven. This annual mileage may be treated as a general mileage budget that guides the use of a vehicle on a day-to-day basis (for example, a household may choose to use a special collector's edition car only sparingly while using the family van for everyday chores). As indicated earlier, vehicle types are defined by body type, fuel type, and vintage of the car. Within each of the body and fuel types, the MDCEV model is capable of simulating the exact make and model of the vehicle, thus providing detailed information regarding the vehicle fleet in the population. By tracking individual vehicles throughout the course of a day, a complete trajectory can be built and this information can then be used to (a) estimate energy consumption and emissions models and (b) offer the baseline data for eco-driving advisory programs.

Following (or in combination with) the development of a vehicle fleet composition model, it is desirable to allocate each vehicle in the fleet to a primary driver in the household (see Vyas et al., 2011). This facilitates a "higher level" allocation of vehicles to drivers with the idea that drivers generally use the vehicle(s) allocated to them, particularly when undertaking drive alone trips. When undertaking joint activities and trips, then it is possible that a tour level vehicle choice is exercised and subsequent model components can effectively simulate such processes. However, the choice of vehicle will be influenced by the "higher level" allocation of vehicles to primary drivers. For this reason, survey data elicits questions about vehicle allocation are very useful (in addition to a complete inventory of all the household fleet vehicles and details about their type and fuel used).

At this point of the simulation cascade on the left side of Figure 1, the model system produces the spatial distribution of all the residents by different social and demographic levels (including race) as well as employment and school locations assigned to each person. In addition, each household is assigned to a housing type. This resembles a complete Census of the resident population and can be done at any level of spatial aggregation. One could also draw samples from this population or proceed to the next step using 100% of the simulated residents. This is particularly convenient when there is a whole range of scenarios or policies to examine. In this case, a sample can be used to quickly narrow down to the ones to study in more detail, and these can be taken through the 100% of simulated residents. It is also possible to focus on a specific subarea (e.g., a city) and perform detailed analysis and modeling while keeping the rest of the region as an evolving background. Thus, the data needs will depend on the spatial resolution and the specific use of the model.

Daily Activity Schedules and Travel Choices

For each synthetically generated household and person within each household, daily activity and travel patterns are created in this block of models. While there are a variety of model systems that have been implemented or developed in the activity-based modeling field, two families of models appear to have emerged. One family of models, which may be labeled as ***tour-based models*** (see Vovsha et al., 2005, Bradley et al. 2010) focus on the generation of tours and the

trips that comprise the tours, as well as the tour/trip characteristics. These models have been successfully implemented in practice and provide a robust framework capable of accounting for inter-dependencies among trips. These frameworks can be expanded to account for interactions among household members, although such extensions have not been very common. The tour based models in practice are largely comprised of a series of multinomial logit and nested logit models that are strung together to form a long chain of models, so that logsums (i.e., sums of exponentiated utility functions) from one set of logit models can feed into another choice model specification at a higher level in the hierarchy. This structure provides the ability to account for accessibility impacts on tour generation, destination choice, mode choice, and time of day choice. Time of day is treated as a discrete choice with the time of day being split into one hour, 30 minute, or 15 minute slices. By introducing logsum terms into the model specifications of higher level models, the tour based modeling paradigm is able to account for key behavioral phenomena of interest. Improvements in accessibility (as represented by changes in logsum values) will impact destination choice and tour generation, not to mention time of day choice and mode choice.

While the tour-based models provide a powerful practical framework for modeling daily activity-travel patterns, they are somewhat restrictive in their ability to simulate the emergent nature of activity-travel demand. Pendyala et al., 2011 review some of their limitations in detail. We favor ***continuous time activity-based model systems*** in which time plays an all-encompassing role, activity durations and time use are explicitly modeled, and spur-of-the-moment activity generation can be explicitly represented. Figure 2 constitutes a simplified representation of a worker's daily activity-travel pattern and also illustrates the temporal resolution survey data should provide.

Figure 2 goes here

In this example, simulation starts at 3 AM on the first day and proceeds for 24 hours until 3 AM on the second day. Within this framework, the continuous time model will first determine the fixed activity pegs (geographical anchors) around which discretionary activities and travel must be undertaken. The fixed activity pegs typically correspond to work activities, the timing of which are modeled in an earlier step based on the work-related characteristics (work duration,

flexibility, etc.) predicted in the long-term choice module. In the most recent versions of the continuous-time activity system for a large mega-region in the US, the fixed activity pegs include joint activities undertaken by individuals in the household (see Bhat, 2011). The activity pegs determine the daily time space diagram of a person's path, based on the notion of a time-space prism (Hägerstrand, 1970, 1989). The time periods corresponding to the fixed activity pegs are locked and do not constitute open time-space prisms in which individuals can pursue discretionary activities and travel. After higher level models establish the locked periods and the open time-space prisms, the simulator generates discretionary activities within open time-space prisms along the course of a day. History dependence is incorporated into the model framework so that activities in the latter part of the day are influenced by activities undertaken earlier in the day. For example, if a person has done two shopping episodes in the earlier part of the day, it is less likely that this person will undertake yet another shopping activity towards the end of the day. In each open prism, an individual can make multiple stops outside home, or simply choose to stay home until travel must be undertaken to the next fixed activity. The number of stops undertaken in each open time-space prism is dependent on the time available in the prism (the size of the prism), which is in turn influenced by the network accessibility and level of service measures. It should be noted that the figure is a rather simplified representation of a pattern. It is technically feasible to have any number of tours or trip chains within an open time-space prism, and within each tour, it is possible to have any number of stops.

For a non-worker, Figure 3 offers another example of a daily activity-travel pattern. In the case of non-workers, the degrees of freedom are greater as there is only home and joint activities (including serve passenger trips) as the fixed pegs around which discretionary activities and travel must be undertaken. In particular, for both workers and non-workers, it is necessary to first establish household responsibilities and child dependencies that may constrain certain periods of the day and limit the flexibility in scheduling and undertaking non-work activities and trips. In the case of non-workers, the simulation process then proceeds through the successive generation of tours with one or more stops in each tour. Individuals can return home for temporary stays during the course of the day. Within each tour, an unlimited number of stops can be made for various purposes, each with a destination, mode, and duration attached to it. The above framework is a hybrid of multiple elements and brings together the best of the behavioral

paradigms in the activity based travel modeling arena. On the one hand, household and individual level day-patterns are imposed such that certain periods of the day that need to be dedicated to activities that are fixed in space and time are set aside. On the other hand, the model framework accommodates the emergent nature of activity engagement and travel demand with the possibility for discretionary activities and travel to be undertaken on-the-fly.

Figure 3 goes here

The examples here can be implemented at the level of the traffic analysis zone (TAZ) and this is the preferred starting point of model implementation in the US because many regions have an already operational four-step model with TAZ populated with data. However, an implementation that treats space in as disaggregate a manner as possible (similar to the continuous disaggregate treatment of time) is ideal. This puts more pressure on land use data and network fidelity. These data are available and often of better quality at the traffic analysis zone, which is the level that many operational models use. But, over time, as the network fidelity gets better, land use models become more detailed, and land use data reliability at small spatial scales improves. As a consequence, the spatial resolution of the activity-based model system is increasingly becoming more disaggregate. In fact, the spatial unit for activity-based model systems is moving to the Census block group, block, or even individual activity locations (parcels) and points on a network. It is also important to note that model systems are being designed to operate at multiple scales, enabling developers to use data from different scales depending on data supply and methods to convert data from one scale to another (downscaling or upscaling). This discussion is clearly pointing to the need for surveys to record all locations visited using longitude and latitude as well as verified addresses, while time should be measured at the finest possible level.

At this stage of the model system, the output contains more complete data than an activity survey diary database (because it recreates the activity-travel patterns of the entire population of a region). This is an important consideration for verification and validation, since the simulation produces more information than is currently available from other sources to use as the gold standard or benchmark. Typical external data used to verify the output are commute statistics among subregions, other local surveys, or counts of persons at specific locations by time-of-day.

Routes, Assignment to Networks, and Emissions

The output of a continuous time activity-based travel model system such as the one just discussed can be used in many different ways. For example, testing of policy scenarios may be studied and their impact assessed by examining the timing decisions of individuals (e.g., advancing or postponing the starting of trips). The model can also be coupled with other algorithms that undertake traffic assignment, producing estimates of the number of cars on a network and/or routes chosen by vehicles from an origin to a destination. Continuous time activity-based models may also be interfaced with algorithms that are able to produce point-by-point stops in the daily activity and travel paths of simulated persons. While this approach to assignment is appealing from the standpoint of predicting each individual's actual path through the network, it should also be noted that the current state-of-the-art in activity based travel demand models covers only personal passenger travel; it does not cover freight travel, non-resident in the region tourist and visitor travel, and possibly some special generator travel such as airport travel. Some applications in this direction exist but are not widely used yet. For the above reason, post processing routines typically convert the activity-travel patterns into origin-destination trip tables and combine these trips with origin-destination trip tables of the other types of trips (obtained using existing classical processes) prior to feeding the entire travel demand into traffic assignment processes. This is the procedure used when a static traffic assignment process is used. If dynamic traffic assignment procedures are used, then it may make sense to retain the detailed trip records coming out of the activity-based travel model and feed individual trips or tours to the dynamic assignment model. Origin-destination trip tables of the other types of trips can be fed to the dynamic assignment model in the usual way; however, routines need to be setup to ensure that there is consistency in how trips from the activity based model and trips from the other origin-destination trip tables are routed through the network along the continuous time axis. Similar processing is also required when using traffic microsimulation techniques (such as the ones embedded in TRANSIMS and MATSIM), with the key difference of converting all travel into travel plans to mimic activity and travel schedules of people. One example of this type of application is TASHA in Toronto (Hao et al., 2010). Energy consumption and emissions are still (even in advanced models) relying on averaging of vehicles characteristics, average link operating speeds, and other approximations. Modeling advances are rapidly made towards developing emission speed profiles by different types of vehicles (Barth

and Boriboonsomsin, 2008) that in turn can be used to develop daily emission profiles for each vehicle and driver in a simulated activity-travel schedule at fine temporal and spatial resolutions.

Agent and Environment Evolution

The application of an activity-based travel model system requires the generation of a synthetic population for the entire region. The activity-travel patterns of the individuals in the synthetic population are simulated using the activity-based travel model. In this regard, there is considerable interest in being able to forecast the characteristics of the population in the future through the evolution of a base year population. The base year population would be taken through a series of models that mimic the lifecycle processes of households and individuals (the agents). These include household emigration and immigration, entry into and exit from labor force, aging, birth and death, marriage and divorce, household formation and dissolution, entry into and exit from school, acquisition of driver's license, and any other socio-economic phenomena that characterize the evolution of the population over time. There are thus two ways to generate a future year synthetic population that also define the type of data needs we face. Marginal distributions on control variables of interest for a future year may be obtained from a land use model, and these future year marginal distributions may be used to generate a synthetic population for the future year in a manner that is similar to that in the base year. Alternatively, the base year synthetic population may be aged and evolved through a series of lifecycle stage transition models on an annual time step to better replicate the demographic processes at play in the region. Life cycle stage transitions are in essence the simulation of turning points in the life of individuals. The population synthesis process within the activity-based model system should then be enhanced to include a series of lifecycle evolution models that would provide a future year population.

An evolution model system should also include processes to explain regional vehicle fleet evolution and predict the future year vehicle fleet as well as the path this fleet follows. In fact, just as a population of households and individuals evolves over time, so does a fleet or population of vehicles. New vehicles are acquired, old vehicles are scrapped, and some vehicles are simply swapped, traded, or replaced. It would be desirable to implement a series of vehicle

transition or transactions models that allow one to evolve the vehicle fleet in response to changes in socio-economic characteristics of the household, the available vehicle types and technologies, the regulatory policy environment, incentives and tax policies, costs of acquiring and maintaining vehicles, and vehicle attributes. In this way, vehicle fleet composition and usage levels can be forecast for any horizon year. The vehicle evolution models should include a suite of components capable of simulating the acquisition, disposal, and replacement of vehicles on an annual cycle.

The environment of firms and establishments (and, therefore, the opportunity space for activity participations of individuals and households) also changes, which, in turn, affects the behavior of individuals and households. One way to capture the evolving nature of the environment of firms and establishments is to adopt similar techniques as for household evolution. We can consider every establishment of firm and take it through transitions of creation, location choice, dissolution, movement to a different location, or any other type of morphing. This is known as Firmographics and it is used to some extent in microsimulation-based land use models. The task of evolving this aspect of the environment is usually in the domain of land use models and their integration with travel demand models.

There is yet another evolutionary aspect of the environment that is surprisingly somewhat neglected in modeling and simulation. A fundamental piece of information we use in travel behavior is travel impedance from a point to another, which is a function of the transportation infrastructure. Changes in the nature of the characteristics of the infrastructure have an impact on impedance and thus on behavior. For instance, schedules of transit and availability of lanes on a highway change over time (e.g., weekday vs weekend, winter vs summer) and they also change from one year to the next (e.g., additions and elimination of routes, opening of a new lane, conversion of lanes to HOV/HOT). However, most models do not consider such varying impedance levels, and do not collect information on travel behavior characteristics in the context of the impedance characteristics being experienced. The end result is differences in infrastructure data vintage and survey data vintage, and the resulting obvious potential threat to validity when one combines the two data sources to estimate behavioral models. In this context, it may be

desirable to have evolutionary models of infrastructure that provide yearly snapshots to parallel all the other models discussed above.

The Units of Analysis

The unit of analysis for an activity-based travel model is the individual, as individual activity travel patterns are simulated through the time-space domain. However, it is necessary to recognize that individual activity-travel patterns are influenced by household interactions and child dependencies. To account for and model these interactions, the household is also a unit of analysis and considered as the behavioral entity (Decision Making Unit). Outcomes are determined and reported at both levels (household and household member). At any point in the simulation, one should be able to collectively report the activity-travel engagement patterns of all individuals in a household and thus obtain a more holistic perspective on the household situation. Different models are estimated at appropriate levels to reflect the behavioral nature of the phenomenon under study. Work location and school location choice models are estimated at the level of the individual person. Vehicle fleet composition and usage models are estimated at the household level. However, there are additional levels within the person and household levels of decision making. Mode choice models are estimated at the tour and trip (segment) level. Destination choice models are estimated at the individual stop/trip level. Household-level joint activity engagement and allocation patterns may be estimated using the MDCEV model at the household level (Bhat, 2011). Activity type choice and generation models are, however, estimated at the individual activity level for activities generated on the fly. Activity duration models are estimated at the individual activity level as well. Vehicle type choice models are estimated at the tour level as one generally needs to retain the same vehicle through the completion of a tour. All this points to the need to have *data on activity locations, activities, stops, trips, tours, vehicles used for each movement, persons that act, and companionship in activities and travel as locations, stops, trips, and activities.* As mentioned earlier, time is to be treated as a continuous entity, while space, although may continue to be treated in a more aggregate fashion with existing data, is preferred to be recorded as longitude and latitude of every location visited during a diary. Continuous time is generally approximated using a one minute resolution, thus creating a total of 1440 possible time slices. Appropriate definitions should also be provided to remove any ambiguity regarding the interpretation or representation

of a tour in contrast to a trip. A tour is best viewed as a series of trips with the origin of the first trip and the destination of the last trip being exactly identical. In other words, a closed chain with intermediate stops constitutes a tour. There have been some alternative definitions for tours/chains, but the closed chain definition is likely to serve the activity based travel model development effort well as one can account for all inter-dependencies across all trips that are linked together in some way.

Total Design Data Needs

Figure 4 represents a somewhat ideal complete-total data provision survey design scheme that we use as benchmark for this paper. A design of this type contains a main household survey (the Core) that collects the base data elements needed for an activity-based model system but also serves other simplified versions of travel demand forecasting. This is the main household, person, and base diary portion of Figure 4. Then, subsamples are randomly selected and invited to participate in more detailed surveys targeting a specific topic area. Using suitable statistical methods (e.g., to account for selectivity bias), the responses can then be expanded to the entire main household sample and those, in turn, can be expanded to the entire population using synthetic population methods when desired. Surrounding the main household and activity diary portion are a variety of "satellite" surveys that shed light on specific behavioral facets and provide data for modeling and simulation. The satellite surveys of Figure 4 are defined by a theme, but there is no theoretical underpinning requiring each satellite component to be a separate entity from the rest. For example, one of the satellite surveys of Figure 4 spans an entire week for a smaller portion of the sample. This core component could also be designed as a wave of a panel and/or a continuous survey design to capture additional behavioral dynamics that are not included in our large-scale models yet, but we know would enhance our predictive ability if made available. In this design, a set of complementary survey components may be added to provide more in-depth behavioral data (about behavioral facets addressing other model components) as well as data for verification and validation. This overall design aims at collecting in-depth data by minimizing the overall cost and time to implement the survey and also provides flexibility in designing relatively independent survey modules (satellites). Each satellite aims at a set of individual objectives and distributes survey burden to a different group

of participants from the main survey, minimizing survey burden and fatigue as well as implementation costs. For these reasons we name this type of survey design a total data needs survey design that covers as many behavioral facets as possible to enable new generations of travel demand forecasting models. We proceed here with a brief description of possible satellite surveys (clockwise).

Figure 4 goes here

One Week Activity (and Travel) Diary: In this component, households are recruited to participate in an entire week diary. This will enable the creation of models that account for day-to-day variation in activity scheduling and travel and attempt to identify shifting of tasks and activities from one day of the week to the next. The survey design will dictate the optimal timing of this component to maximize response rate and completion rate, and minimize any biases. It is important that a survey of this type benefit from a design that is able to capture the behavioral processes of scheduling activities, and planning and subsequent re-scheduling modifications as in Auld et al. (2009).

In Depth Car ownership and Use: In this component, we envision the design of an in-depth survey to identify the determinants for each of the car ownership, car type (e.g., new/used, model, make, and fuel type), and car assignment decisions. In the car assignment data collection, both the primary and secondary drivers should be identified. Questions should also be created to identify determinants of changes in car ownership, type, and assignment of cars to household members. Particular emphasis should be given to policy controlled determinants (e.g., taxation, incentives). One approach to study this latter part is using combinations of revealed and stated preference surveys.

Location Choice and Activity Sattisfaction: Destination choice in conventional models is treated as a naïve selection among comparable objects. From environmental psychology, however, we know that places have symbolic and other meanings that travel behavior models neglect (see sense of place in Deutsch and Goulias, 2012). This component identifies how

destinations are perceived and what role these perceptions play in their selection. It also aims at quantifying the contribution to subjective well-being experienced for each activity and travel episode (see for an example Goulias et al., 2012). We would also suggest the design of a small scale survey following the day reconstruction method (DRM). There are two key objectives for this component: (a) provide a benchmark for the diary instrument; and (b) create an assessment of activities (including trips) and subjective experiences that is able to capture preferences, satisfaction, and perceived quality of life. This second set of objectives will enable estimation of choice models with latent variables and classes that are by far richer and more informative than their counterpart observed variable discrete choice models.

Residence, Workplace, and School Location Choice: This is a critical survey component for behaviorally integrated land use travel demand models. We expect this component to be an in-depth survey to identify the determinants for each of the residential, workplace, and school choices (see Kortum et al., 2012). Both primary locations and secondary locations should be examined in more detail than typical household surveys and data collected to estimate choice models for each facet.

Retrospective and Prospective Location Choices: Questions should also be created to identify determinants of change for each location examining behavior retrospectively and prospectively. Particular emphasis should be given to policy controlled determinants. This portion is shown in Figure 4 as a separate survey component because of the possible need to also add questions about personal biography of each household member using techniques that are not used by typical household surveys (e.g., ethnography).

Toll Willingness to Pay: In this supplemental survey, we envision identifying attitudes and willingness to pay for tolls on highways. The data in this component can be used to develop behavioral equations of the willingness to pay, which, in turn, enable the large scale regional

simulation models to develop pricing strategies (Bhat and Castellar, 2002; Bhat and Sardesai, 2006).

Long distance travel: Travel models in Mega regions and statewide applications also need models that are able to capture what is called interregional travel and long-distance travel. Many of the trips in this class are business related, leisure related, or simply long commutes. Regional forecasting applications need data to estimate this type of trip making, but also need data to correlate long distance travel and short distance travel. This data component aims to accomplish exactly this objective, and to enable the study of trade-offs people make when they engage in travel that, for example, requires an overnight stay outside the home base. In addition, it is also desirable to study the relationship between land use and the propensity to make long distance travel.

Expenditures and budgeting survey: Annual, monthly, or even weekly expenditures for activity participation, travel, and vehicles and housing units maintenance ownership and energy consumption are not collected in typical travel surveys. The most recent policy actions, however, require linking housing to transportation demand. In addition, efforts to develop more complete household Greenhouse footprints will increase and add to the move toward developing models of comprehensive accounting of energy demand. This component will provide the data needed to enable a direct association between travel and at home energy consumption to eventually create models of the type in Fissore et al. (2011).

In-Depth Mode Supplement and Active Living Questions: Mode choice is of paramount importance and particularly when linked to destination choice. This add-on survey collects information about detailed reasons for not using specific modes, including non-motorized modes for active living studies. The survey objective is to identify situational constraints, attitudes, and predispositions in favor or against modes such as walk, bike and public transportation. Moreover, collecting information about the chosen and not chosen modes enables the creation of

models to study policy actions that go beyond the time-cost-comfort analysis. It is also possible to add a stated choice, intentions, and preference components to this module. Equally important is also the added detail of collecting data about walking and biking either as a main mode for each trip or as an access mode to another main mode (e.g., walking from a parking lot to an office, biking to a bus stop and then taking the bus).

GPS and GPS OBD (verification, special days, emissions): This is a GPS household member tracking component to: a) develop a database to correlate destinations to routes and identify a typology of different types of routes and stop making patterns; b) develop a route choice model; c) estimate the level and nature of misreported trips by different modes of the main two-day activity diary; d) verify day-to-day behavioral change in other survey components and day of the week effects; and e) provide detailed operating characteristics of the household vehicles. This component for persons carrying GPS devices (wearable GPS) can also be supplemented with an online diary and vehicle-mounted GPS (week long to capture day to day variation) and On-Board Diagnostics devices (to identify driving patterns and correlate/link them with emissions models).

Panel of Households and Persons and Multi-Day Activity: Repeated observation of the same people over time provides a unique source of information for understanding change in behavior and develop more accurate travel demand models. Examples of this type of surveys includes designs such as Mobidrive (Axhausen et al., 2002) to identify weekly rhythms in activity scheduling and travel and household panel surveys that enable disentangling temporal ordering and causality in behavior (see the edited volume by Golob et al., 1997).

Figure 4 cannot be implemented in its entirety within the resources available for most surveys and requires modification to meet some additional data needs for model components of the specific region or a state/country developing their model systems. The minimum data elements required are household composition information, person characteristics (age, gender, education, employment, drivers license, marital status), and vehicle data. A base diary activity-travel diary

(two-day is desirable to study day-to-day variation, but a single day diary is current practice in the US) shall include a complete record of each person's daily schedules including all activities engaged in and all trips made (with a record of their assembly into tours), locations visited, the persons with whom each activity and trip were made, and activities carried out at home and at other places. Ideally, this diary will be for a pre-specified pair of days for all persons in the household and will spread over a 12–month period (to mimic the American Community Survey) with a uniform distribution of interviews throughout the survey period. Development of the next generation activity-based model(s) requires a greatly enhanced activity diary of complete households. The additional travel behavior indicators needed for the new models are usually incorporated in activity diaries. The usual design in this setting is a household questionnaire that includes social and demographic information of each household member, housing characteristics and auto ownership. The activity diary portion of the questionnaire includes but is not limited to respondents' activities, travel, and characteristics of the surrounding environment (including parking locations and parking rates). The overall survey design will also enable the study of toll modes and willingness to pay, gather additional data on walk/bike modes, Transportation Demand Management program participation, auto ownership (including fuel efficiency and fuel usage), and enable the study of equity and environmental justice issues. More detailed listing of the variables needed are included in the references to this paper.

Current activity-based models are based on data that are stitched together from a variety of surveys. Assumptions about validity of this action are usually untested due to necessity and lack of suitably designed surveys. Attempts, however, are made to design decennial Household Travel Surveys (e.g., the CHTS) and fill important gaps of information supporting large scale models. Similar attempts are also made to develop the data required to model the dynamically changing environment in which simulated agents live. In parallel, new model development continues to use whatever data are available with substantial improvements in our ability to simulate policies and validate models using internal to the new survey and external information. In data collection, however, and mainly with a push by the recent legislative innovation of coordinating land use policies with transportation policies we see positive change.

A much simpler version of Figure 4 is being explored for the California Household Travel Survey (CHTS, www.catravelsurvey.com). CHTS is currently (May 2012) in its initial months

with a possible delivery of data in March 2013. CHTS includes a core (household and single day diary covering the entire state with additional observations for some of the CA regions), a three-day wearable GPS and a seven day vehicle GPS with a small sample of On-Board Diagnostics (OBD), a long-distance (trips longer than 50 miles) travel log to capture trips made in two weeks preceding the diary day, a car ownership and type revealed and stated preference survey, and an "augment" survey for the Southern California region with questions about location choices. This shows survey practice is already attempting to satisfy some of the data needed for large scale regional simulation models but far from complete.

Although theoretically a satellite design of a large scale survey may solve the problem of the curse of dimensionality, respondent burden and fatigue, and exploding data collection budgets, it is unknown if a survey of this type is feasible from the survey data collection management/operations viewpoint and if multiple contractors (in CHTS there are three different contractor teams collecting data without a contractual obligation and pressure from the funding agencies to coordinate and collaborate) are able to coordinate their data collection efforts and provide a harmonized database for modeling and simulation. From this viewpoint, the California Household Travel Survey is a real life experiment in a state that requires policy analysis supported by this type of data in a timely fashion to demonstrate strategies of meeting Greenhouse gas emissions targets set by policy. For an example of these targets see the California Air Resources board website http://www.arb.ca.gov/cc/sb375/final_targets.pdf. The policy environment, substantial progress in modeling and simulation, and willingness to innovate in data collection will give us many opportunities to rethink about data needs and model development in new ways and at a faster pace than in any past epochs.

Appendix A Excerpt from California Legislative Initiative

Senate Bill 375 (SB 375) was enacted to reduce greenhouse gas emissions from automobiles and light trucks through integrated transportation, land use, housing and environmental planning. Under the law, SCAG is tasked with developing a Sustainable Communities Strategy (SCS), a newly required element of the 2012 Regional Transportation Plan (RTP) that provides a plan for meeting emissions reduction targets set forth by the California Air Resources Board (ARB).

On September 23, 2010, ARB issued a regional 8% per capita reduction target for the planning year 2020, and a conditional target of 13% for 2035.

SCS Requirements

According to SB 375, "each metropolitan planning organization shall prepare a sustainable communities strategy, including the requirement to utilize the most recent planning assumptions considering local general plans and other factors. The Sustainable Communities Strategy shall:

- identify the general location of uses, residential densities, and building intensities within the region;
- identify areas within the region sufficient to house all the population of the region, including all economic segments of the population, over the course of the planning period of the regional transportation plan taking into account net migration into the region, population growth, household formation and employment growth;
- identify areas within the region sufficient to house an eight-year projection of the regional housing need for the region;
- identify a transportation network to service the transportation needs of the region;
- gather and consider the best practically available scientific information regarding resource areas and farmland in the region;
- consider the state housing goals specified in Sections 65580 and 65581;
- set forth a forecasted development pattern for the region, which, when integrated with the transportation network, and other transportation measures and policies, will reduce the greenhouse gas emissions from automobiles and light trucks to achieve, if there is a feasible way to do so, the greenhouse gas emission reduction targets approved by the state board;
- allow the regional transportation plan to comply with the federal Clean Air Act“

Acknowledgments

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

References

Auld, J., Mohammadian, A., & Doherty, S. (2009) Modeling Activity Conflict Resolution Strategies Using Scheduling Process Data. Transportation Research Part-A: Policy and Practice, Vol. 43(4), 2009, pp. 386–400.

Axhausen, K.W., Zimmermann, A. , Schönfelder, S. , Rindsfüser, G. & Haupt, T. (2002) Observing the rhythms of daily life: A six-week travel diary, *Transportation*, **29** (2) 95-124.

Barth M. & Boriboonsomsin, K. (2008) Real-World CO2 Impacts of Traffic Congestion. http://www.uctc.net/research/papers/846.pdf

Bhat, C.R., Guo, J.Y., Srinivasan, S., & Sivakumar, A. (2004), "**Comprehensive Econometric Microsimulator for Daily Activity-Travel Patterns**," *Transportation Research Record,* Vol. 1894, pp. 57-66.

Bhat, C.R., & Castelar, S. (2002), "A Unified Mixed Logit Framework for Modeling Revealed and Stated Preferences: Formulation and Application to Congestion Pricing Analysis in the San Francisco Bay Area", *Transportation Research Part B*, Vol. 36, No. 7, pp. 593-616.

Bhat, C.R., & Sardesai, R. (2006), "**The Impact of Stop-Making and Travel Time Reliability on Commute Mode Choice**," *Transportation Research Part B*, Vol. 40, No. 9, pp. 709-730

Bhat, C.R., Goulias, K.G., Pendyala, R.M. , Paleti, R., Sidharthan, R., Schmitt, L., & Hu, H. (2011) A Household-Level Activity Pattern Generation Model for the Simulator of Activities, Greenhouse Emissions, Networks, and Travel (SimAGENT) System in Southern California, Paper Accepted for Presentation at the 2012 Transportation Research Board Meeting.

Bowman J. (2009) Historical Development of Activity Based Model Theory and Practice. http://jbowman.net/papers/2009.Bowman.Historical_dev_of_AB_model_theory_and_practice.pdf

California Government SB 375 (2011) http://www.leginfo.ca.gov/pub/07-08/bill/sen/sb_0351-0400/sb_375_bill_20080930_chaptered.pdf. Accessed September 2011.

Deutsch K. and Goulias K.G. (2012) Understanding Places Using a Mixed Method Approach. Paper 12-2984 presented at the *91st Annual Meeting of the Transportation Research Board, Washington, D.C.,* January 22-26, 2012. (in press)

Donnelly R., Erhardt, G.D. , Moeckel, R. & Davidson, W.A. (2010) Advanced Practices in Travel Forecasting. National Cooperative Highway Research Program, Synthetsis 406, Transportation Research Board, Washington.D.C.

Ferdous, N., Sener, I.N. , Bhat, C.R. & Reeder, P. "Tour-Based Model Development for TxDOT: Implementation Steps for the Tour-based Model Design Option and the Data Needs," Report 0-6210-1, prepared for the Texas Department of Transportation, October 2009

Fissore, C., Baker, L. A., Hobbie, S. E., King, J. Y., McFadden, J. P. , Nelson, K. C. & Jakobsdottir, I. (2011) Carbon, nitrogen, and phosphorus fluxes in household ecosystems in the Minneapolis-Saint Paul, Minnesota, urban region. Ecological Applications 2011 21:3, 619-639

Golob T., Kitamura, R. , & Long, L. (1997) Panels for Transportation Planning: Methods and Applications Kluwer.

Goulias K.G. (2007) Activity Based Travel Demand Model Feasibility Study. Final Report Submitted to the Southern California Association of Governments. Contract Number 07-046-C1. Work Element Number 07-070.SCGC08. June, Solvang, CA.

Goulias K.G. & Morrison, E. L. (2010) Pre-Survey Design Consultant for the Year 2010 Post-Census Regional Travel Survey. Final Summary Report Project Number 10-046-C1(April 2010 to July 2010). Submitted to Southern California Association of Governments and Caltrans. June, Solvang, CA.

Goulias, K.G., Bhat, C.R., Pendyala, R.M. , Chen, Y. , Paleti, R. , Konduri, K.C., Lei, T. , Tang, D., Yoon, S.Y. , Huang, G. , & Hu, H. (2011) Simulator of Activities, Greenhouse Emissions, Networks, and Travel (SimAGENT) in Southern California, Paper Accepted for Presentation at the 2012 Transportation Research Board Annual Meeting.

Goulias, K.G., Ravulaparthy S., Yoon S.Y., & Polydoropoulou, A. (2012) An exploratory analysis of the time-of-day dynamics of episodic hedonic value of activities and travel. Paper presented at the 2012 International Association for Travel Behavior Research Conference, July 15-20, Toronto, Canada.

Hägerstrand T. (1970) What about people in Regional Science? Papers in Regional Science, 24(1), 6-21.

Hägerstrand T. (1989) Reflections on “what about people in regional science?” Papers in Regional Science, 66(1), pp 1-6.

Kortum, K., Paleti, R. , Bhat, C.R. & Pendyala, R.M. (2012), "A Joint Model of Residential Relocation Choice and Underlying Causal Factors," Paper presented at the 2012 TRB Conference.

Pendyala R. (year) A Household Travel Survey Data Collection Plan. Report for Maricopa County, Tempe, AZ.

Pendyala, R.M., Bhat, C.R. , Goulias, K.G. , Paleti, R. , Konduri, K.C., Sidharthan, R., Hu, H., Huang, G. & Christian, K.P. (2011) The Application of a Socio-Economic Model System for Activity-Based Modeling: Experience from Southern California, Paper Accepted for Presentation at the 2012 Transportation Research Board Annual Meeting and publication in the Transportation Research Record.

Rossi T., Bowman, J., Vovsha, P. , Goulias, K. G., & Pendyala, R. (2010) CMAP Strategic Plan for Advanced Model Development. Final Report of the CMAP Advanced Travel Model Cadre. Chicago, Il.

Vyas, G., Paleti, R., Bhat, C.R., Goulias, K.G. , Pendyala, R.M. , Hu, H. , Adler, T.J., & Bahreinian, A. (2011) A Joint Vehicle Holdings (Type and Vintage) and Primary Driver Assignment Model with an Application for California, Paper Accepted for Presentation at the 2012 Transportation Research Board Annual Meeting.

Yagi, S., & Mohammadian, A. (2010) An Activity-Based Microsimulation Model of Travel Demand in the Jakarta Metropolitan Area, in the Journal of Choice Modelling, , Vol. 3, No. 1, pp. 32-57.

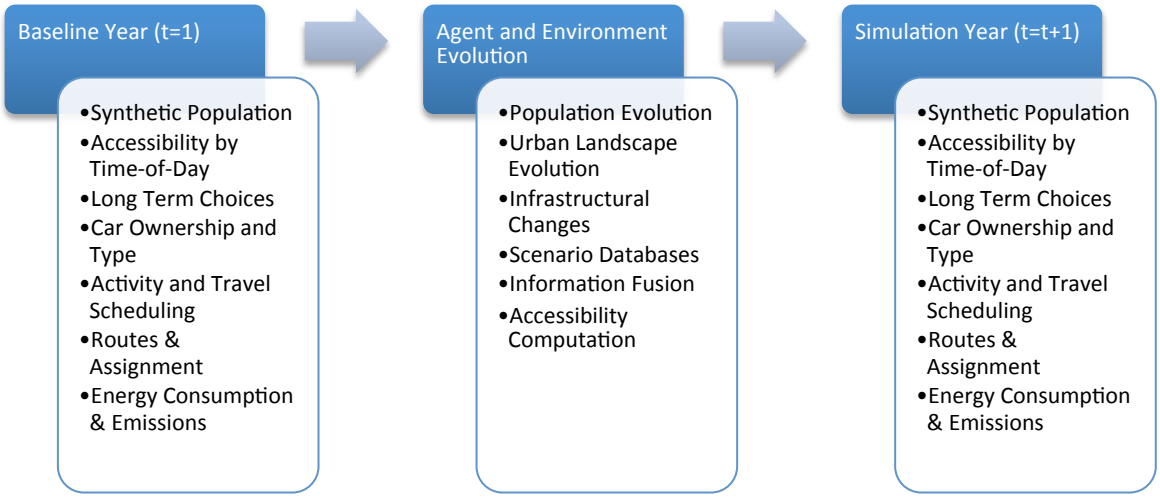

Fig. 1 A Schema of a Typical Model

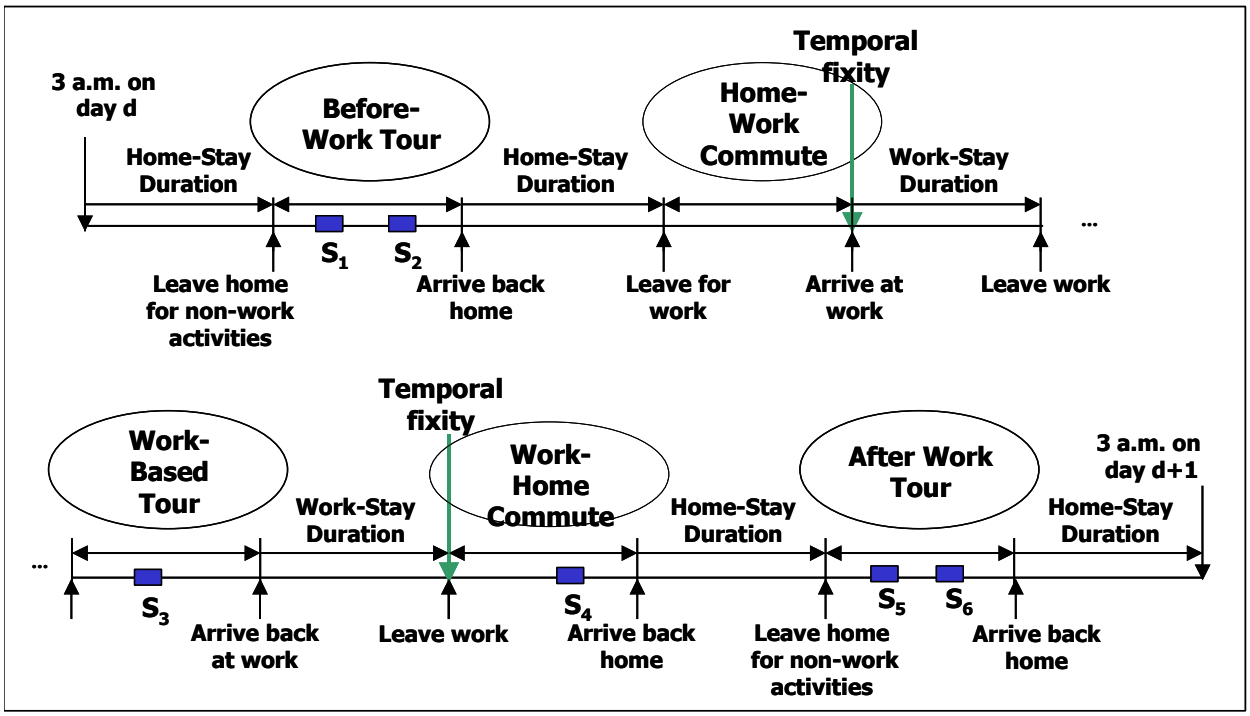

Fig. 2. Representation of a Worker's Daily Activity-Travel Pattern
(Reproduced from Bhat et al., 2004)

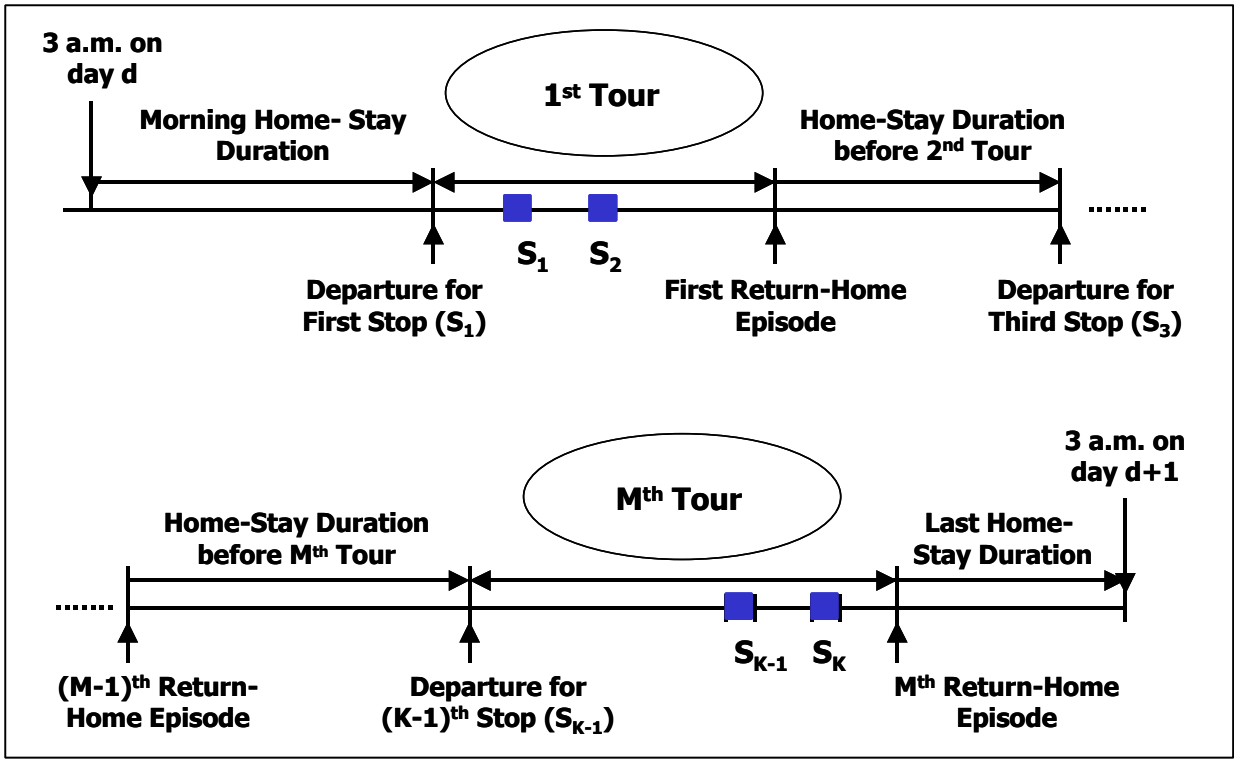

Fig.3. Representation of a Non-Worker's Daily Activity-Travel Pattern
(Reproduced from Bhat, et al., 2004)

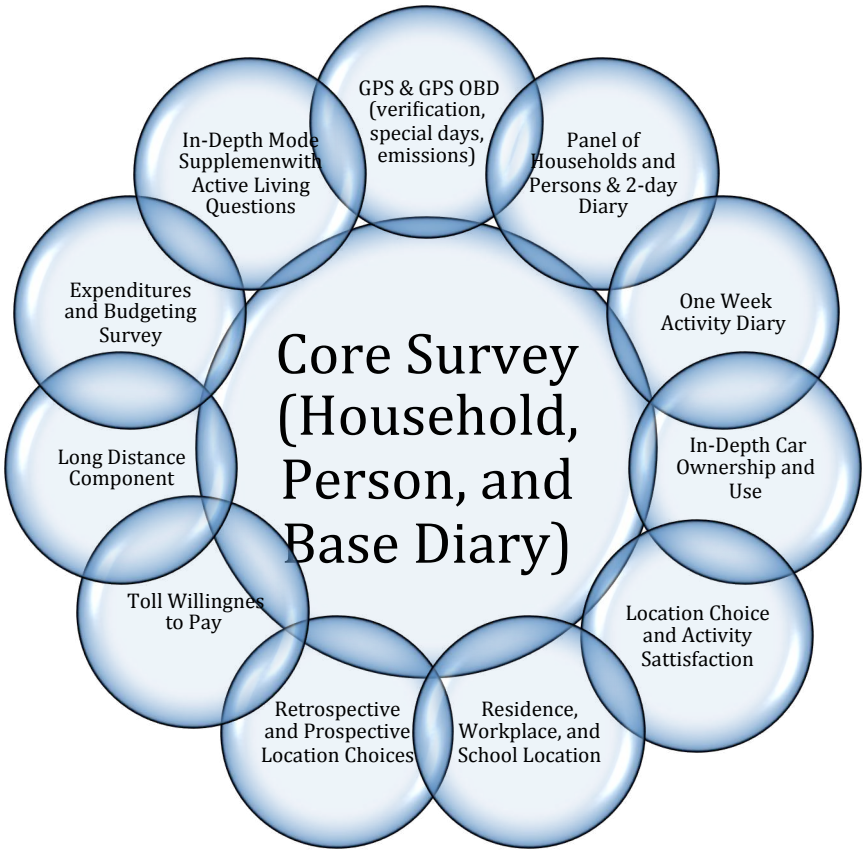

Fig. 4 The Data Collection Overall Scheme