Activity-Based Model for a Medium Sized City: Sacramento

Mark A Bradley, 524 Arroyo Ave., Santa Barbara, CA 93109, USA. +1-805-564-3908, mark_bradley@cox.net

John L. Bowman, Ph. D., 28 Beals Street, Brookline, MA 02446, USA, +1-617-232-8189, John_L_Bowman@alum.mit.edu, website: http://JBowman.net

Bruce Griesenbeck, Principal Transportation, Analyst, Sacramento Council of Governments, +1-916-340-6268, bgriesenbeck@sacog.org

ABSTRACT

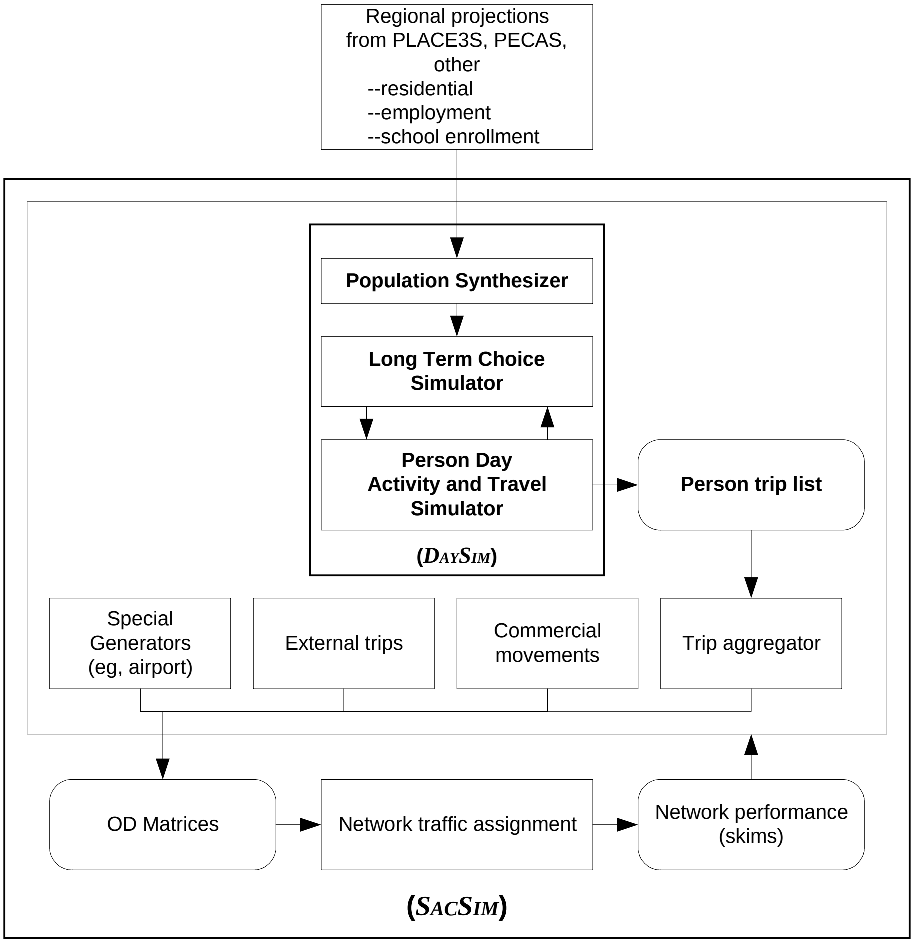
This paper presents the activity-based regional travel forecasting model system being used by the Sacramento, California, Council of Governments. The SACSIM model system represents travel in the context of an integrated disaggregate econometric model of each resident's full-day activity and travel schedule. Sensitivity to neighborhood scale is enhanced through disaggregation of the modeled outcomes in three key dimensions: purpose, time, and space. Each activity episode is associated with one of seven specific purposes, and with a particular parcel location at which it occurs. The beginning and ending times of all activity and travel episodes are identified within a specific 30-minute time period. The model system has been calibrated and tested for a base year of 2000 and for forecasts to the years 2005 and 2032. The paper summarizes the model system structure, explains the integration with the traffic assignment model, discusses issues with preparing input data for such a model system, and presents application results.

INTRODUCTION

This paper presents a regional travel forecasting model system called SACSIM, used by the Sacramento (California) Area Council of Governments (SACOG) for planning and air quality analysis. SACSIM includes an integrated econometric microsimulation of personal activities and travel with a disaggregate treatment of activity purpose, time and location.

Figure 1 shows the major SACSIM components. The Population Synthesizer creates a synthetic population, comprised of households drawn from the U.S. Census Public Use Microdata Sample (PUMS) and allocated to parcels. Long-term choices (work location, school location and auto ownership) are simulated for all members of the population. The Person Day Activity and Travel Simulator then creates a one-day activity and travel schedule for each person in the population, including a list of their tours and the trips on each tour. These components, together called DaySim and implemented in a single custom software program, consist of a hierarchy of multinomial logit and nested logit models. The models within DaySim are connected by adherence to an assumed conditional hierarchy, and by the use of accessibility logsums. The trips predicted by DaySim are aggregated into matrices and combined with predicted trips for special generators, external trips and commercial traffic into time- and mode-specific trip matrices. The network traffic assignment models load the trips onto the network. Traffic assignment is iteratively equilibrated with DaySim and the other demand models.

Figure 1: New SACOG Regional Travel Forecasting Model System



As shown here, the regional forecasts are treated as exogenous. In subsequent implementations, it is anticipated that SACSIM will be integrated with PECAS, Sacramento's land use model (Abraham, Garry and Hunt, 2004), so that the long range PECAS forecasts will depend on the SACSIM forecast.

DAYSIM OVERVIEW

DaySim follows the day activity schedule approach developed by Bowman and Ben-Akiva (2001). Its features include the following:

- Using microsimulation, the model predicts outcomes for each household and person, producing activity/trip records comparable to those from a household survey (Bradley, et al, 1999).
- It simulates choices at four integrated levels—long-term, day, tour and trip.
- The models of longer term decisions and activity/tour generation are sensitive to network accessibility and land use variables.

- There are seven activity purposes for tours and intermediate stops (work, school, escort, shop, personal business, meal, social/recreation).

- A person's work tour destination for the day can differ from their usual work location.

- In location choice it predicts an individual parcel.

- It predicts trip and activity start and end times to the nearest 30 minutes, using an internally consistent scheduling structure that is sensitive to differences in travel times across the day (Vovsha and Bradley, 2004).

- The model is highly integrated, including the use of logsums and approximate logsums in the upper level models, encapsulating differences across different modes, destinations, times of day, and types of person.

Table 1 lists DaySim's component models. They are numbered hierarchically in the table and in subsequent references. The hierarchy embodies assumptions about the relationships among simultaneous real world outcomes. In particular, outcomes from models higher in the hierarchy are treated as known in lower level models. It places at a higher level those outcomes that are thought to be higher priority to the decision maker. The model structure also embodies priority assumptions about the relative priority of tours in a pattern, and of stops on a tour. The formal hierarchical structure provides what has been referred to by Vovsha, Bradley and Bowman (2004) as downward vertical integrity.

Table 1. Component Models of DaySim

Model #	Model Name	Level	What is predicted
1.1	Synthetic Sample Generator	Household	Household location, size, composition and income; person age, gender, employment status and student status
1.2	Regular Workplace Location	Worker	Workplace location parcel
1.3	Regular School Location	Student	School location parcel
1.4	Auto Ownership	Household	Number of autos
2.1	Daily Activity Pattern	Person-day	0 or 1+ tours for 7 activity purposes. 0 or 1+ stops for 7 activity purposes
2.2	Exact Number of Tours	Person-day	For purposes with 1+ tours, 1, 2 or 3 tours.
3.1	Tour Primary Destination	(Sub)Tour	Primary destination parcel
3.2	Work-Based Subtour	Work Tour	Number and purpose of subtours made during a work tour
3.3	Tour Main Mode	(Sub)Tour	Main tour mode

3.4	Tour Time of Day	(Sub)Tour	The time period arriving and the time period leaving primary destination
4.1	Intermediate Stop Generation	Half Tour	Number and activity purpose of intermediate stops made on the half tour
4.2	Intermediate Stop Location	Trip	Destination parcel of intermediate stop, conditional on tour origin, destination, and location of stops nearer to the tour destination
4.3	Trip Mode	Trip	Trip mode
4.4	Trip Departure Time	Trip	Departure time 30 min. period, conditional on time windows remaining from previous choices

Just as important as downward integrity is upward vertical integrity, achieved by using composite accessibility variables to explain upper level outcomes. This makes the upper level models sensitive to attributes that are known only at the lower levels of the model, especially travel times and costs. It also captures non-uniform cross-elasticities among lower level alternatives.

When there are billions of alternatives (as in the entire day activity schedule model), the preferred measure of accessibility, the expected utility logsum, is computationally infeasible. So, for SACSIM, approaches were developed to approximate true logsums. Approximate logsums are calculated in the same basic way, by summing the exponentiated utilities of multiple alternatives, but ignoring minor differences among decisionmakers and using a carefully chosen subset or aggregation of the available alternatives. The approximate logsum is pre-calculated and used by several of the model components, and can be re-used for many persons.

Another simplifying approach involves simulating a conditional outcome. For example, in the tour destination choice model, where time-of-day is not yet known, a mode choice logsum is calculated based on an assumed time of day, where the assumed time of day is determined by a probability-weighted Monte Carlo draw. In this way, the distribution of potential times of day is captured across the population rather than for each person, making the destination choice model sensitive to time-of-day changes in travel level of service.

In many other cases within the model system, true expected utility logsums are used.

COMPONENT MODELS OF DAYSIM

Following are highlights of some of the DaySim component models. Similar models are grouped together, for ease of presentation. For more details, see SACSIM Technical Memos (Bowman and Bradley, 2005-6), available at http://JBowman.net .

Day activity pattern (2.1-2.2)

This model is a variation on the Bowman and Ben-Akiva approach, jointly predicting the number of home-based tours a person undertakes during a day for seven purposes, and the occurrence of additional stops during the day for the same seven purposes. The pattern choice is a function of many types of household and person characteristics, as well as land use and accessibility at the residence and, if relevant, the usual work location. The main pattern model (2.1) predicts the occurrence of tours (0 or 1+) and extra stops (0 or 1+) for each purpose, and a simpler conditional model (2.2) predicts the exact number of tours for each purpose. The "base alternative" in the model is the "stay at home" alternative where all 14 dependent variables are 0 (no tours or stops are made).

Many household and person variables were found to have significant effects on the likelihood of participating in different types of activities in the day, and on whether those activities tend to be made on separate tours or as stops on complex tours. They include employment status, student status, age group, income group, car availability, work at home dummy, gender, presence of children in different age groups, presence of other adults in the household, and family/non-family status. For workers and students, the accessibility (mode choice logsum) of the usual work and school locations is positively related to the likelihood of traveling to that activity on a given day. For workers, the accessibility to retail and service locations on the way to and from work is positively related to the likelihood of making intermediate stops for various purposes.

Simpler models were estimated to predict the exact number of tours for any given purpose, conditional on making 1+ tours for that purpose. Compared to the main day pattern model, the person and household variables have less influence but the accessibility variables have more influence. This result indicates that the small percentage of people who make multiple tours for any given purpose during a day tend to be those people who live in areas that best accommodate those tours. Other people will be more likely to participate in fewer activities and/or chain their activities into fewer home-based tours.

Number and purpose of work-based tours (3.2)

For this model, the work tour destination is known, so variables measuring the number and accessibility of activity opportunities near the work site influence the number and purpose of work-based tours. This model is very similar in structure to the stop participation and purpose models described next.

Stop participation and purpose (4.1)

For each tour, once its destination, timing and mode have been determined, the exact number of stops and their purposes is modeled for the half-tours leading to and from the tour destination. For each potential stop, the model predicts whether it occurs or not and, if so, its activity purpose. This repeats as long as another stop is predicted. The outcomes of this model are strongly conditioned by (a) the outcome of the day activity pattern model, and (b) the outcomes of this model for higher priority tours. For the last modeled tour, this model is constrained to accomplish all yet-unaccomplished intermediate stop activity purposes prescribed by the activity pattern model.

The estimation results indicate that accessibility measures are important in determining which stops are made on which tours, as well as the exact number of stops. An important feature of this model system is that it does not predict the number and allocation of stops completely at the upper pattern level, as is done in the Portland and SFCTA models, or completely at the tour level, as is done in other models such as those in Columbus and New York. Rather, the upper level pattern model predicts the likelihood that ANY stops will be made during the day for a given purpose, at a level where the substitution between extra stops versus extra tours can be modeled directly. Then, once the exact destinations, modes and times of day of tours are known, the exact allocation and number of stops is predicted using this additional tour-level information. We think that this approach provides a good balance between person-day-level and tour-level sensitivities.

LOCATION MODELS

Usual work (1.1) and school (1.2) locations, tour destinations (3.1)

The dependent variable in the usual location and tour destination models is the parcel address where the activity takes place. Since over 700,000 parcels comprise the universal set of location choice alternatives in the SACOG six-county region, it is necessary to estimate and apply the location choice models using a sample of alternatives. Using two-stage importance sampling with replacement, first a zone is drawn according to a probability determined by its size and impedance, and then a parcel is drawn within the zone, with a size-based probability.

Several important differences exist among these models. For the usual work and school location models, auto ownership is assumed to be unknown, based on the assumption that auto ownership is mainly conditioned by work and school locations of household members, rather than the other way around. For the tour destinations, auto ownership levels and usual locations are treated as known. For the two usual location models (work and school), the home location is treated as a special location, because it occurs with greater frequency than any given non-home location, and size and impedance are not meaningful attributes. Because most work tours go to the usual work location, the work tour destination model has this as a special alternative.

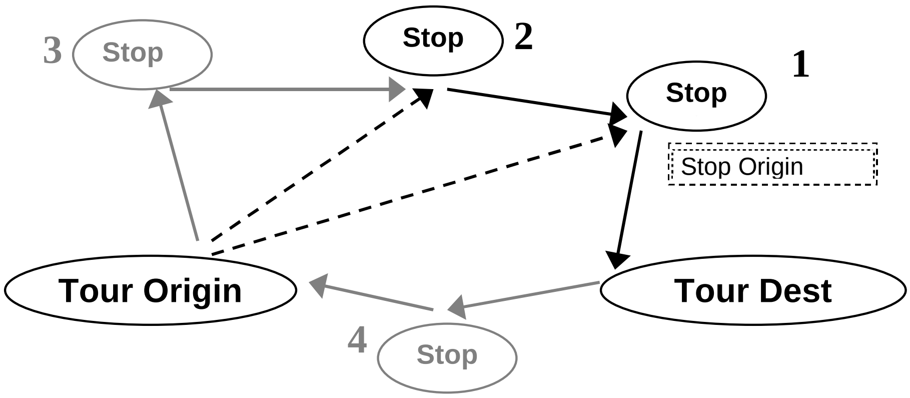
Two important variables in all of these models are the disaggregate mode choice logsum and network distance. The logsum represents the expected maximum utility from the tour mode choice, and captures the effect of transportation system level of service on the location choice. Distance effects, independent of the level of service, are also present to varying degrees depending on the type of tour being modeled. In nearly all cases, sensitivity to distance declines as distance increases.

In most cases the models include an aggregate mode-destination logsum variable at the destination. A positive effect is interpreted as the location's attractiveness for making subtours and intermediate stops on tours to this location. A mix of parking and employment, at both the zone and parcel level, as well as street connectivity in the neighborhood, attract workers and tours for non-work purposes. Also, parcel-based size variables and zone-level density variables affect location choice.

Intermediate stop location (4.2)

Figure 2 depicts the modeling of location choice for the stop numbered 2 in a tour with four intermediate stops. Solid arrows represent the temporal sequence of stops. Numbers indicate the order in which the stops are modeled. When modeling the stop 2 location choice, the tour origin and stop 1 location (stop origin) are known, so the model considers the extra time and cost required to get to stop 1 via stop location 2 instead of going directly. So the choice of location involves comparing, among competing locations, (a) the impedance of making a detour to get there, given the tour mode, and (b) the location's attractiveness for the given activity purpose. The model is a multinomial logit (MNL).

Figure 2: Modeling the location of intermediate stop 2.



Trip characteristics used in the model include stop purpose, tour purpose, tour mode, tour structure, stop placement in tour, person type, and household characteristics. The most important characteristics are the tour mode and the stop purpose. The tour mode restricts the modes available for the stop, and this affects the availability and impedance of stop locations. The availability and attractiveness of stop locations depend heavily on the stop purpose. Tour characteristics also affect willingness to travel for the stop, and the tendency to stop near the stop or tour origin. These trip and tour characteristics tend to overshadow the effect of personal and household characteristics in this model.

The main impedance variable is generalized time, as well as its quadratic and cubic forms, to allow for nonlinear effects. Additional impedance variables used in the model include **travel time as a fraction of the available time window**, which captures the tendency to choose nearby activity locations if there are tight time constraints on the stop, and **proximity** variables (inverse distance), which capture the tendency to stop near either the stop origin or the tour origin.

MODE CHOICE MODELS

Tour main mode (3.3)

The tour mode choice model uses eight modes, although some of them are only available for specific purposes:

(1) DT- Drive to Transit: Available only in the Home-based Work model, for tours with a valid drive-to-transit path.

(2) WT- Walk to Transit: Available in all models except for Home-based Escort, for tours with a valid walk to transit path.

(3) SB: School Bus: Available only in the Home-based School model, for all tours.

(4) S3- Shared Ride 3+: Available in all cases.

(5) S2- Shared Ride 2: Available in all cases.

(6) DA- Drive Alone: Available in all models except for Home-based Escort, for tours made by persons age 16+ in car-owning households.

(7) BI- Bike: Available in all models except for Home-based Escort, for all tours with round trip road distance of 30 miles or less.

(8) WK- Walk: Available in all models, for all tours with round trip road distance of 10 miles or less.

The models' values of time and out-of-vehicle/in-vehicle time ratios are shown in Table 2.

Two land use variables, significant in many of the models, increase the probability of walk, bike and transit:

Mixed use density: This is defined as the geometric average of retail and service employment (RS) and households (HH) within a half mile of the origin or destination parcel (RS * HH / (RS + HH)). This value is highest when jobs and households are both high and balanced. High values near the tour origin tend to encourage walking and biking, while high values near the tour destination more often encourage transit use.

Table 2: Tour Mode Choice Level of Service Coefficient Summary

Model	Value of time ($/hr)	Ratio Walk to In-Vehicle	Ratio Wait to In-Vehicle
Home-Based Work	$11.20	2.95	2.50
Home-Based School	$6.00	2.20	2.20
Home-Based Escort	$7.50	3.00	N/A
Home-Based Other	$7.50	2.72	2.72
Work-Based	$7.50	2.84	2.84

Intersection density: This is defined as the number of 4-way intersections plus one half the number of 3-way intersections minus the number of 1-way "intersections" (dead ends and cul de sacs) within a half mile of the origin or destination parcel. Higher values tend to encourage walking for School and Escort tours, where safety for children is an issue, and also to encourage walking, biking and transit for Home-Based Other tours.

Trip mode (4.3)

The trip-level mode is conditional on the predicted tour mode, but now uses a specific OD pair and a time anchor, and also the trip mode for the adjacent, previously modeled trip in the chain. The majority of tours use a single mode for all trips, so this model only explains the small percentage of trips that are made by modes other than the main mode.

TIME OF DAY MODELS

DaySim employs a method of modeling time of day developed by Vovsha and Bradley (2004). These models are able to capture the effects of transport level of service on the time-of-day choice, including the tendency to shift travel away from periods when travel congestion is heavy.

The time of day models explicitly model the 30 minute time periods of arrival and departure at all activity locations, and hence for all trips between those locations. This provides an approximate duration of the activity at each activity location. The model uses 48 half-hour periods in the day, beginning at 3:00AM.

Tour primary destination arrival and departure time (3.4)

For each tour, the model predicts the person's time of arrival and subsequent time of departure at the primary destination. It includes as alternatives the 1716 possible combinations of the 48 half-hour time periods. Since entire tours, including stop outcomes, are modeled one at a time, first for work and school tours and then for other tours, the periods away from home for each tour become unavailable for subsequently modeled tours.

Intermediate stop arrival or departure time (4.4)

For each intermediate stop made on any tour, this model predicts either the time that the person arrives at the stop location (on the first half tour), or else the time that the person departs from the stop location (on the second half tour). On the second (return) half tour, we know the time that the person departs from the tour destination, and, because the model is applied after the stop location and trip mode have been predicted, we also know the travel time from the primary destination to the first intermediate stop. As a result, we know the arrival time at the first intermediate stop, so the model only needs to predict the departure time from among a maximum of 48 30-minute alternatives. This procedure is repeated for each intermediate stop on the half tour. On the first (outbound) half tour, the stops are simulated in reverse order from the primary destination back to the tour origin, so we know the departure time from each stop and only need to predict the arrival time. As stops within a tour are modeled, the periods occupied by each modeled stop become unavailable for subsequently modeled stops and tours.

SACSIM EQUILIBRATION

Figure 1 shows a cyclical relationship within SacSim between network performance and trips: DaySim and the auxiliary trip models use network performance measures to model person-trips, which are then loaded to the network, determining congestion and network performance for the next iteration. The model system is in equilibrium when the network performance used as input to DaySim and the other trip models matches the network performance resulting from assignment of the resulting trips. 'Network performance' as used here refers to times, distances, and costs measured zone-to-zone along the paths of least generalized cost.

To achieve equilibrium, SacSim uses the method of convex combinations that is employed by almost all convergent trip-based model systems, for which the theory of system equilibrium is well-developed (Evans, 1976). Equilibrium assignment iteration loops (a-iterations) are nested within iterations between the demand and assignment models (da-iterations). During each a-iteration, link volumes are estimated and combined in a convex combination with link volumes from the prior a-iteration, and from the prior da-iteration, in order to prevent link volume oscillation between a-iterations and between da-iterations. The user can tune the system for rapid convergence by controlling the parameters of the convex combinations. In order to further speed up convergence, DaySim, which runs once for each da-iteration, can be set to simulate activity and travel schedules for only a small fraction of the population in the early iterations, scaling up the results to represent the schedules of the travel of the entire population. Bowman, et al (2006) report convergence results from tests employing various iteration sequences of convex combinations and DaySim population fractions. In the future, some applications of SacSim may need to reduce the randomness of trip forecasts below what is inevitable from the Monte Carlo DaySim process at full sampling. This can be achieved by supersampling, which is running DaySim on a sample larger than the population and down-scaling the results, or by running DaySim multiple times and averaging the results.

APPLICATION ISSUES

The biggest application issue has been the development of forecast year parcel/point datasets required by SACSIM. Development of the model was based on parcel/point data from Year 2000 surveys and inventories of population, employment and land use. Application of the model was based on synthesized datasets for the model base year (2005) and for all forecast years for the MTP.

The primary parcel/point data source was SACOG's parcel-based land use database, called Place3s. Place3s is a GIS-based land use scenario generator (Allen, et al. 1996). Scenarios built at parcel level, with land uses characterized by "place type", which includes assumptions about the type, density, and mix of uses. SACOG uses a palette of about 50 place types. Total development levels are controlled by aggregate county-level econometric forecasts adopted by the SACOG Board for use in the development of the MTP. Place3s was used to estimate dwelling units and employment (9 sectors) at parcel level.

Even with the basic demographic variables forecasted at parcel level, other datasets which are very important for predicting travel behavior do not come naturally from Place3s, and were prepared separately. These variables include: small-area demographics needed to control the development of synthetic populations; K12 schools, colleges and universities; some sectors of employment (e.g. medical employment not associated with hospitals and large medical centers); paid off-street parking facilities; transit stops; and street-pattern variables.

Demographics to control the development of synthetic populations were built up from the Place3s parcel-level estimates for dwelling units. The control variables for the population synthesis are household size (1,2,3,and 4+ persons); workers per household (0,1,2, and 3+ workers); income level (5); and age of head-of-household (over/under 55 years). Demographic profiles based on control variables for three dwelling unit structure types (single family, multi-family 2-4 units, and multi-family 5+ units) were drawn from Year 2000 Census tabulations for regional analysis districts within the region. The profiles are applied to the Place3s estimates of dwellings by type at traffic analysis zone level. The resulting files are used directly by SACOG's 4-step travel model (SACMET), and are used as control files for the SACSIM population synthesis.

School locations and types are built up at point-level from a Year 2005 inventory of schools to future years by adding future schools. For K12 schools, future school needs are calculated at zone-level by tallying growth in school-age children in the synthetic populations. For example, the Year 2035 land use forecasts require about 300 new K12 schools. Where possible, future school sites are identified in local agency general plans and school district plans. In practice, only a minority of future K12 sites are explicitly identified in planning documents, and the majority of future K12 sites are manually identified based on the location of residential growth and judgement. Future colleges and universities are based on known plans for these facilities.

Place3s estimates medical employment associated with hospitals and large medical centers. All other medical employment associated with smaller clinics, private offices, and other medical-related uses are included within estimates of office and service employment sectors. Other medical employment is split out from these more aggregate categories based on proximity of parcel to the hospitals and large medical centers. For parcels very near hospitals/medical centers, a higher percentage of the total office/service employment is medical; as distance increases, the percentage decreases. Rates for this post-processing were based on Year 2005 employment inventories.

Paid off-street parking facilities are built up at point-level from a Year 2005 inventory in a manner similar to the build-up of K12 schools. The growth in paid off-street parking spaces is calculated at zone level, based on the growth in employment by density range. In general, paid off-street parking is directly related to density of development: as the density of development on a parcel increases, the likelihood of paid off street parking, and prices charged, increases. The "yields" of paid off-street parking are calculated at zone-level based on the amount of growth in several density ranges, with facility locations identified based on judgement within each zone. The yield rates were computed from a Year 2005 inventory of parking facilities, and Year 2005 Place3s development density estimates. Paid parking is also related to special uses, like colleges/universities and hospitals, and facilities are added at future locations of these uses.

Proximity to transit is measured as orthogonal distance from parcel to the nearest transit station or stop in SACSIM. Transit stops are also built up at point level from a Year 2005 inventory of transit stops. New future transit stop points are based on a comparison of forecast year and Year 2005 transit networks from the travel demand model. Where there are new transit lines are added, new stops are added to the inventory. In areas with little or no change in transit service, the Year 2005 stop inventory is used. For rail and express bus facilities, stations and stops as coded in the travel demand model are used directly. For fixed route bus services, the travel demand model stops under-predict actual stops. This is because zone-based travel models do not include sufficient detail to capture the stop-spacing for local bus routes, especially in urban areas. In these areas, stops points are synthesized along the bus routes and added to the Year 2005 inventory points.

Street pattern variables are used in several location and mode choice models in SACSIM, and are strongly related to non-motorized mode choice. The key street pattern variables are the buffered densities or numbers of intersections of three types: 1-leg intersections (e.g. cul-de-sacs); 3-leg intersections (e.g. a “T”); and 4+-leg intersections (e.g. a four-way intersection). Higher levels of 1-leg intersections are associated with lower likelihood of trip linking and non-motorized modes of travel; higher levels of 3- and 4+- leg intersections are associated with higher likelihood of trip linking and non-motorized travel modes. While future densities and mixes of use in growth areas are captured in the Place3s land use scenarios, future street pattern is not. Street patterns profiles for growth areas are “borrowed” from Year 2005 observed street patterns by place type and density level.

Each one of these data issues required significant time and effort to address. However, with the exception of transit stops, the data are prepared only once for each land use data run, and the process is becoming more routinized and efficient. Virtually all of these issues needs to be addressed for zone-based models, but the aggregate nature of the zones allows for the data to be developed with less rigor and hand-wringing. The discipline of developing the datasets at parcel/point level simply requires that all the assumptions be laid out explicitly.

This correct reference:

Bowman, J.L., M.A. Bradley and J. Gibb, **The Sacramento Activity-Based Travel Demand Model: Estimation And Validation Results**, presented at the European Transport Conference, September 18-20, 2006, Strasbourg, France, 2006.

should replace this incorrect reference:

Bowman, J. L., Bradley, M. A. (2006) ‘The Sacramento Activity-based Travel Demand Model: Estimation and Validation Results’ Presented at the 2006 European Transport Conference, Strasbourg, France, October, 2006.