White Paper for the Conference on Innovations in Travel Demand Modeling, Austin, TX, May 2006

A Summary of Design Features of Activity-Based Microsimulation Models for U.S. MPOs

Mark Bradley & John Bowman

This short paper provides a concise summary of important design features of various activity-based model systems that have been implemented or have recently been designed for planning agencies in the U.S. The models described are those for Portland, San Francisco, New York, Columbus, Atlanta, Sacramento, Bay Area, and Denver. We selected these models because they are in the same "family" of activity-based models, and one or both of the authors have been involved in the design of all of them except for New York. We have not included some other examples of activity-based models in the U.S., such as the CEMDAP model for Dallas, the FAMOS model for Tampa Bay, the TranSIMS model for Portland, or the TLUMIP model for the State of Oregon. Although those models share some of the features discussed here, we are not familiar enough with them to compare them at the level of detail included here, although that could be a useful extension of this paper.

All of the model systems described in this paper share a similar overall structure, with a hierarchy of levels from "top" to "bottom", with the lower choices predicted conditional on higher level choices. The levels are:

- **Population synthesis** (geographic allocation of households)
- **Longer term decisions**: auto ownership and (in some cases) work and school locations
- **Person/household-day level**: number of tours and activities made for various purposes
- **Tour-level**: The main destination and mode, begin/end times, and number of stops
- **Trip-level**: Intermediate stop location, and the mode and departure time of each trip

Within this structure, there are several important design features that distinguish the models, and these are summarized in the table below. The models are listed in the table more or less chronologically, with the earliest ones at the left and the later ones at the right. At the time of writing, the Bay Area (MTC) and Denver (DRCOG) models are in the design stage, and so the design characteristics shown for those models are those that are currently envisioned. Each paragraph below is a more detailed annotation of a row in the comparison table.

Controls/categories for population synthesis: All of the model systems simulate persons one by one, and require a representative sample of households and persons for the base year and forecast years. All of the regions use zone-level data and forecasts of household size and income as control variables for sampling households from the regional PUMS households. In addition, most of the regions have used the number of workers in the household as a third control variable, both because it is important behaviorally, and because CTPP Table 1-75 provides a useful 3-way joint distribution of household size, number of workers and income for 2000. The Portland (METRO) and San Francisco (SFCTA) models have also used age of head of household as a control variable, and Atlanta (ARC), Bay Area and Denver. Are all considering using age or age-related variables as well (e.g. presence of children and/or senior citizens). The sample generation software created for Atlanta has a flexible system for designating and combining control variables, as well as facilities for testing how well the synthetic population matches other variables which have not been controlled for explicitly. An important test will be how well the age distribution is matched when age is not one of the explicit control variables.

"Usual" work & school locations modeled at the top level?: There is a recognition that the choice of where to work and where to go to school are longer-term decisions that are not adjusted day to day, similar to the choice of residence (which is implicitly modeled in the synthetic sample). In most of the models, and all of the more recent ones, the “usual” work and school places are modeled at the “top” level, meaning that these are predicted before predicting any choices specific to the travel day. The home location is typically one of the alternatives in the choice set, for people whose main workplace is at home or who are home-schooled. Note that certain types of individuals such as construction workers or traveling salespeople may not have a “usual” workplace. Also note that this model formulation requires that data be collected on each worker’s most frequent work location, even if that person does not visit that location on the survey diary day(s). The destination for any particular work tour will most often be the “usual” work location, but may be another location instead (a business meeting, for example), and that choice is modeled accordingly at the tour level. School tours nearly always go to the usual school location, so a separate school tour destination model may not be needed. In the future, it would be ideal for the population synthesis and longer term models to be replaced by a dynamic, integrated land use model that includes joint prediction of residential and workplace (re)location decisions.

Number of out-of-home activity purposes: The simplest purpose segmentations are in the first version of the Portland model, with 3 purposes (work/school, maintenance and discretionary), and in San Francisco, also with 3 purposes (work, school, other). Most other model systems have included at least 7 activity purposes, being work, school, escort (serve passenger), shopping, meals, personal business (or “other maintenance”), and social/recreation (or “other discretionary”). In some cases, social visit has been separated from recreation. The main reasons for splitting out the meal activity are that it tends to be done at certain types of locations, and has very specific time-of-day and duration characteristics. The escort activity also tends to be to specific locations at specific times in terms of driving children to/from school. Note that in tour-based models we do not need to treat non-home-trips as if they are separate “purposes”, although all of the systems do have separate tour level models for work-based tours (often called “subtours” because they are tours within tours). In most of the model systems, the division of the school purpose into university, K-12 and pre-school is made in the lower level models based on the age and enrolment type of the particular person in the sample.

Number of in-home activity purposes: In the Portland models, in-home activities are distinguished between 3 purposes (work/school, maintenance and discretionary), but this distinction is only made for the “primary” activity of the day, and is only predicted in cases when the person has no out-of-home activities. This distinction did not appear to add substantially to the explanatory of the models. That fact, coupled with the fact that most survey respondents’ are reluctant to provide much detail about their in-home activities, explains why none of the other models distinguish between types of in-home activities. Since some of the models predict which people work primarily at home, that provides some substitution between in-home and out-of-home work. It does not, however, handle the phenomenon of part-time telecommuting, which is the focus of some TDM policies. As a result, there is some interest in predicting work-at-home as a separate activity type in the Bay Area model if the data will support it.

Day pattern type linked explicitly across HH members?: This and the following three paragraphs are concerned with the modeling of explicit linkages between the predicted activities and travel of different members of the same household. All of the models treat such linkages implicitly through the use of a wide variety of person type and household composition variables, and indeed one of the main advantages of the microsimulation approach is the ability to reduce aggregation bias by including such case-specific variables. Using explicit linkages takes that ability one step further and reduces aggregation bias even more. One of the key linkages is a fairly simple one. If each person’s full day activity pattern is classified into three main types—stay at home, go to work/school, or travel for some other purpose—then we see strong similarities between the patterns of members of the same household, even stronger than the similarities that would be predicted indirectly. The Columbus model system includes a sequential model of these linkages, simulating children first, and then adults conditional on what the children do. The Atlanta model system includes a similar model that is estimated simultaneously across all household members, avoiding the need to assume the order in which they are simulated and thus the direction of causality. A similar model is planned for the Bay Area system.

Joint activities linked explicitly across HH members?: Joint activities are cases in which two or more household members travel together to and from an activity location, and participate in the same activity while at that location. In the lower level models such as mode and destination choice, it is best to model such cases as a single joint decision, rather than as independent decisions made by different people. The Columbus and Atlanta model systems include models of household joint activity generation and participation. The application of the Columbus model has shown that predicting joint travel can have significant implications for mode choice, so this type of model has been recommended for the Bay Area model. However, in a wider sense the "jury is still out" as to what extent the additional accuracy of explicitly modeling household interactions will merit the additional complexity. For that reason, such models will not be included in the Denver system, at least in the initial version.

"Escort" trips linked explicitly across HH members?: Another type of joint travel is the case where two or more household members travel together to and/or from an activity location, but do not participate in the same activity there. The most common example is a parent driving a child to school and then either returning home (an escort tour) or else driving on to work (an escort stop on a work tour). Because these types of tours are partly joint and partly independent, it can be very complex to explicitly link them across persons. For that reason, explicit modeling of escort linkages has not been done in any of the applied models or recommended for the models under design. Most of the models, however, do include a separate "escort" purpose, so that the most important special characteristics can be captured—particularly that fact that the mode is nearly always auto, with the exception of infrequent cases of walk escort. Also, childrens' school locations can easily be included as special alternatives in the parents' escort tour destination choice sets, so that at least the location is accurate, even if the exact trip timing and car occupancy are not matched.

Allocated activities divided explicitly among HH members?: Certain types of activities such as grocery shopping, escorting, and some other "maintenance" chores, are likely to be allocated across individuals in a household, showing a negative correlation across frequencies within a household-day. The Columbus and Atlanta model systems include explicit models of the generation of these activities at the household level and then allocation to particular individuals. In the Atlanta case, this model was estimated jointly with the household joint travel generation model. Compared to explicitly linking people who make joint tours together, predicting which people within a household perform allocated activities appears less important to the model results—we are not changing anything fundamental about the tours, just which person makes them. So, in terms of the tradeoff between accuracy and complexity, these models seem less crucial than the joint travel models, and thus they have not been recommended for the Bay Area models. In addition, the limited number of activity categories offered in most surveys makes it rather difficult to determine which activities are most likely to be allocated. For example, grocery shopping is mainly an allocated activity, while shopping for a good book to read is an individual activity, but both are usually coded the same.

Level at which intermediate stop purpose and frequency are modeled: When ordering the models in an activity-based system from "top" to "bottom", it is not always clear which decisions should be modeled conditional on which other decisions. A prime example is the generation of intermediate stops made during tours. Are activities planned and combined into trip chains when a person is planning their day, in which case the mode, timing and location of the tours may depend on which stops they contain? Or, conversely, do people make tours, and then decide during the tour how often and where to make stops depending on their mode and location? Clearly, both of these describe real behavior, and which description is more accurate depends on the particular person and the types of activities they are carrying out. The Portland and San Francisco models follow closely the original Bowman and Ben-Akiva day pattern approach, in which the number and purpose of any intermediate stops are predicted at the person-day level before any particular tours are simulated. In contrast, the Columbus, New York and Atlanta models predict only the number and purpose of tours at the person-day level, and then the number and purpose of intermediate stops on any particular tour are predicted at the tour level once the tour destination, time of day and main mode are known. In the Sacramento models, an intermediate approach is used. Some information about stop-making is predicted at the person-day level, predicting whether or not any intermediate stops are made for each activity purpose during the day (7 yes/no variables). These are predicted jointly with the choice whether or not to make any tours for each of the activity purposes (7 more yes/no variables), thus capturing some substitution effects between the number of tours and the number of trips per tour. Then, when each tour is simulated, the exact number of purpose of stops on each tour is predicted conditional on the mode and destination of that tour and conditional on what types of stops still need to be simulated to fulfill the person-day level prediction. There is no obvious behavioral reason for this structure, other than that it "balances" the model sensitivities between the two types of behavior described above. A similar approach is planned for Denver and recommended for the Bay Area.

Number of network zones used: The next three paragraphs discuss spatial aspects of the model systems. In all cases, the zone system used for model development and application is the same as was also used for trip-based models. The auto and transit networks and assignments are also the same as used in the trip-based models. This fact has facilitated the transition to activity-based models, but at the same time, the microsimulation framework can also be used with more detailed spatial systems, and would support more accurate traffic simulation methods as well.

Smaller spatial units used below zones? : Because the microsimulation framework is not tied as strongly to zone definitions, it is possible to use the zones only to provide the road and transit path level of service variables, while variables related to land use, parking, and walk access (which do not need to be stored as matrices) can specified at a finer level. The Portland model uses such an approach for roughly 20,000 "blocks", while the Sacramento models use over 700,000 parcels. An intermediate approach, which has been recommended for the Denver and Bay Area models, is to divide zones with heterogeneous transit and walk accessibilities into more homogeneous "subzones", but with assignments and skims still done at the larger zone level.

Simultaneous mode and destination choice model estimation?: It has become a sort of tradition in modeling to condition mode choice upon a known destination, sometimes using a sequential nested structure where the mode choice logsum is used in the destination choice model. That is probably appropriate for purposes such as work and school. For purposes such as shopping, however, the choice of store may depend more upon the mode used than vice-versa. Simultaneous estimation of mode and destination choice allows the modeler to test different nesting hypotheses. Such an approach was used in the Portland model, and may be used in Denver as well.

Network and modeled time periods: Most 4-step models only use two times of day—peak and off-peak, and use fixed time-of-day factors. All of the activity-based models contain tour time of day models that allow some sensitivity of time of day choice to network conditions. All of the models have used at least 4 network assignment periods—AM peak, midday, PM peak and off-peak. In some cases, free flow conditions are assumed for off-peak, so no traffic assignment is needed for that period. In some models, a fifth period has been added by splitting the off-peak period into early morning and evening/night. The more recent models, beginning with Columbus, use more precise time windows in order to schedule each tour and trip consistently during the day. This involves keeping track of the available time windows remaining after "blocking out" the time taken by each activity and associated travel. The time windows can also be used in the activity generation models. The Sacramento model and perhaps other models are moving to half-hour periods to provide even more detail. The main constraint on how small the time periods can be is the adequacy of the self-reported times in the diary survey data. There is evidence that people often round clock times to 10, 15 or 30 minute intervals.

Tour time of day relative to mode and destination choice models: It is not obvious whether activity and departure times should be predicted before mode and destination choice, between them, or after both. There is some empirical evidence that shifts in time of day occur at two levels: the choice among broad periods of the day (e.g. morning, afternoon, etc.) is made fairly independently of accessibility, while smaller shifts of up to an hour or two are more sensitive to travel times and costs—the peak-spreading effect. Since all of the models use broad network time periods, the tendency has been to model the choice of these periods for tours at a fairly high level above mode and destination choice (although in most cases the usual destination for work and school tours has already been predicted). In some models, time of day choice is predicted between the destination and mode choice levels, which allows the use of destination-specific mode choice logsums in the time of day model, but requires that the destination choice model assume (or stochastically select) a specific time of day for the impedance variables.

Departure time choice modeled separately at the trip level?: Perhaps the placement of the model that predicts the choice of times for the overall tour is not as crucial if there is a separate model that predicts the departure time for each trip to the more detailed periods, conditional on the mode and OD of each trip. Some of the model systems include such a model as the "lowest" model in the system. It is also possible to include such a model for car trips only, in order to predict the shape of the demand profile within the broader peak periods.

Accessibility measures in the upper level models: Last, but certainly not least, is the issue of how to most accurately include accessibility and land use effects in the upper level models. Calculation of full logsums across all possible nests of lower level alternatives is clearly infeasible with so many levels of choices. The earliest Portland models came the closest to including "proper" individual-specific logsums, but the structure of that model was relatively simple, and the effect on model run-time was severe. The San Francisco models include mode-specific measures with set boundaries, such as the number of jobs accessible within 30 minutes by transit. The rather arbitrary cutoff boundaries in such measures can cause unexpected sensitivities when applying the models. The New York and Columbus models use mode-specific travel time decay functions that approximate the logsum from a simple destination choice model. Such measures perform better, but still have the problem that they are mode-specific, and that auto and transit accessibility tend to be correlated, so it is difficult to estimate model parameters for both of them. A method that solves this problem and is more consistent with discrete choice theory is to approximate joint mode/destination choice logsums. However, the mode choice logsums tend to vary widely across the population, so it is best to calculate different accessibility measures for different population segments. The Sacramento models use such an approach, with aggregate accessibility logsums for each combination of 7 travel purposes, 4 car availability segments, and 3 walk-to-transit access segments—as those tend to be the most important segmentation variables in the mode choice models. The Sacramento approach is described in more detail in another paper for this conference by Bowman and Bradley.

References for further details

Include a list of references… mostly TRB papers.

Model design feature	Portland Metro I / II	San Fran. SFCTA	New York NYMTC	Columbus MORPC	Atlanta ARC	Sacramento SACOG	Bay Area MTC (*)	Denver DRCOG (*)
Controls / # categories for population synthesis	4 hh size 4 income 4 age	4 hh size 3 # workers 4 income 3 age	5 hh size 4 # workers 4 income	5 hh size 4 # workers 4 income	100+ comb. of hh size, # workers, income, age	4 hh size 4 # workers 4 income	4 hh size 4 # workers 4 income Age (?)	4 hh size 3 # workers 4 income age (?)
"Usual" work & school locations at top level?	No / Yes	Yes	No	No	Yes	Yes	Yes	Yes
Number of out-of-home activity purposes	3 / 8	3	4	7	8	7	7 or 8	7 or 8
Number of in-home activity purposes	3	1	1	1	1	1	1 or 2	1
Day pattern type linked explicitly across HH?	No	No	No	Yes, sequential	Yes, simultaneous	No	Yes, simultatneous	No
Joint activities linked explicitly across HH?	No	No	No	Yes	Yes	No	Yes	No
"Escort" trips linked explicitly across HH?	No	No	No	No	No	No	No	No
Allocated HH activities allocated explicitly?	No	No	No	Yes	Yes	No	No	No
Level where stop purpose and frequency are modeled	Person-day	Person-day	Tour	Tour	Tour	Person-day and tour	Person-day and tour	Person-day and tour
Network zones (approx.)	1,250	1,900	6,000	2,000	2,500	1,300	1,600	2,800
Smaller spatial units used below zones?	No / Yes 20K blocks	No	No	No	No	Yes, 700K parcels	Transit access subzones (?)	Transit access subzones (?)
Mode and destination model estimation	Simultaneous	Sequential	Sequential	Sequential	Sequential	Sequential	Sequential	Simultaneous for non-work
Network time periods	5 per day	5 per day	4 per day	5 per day	4 per day	4 per day	5 per day	8 per day
Modeled time periods	5 per day	5 per day	4 per day	1 hour	1 hour	30 min	30 min (?)	30 min (?)
Use of time window duration in scheduling?	No	No (may be added)	No	Yes	Yes	Yes	Yes	Yes
Tour time of day relative to mode and destination	Above both	Above both	Between them	Between them	Between them	Between them	Between them	Above both for non-work
Departure time modeled separately at trip level?	No	No (may be added)	No	No	Yes, lowest model	Yes, lowest model	Yes, lowest model	Yes, lowest model
Accessibility measures in upper level models	Person-specific mode / dest logsums	Jobs reached by zone/ mode/ time band	Dest choice logsums by zone / mode / segment	Dest choice logsums by zone / mode / segment	Dest choice logsums by zone / mode / segment	Mode & dest logsums by zone / segment	Mode & dest logsums by zone / segment	Mode & dest logsums by zone / segment

(*) These model systems are currently in the model design phase.