ACTIVITY-BASED MODEL DESIGN AND WORK PLAN

Draft Final Report

Incorporating:

Technical Memorandum 1—State of the Practice
Technical Memorandum 2—Recommended Model Design
Technical Memorandum 3—Implementation Plan

December 4, 2008

Prepared by:

BRADLEY, BOWMAN & CASTIGLIONE

Mark A. Bradley
524 Arroyo Ave.
Santa Barbara, CA 93109
805-564-3908
Mark_Bradley@cox.net

John L Bowman, PhD
28 Beals Street
Brookline, MA 02446
617-232-8189
John_L_Bowman@alum.mit.edu
www.JBowman.net

Under Subcontract to:

Gannett Fleming

TABLE OF CONTENTS

TECHNICAL MEMORANDUM 1—STATE OF THE PRACTICE

STATE OF THE PRACTICE

Introduction

The purpose of this technical memorandum is to gather information that will be useful in designing the proposed activity-based (AB) model system for the Tampa Bay Region. It includes a historical overview of the development for practical use of AB models in the United States, a comparative summary of their features, lists of lessons learned by the current project team and by US agencies currently using their AB model for policy analysis, a list of requirements specific to the FDOT and the Tampa Region, a list of data requirements with description of data limitations that must be dealt with.

Activity Based Models and Trip-Based Models: A Comparison

We compare a trip-based model system to an activity-based model system because the Tampa Bay region and just about everybody else starts with a trip-based model system, and because much of the framework and components are the same. Figure 1 shows the framework of a trip-based model system. The transport model system takes inputs from the urban forecasting and planning process. These include transport networks representing potential future scenarios, zonal attributes representing forecasts of size and distribution of employment and population in the region, and socioeconomic attributes of the population. It then predicts zone-to-zone trip flows and assigns those trips to the network. This occurs iteratively to achieve a final result where inputs to the trip demand models are consistent with predicted link flows and travel speeds coming from assignment of the demand they generate. In some cases, accessibility measures from the transport model system serve as input to the land use forecasting models. The trip demand includes trips associated with special generators, such as airports, trips carried out by persons who live outside the region, trips conducted by residents of the region, and commercial movements. These trips are usually modeled by separate demand models, and the results are combined into trip matrices for highway and transit assignment. Here is where we can start talking of the difference introduced by an activity-based model.

In an activity-based model system, shown in Figure 2, only the trips of residents are modeled using an activity-based model. These are produced by a household travel demand simulator. It generates a synthetic population representing the future population of the region, or uses a synthetic population generated by a disaggregate land use model. Then it predicts activities and trips for every member of each synthetic household. In some cases, the activity-based model uses a large database of parcel attributes instead of only zonal attributes. It takes a lot of resources to generate the parcel database, both for model development and for forecast scenarios. But it is very desirable because it provides substantially better information about transit accessibility, travel times for non-motorized and short trips, and destination attractiveness in heterogeneous zones. When parcel data is available the models use information about each individual parcel as well as information about the nature of a buffer zone surrounding the parcel, such as the number and type of intersections and parking spaces. Parcel buffer attributes represent neighborhood effects more realistically than zonal attributes because they are centered on the parcel. The main output of an activity-based travel demand simulator is a detailed itinerary for every person in every household. It comes in the form of tables or lists that identify attributes of each household, each person in that household, each person's day, each travel tour in their day, and each trip on that tour. These trips are then aggregated and combined with the other predicted trips into matrices for assignment.

So, at this point we see all the components of an activity-based model set needed for transport forecasting. It looks much like a trip-based model system, and uses many of the same components.

Figure 1: Trip-based model system

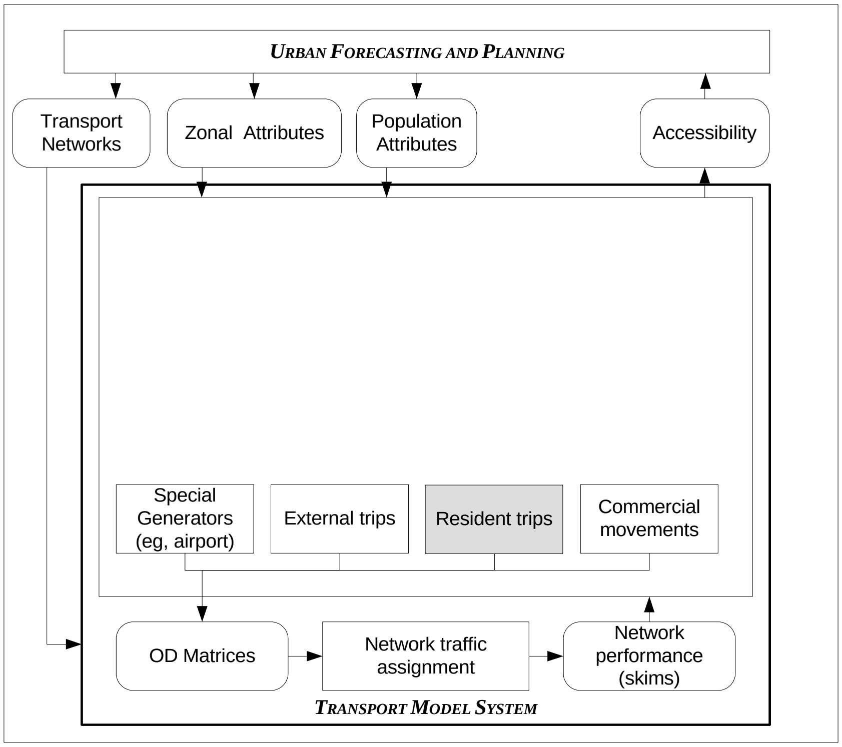
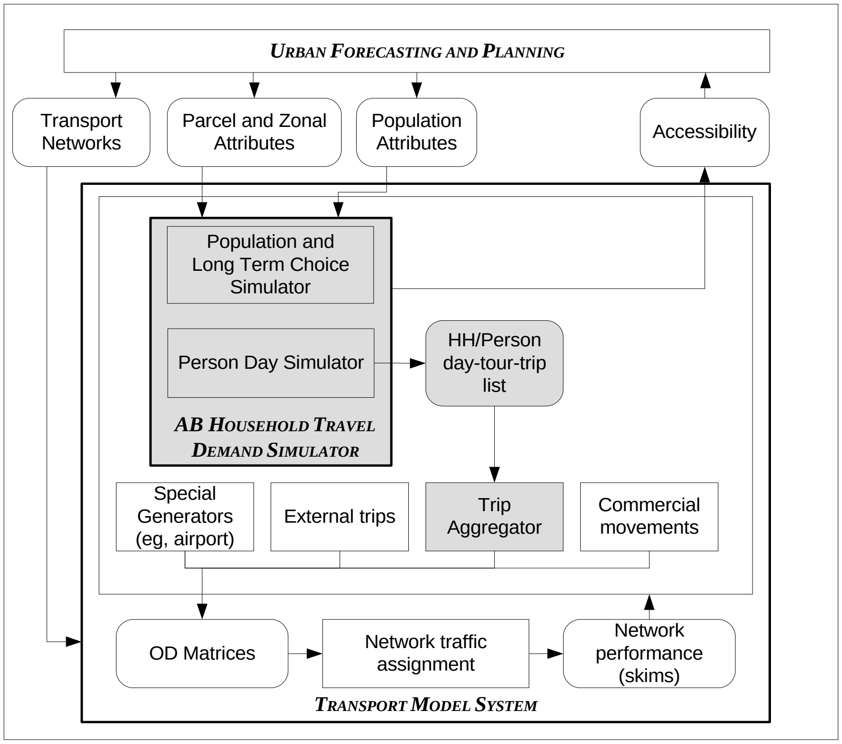
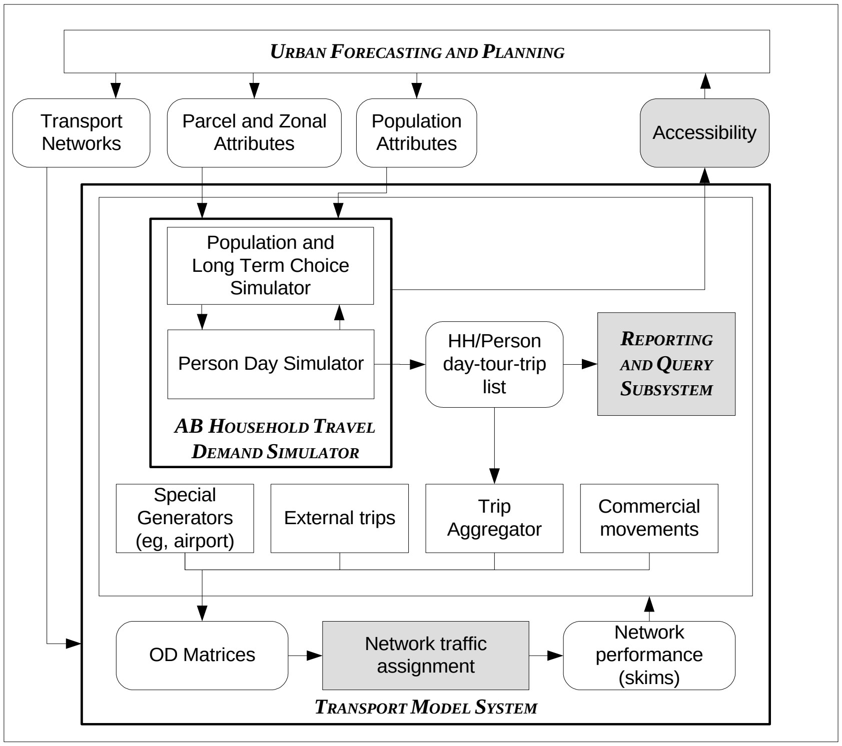


Figure 2: Activity-based model system



But a couple additional things are needed to unlock its power. These are highlighted in Figure 3. In order to leverage the simulator's ability to predict travel by time of day it may be necessary to enhance the highway and transit assignment components to assign trips for more time periods. The activity-based model output can also be aggregated in a great variety of ways, in order to extract information that is not available from trip-based models, especially the impact of a policy on various population subsets. So, to get at this information, it is desirable to create a reporting and query system that can be used to produce standard reports, generate custom queries, and display results on GIS maps. Finally, the activity-based simulator has the potential to supply better measures of accessibility as input to land use models. The existing activity-based models have not yet exploited this potential, but they are rapidly heading in that direction.

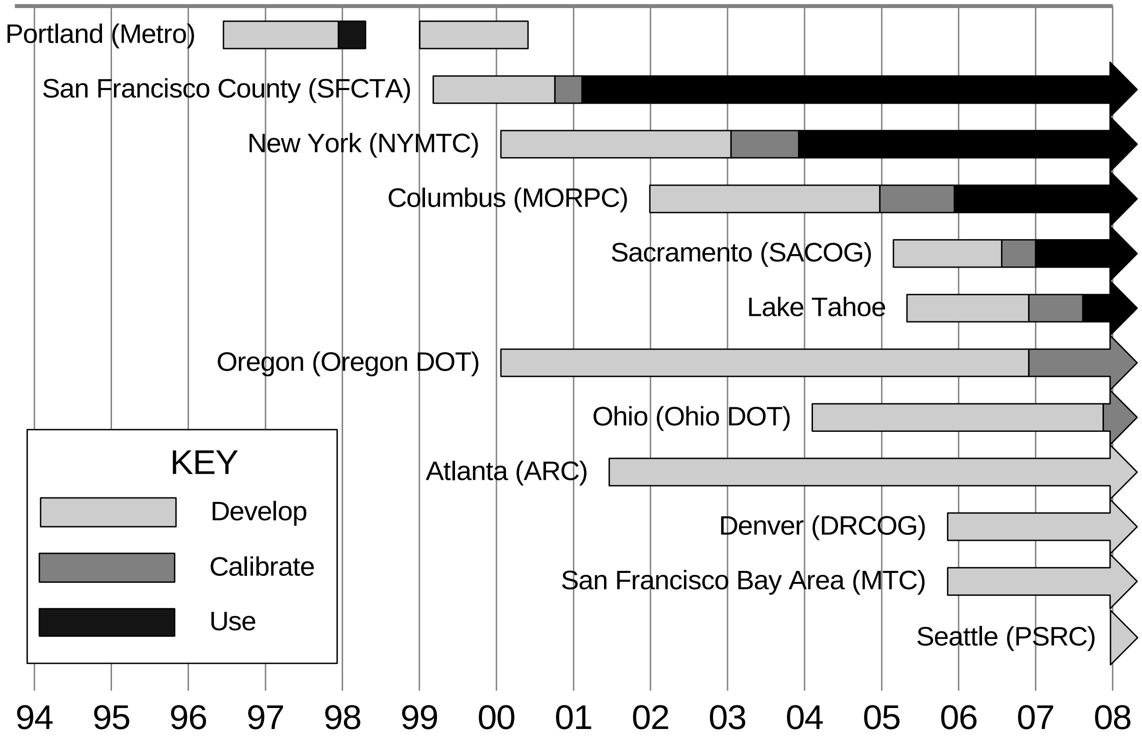
Figure 3: Activity-based model system



Development History

Figure 4 shows a timeline of the development and use of AB models. It includes only US projects, where a sponsor expressed the intent to implement and use the model, and development has started. This section briefly describes the innovations that have been implemented in these projects.

Figure 4: Timeline of Activity Based Model Implementations in the United States



The Metro model was the first to be implemented and used for policy analysis. It was based directly on the Bowman and Ben-Akiva activity schedule approach developed at MIT, using a full-day activity pattern, conditional tour models, and sensitivity at the day level via logsums from the tour models. It introduced work-based subtours, at-home activities and detailed activity purposes, and integrated the AB model with the traffic and transit assignment models.

The San Francisco County model used the same basic design. It was the first of the models to be calibrated, and then used on an ongoing basis for policy analysis. Along the way, innovative procedures were developed for doing that analysis. In a recent major project, the SFCTA model was enhanced to support road pricing, expand its geography, and add mode and temporal detail. It continues to be enhanced.

In New York, a different approach was used for integrating the tour models. Within each household, the simulated tour choices explicitly depended on the purpose of tours already simulated for this and other persons in the household. The NYMTC model has also been used for innovative analyses, some of which would not be possible with a traditional 4-step model.

The Columbus model started with the NYMTC framework and enhanced it substantially, with a strong emphasis on implementing explicit household interactions and detailed time of day modeling.

The Sacramento model also used the Bowman and Ben-Akiva activity schedule approach. It reformulated the day activity pattern, introduced parcel-level spatial resolution, demonstrated the possibility of rapid development and deployment, and used innovative techniques for rapid equilibration of AB microsimulation models.

The Lake Tahoe project was the first implementation for a small MPO, and the first to transfer and recalibrate a model built for another region (MORPC).

The Oregon model was the first activity schedule model to be implemented for an entire state, and it was also integrated into a land use model system.

Ohio imported the Oregon statewide model and enhanced it to include long distance inter-regional trips.

The Atlanta model, which will be based on the MORPC design, hasn’t been fully implemented yet, but they have implemented a flexible population synthesizer, and the design includes other innovations.

DRCOG, MTC and PSRC are the most recent locations where new development projects are under way. PSRC is the first staged implementation in which the first stage involves integrating a day activity pattern model with the existing trip-based model system.

A Comparative Summary of Features

This section provides a concise summary of important design features of various AB model systems that have been implemented or have recently been designed for planning agencies in the U.S. The models described are those mentioned in the previous section, with a few exceptions. Lake Tahoe is excluded because it is essentially a clone of MORPC, and Oregon and Ohio are excluded because they are statewide models. A model for Dallas (CEMDAP) is included because, although no agency has expressed a commitment to use it, it is in an advanced state of development. We have also excluded the FAMOS model for Tampa Bay, which is still primarily a laboratory project, and the TranSIMS model, for which the AB component is not well developed. Finally, we have included the PSRC model system with the Activity Generator incorporated because of its importance to PSRC, even though it remains in part a trip-based model system.

- Except for the PSRC model system, all of the model systems described in this section are similar in several important aspects: represent an entire day of activities and travel for each member of a synthetic population, using stochastic microsimulation
- consist of an integrated system of econometric models
- include traditional traffic and transit assignment components

In addition, the integrated econometric model systems are similar in overall structure, with a hierarchy of levels from "top" to "bottom", where lower level choice predictions are conditioned by those at higher levels, and higher level choices are influenced by accessibility measures that capture the effect of choice opportunities occurring at lower levels. 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: choices that span the entire day for one or more persons in the household
- Tour-level: The main destination, travel mode, begin and end times, and number of stops for each tour
- Trip-level: Intermediate stop location, and the mode and departure time of each trip

Within this structure, there are several important design features and other aspects that distinguish the models, and these are summarized in Table 1 below. At the time of this writing, the Bay Area (MTC) and Denver (DRCOG) models are in the development stage, and the San Francisco model is undergoing a major upgrade, so the design characteristics shown for those models may be in a state of flux. Each paragraph below is a more detailed annotation of a row in the comparison table.

Implementation status: The Metro model was implemented and used once without complete calibration, and subsequently not used. Four of the models, San Francisco (SFCTA), Sacramento (SACOG), New York (NYMTC) and Columbus (MORPC) are in ongoing use, with ongoing maintenance and improvement. The Dallas model (CEMDAP) has been implemented for validation purposes in a laboratory setting. The remaining models are in various stages of development.

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). San Francisco (SFCTA) is also using controls for presence of children, single vs multi-family dwelling, race/ethnicity, and is explicitly synthesizing residents in group quarters housing. 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 explicitly controlled. SFCTA, MTC, DRCOG, MTC and PSRC are all using derivatives of the ARC population synthesizer.

"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 school location should be modeled as a long-term choice and 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 San Francisco model, 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 and 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. None of the other models distinguish between types of in-home activities. Some of the models predict which people work primarily at home, providing some substitution between in-home and out-of-home work. They do 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: All of the models treat linkages across household members implicitly through the use of a wide variety of person type and household composition variables. However, some of them have begun to use explicit linkages between the predicted activities and travel of different members of the same household, which makes microsimulated activity and travel itineraries more consistent among household members. This and the following three paragraphs are concerned with the modeling of these explicit linkages. 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 the fact that the mode is nearly always auto, with the exception of infrequent cases of walk escort. Also, children's 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 of frequencies and duration across household members within a household-day. The Columbus and Atlanta model systems assume that activities for certain purposes are conducted on behalf of the household, and 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, these models seem less crucial than the joint travel models. In addition, it is difficult to reliably determine, from existing surveys, 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. So, without better survey data designed to distinguish activities by whether they achieve household or personal objectives, the quality of models that attempt to allocate household activities is questionable.

Level at which intermediate stop purpose and frequency are modeled: When ordering the models in an AB 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 presence (and, in the case of Portland, basic purpose) of intermediate stops are predicted at the person-day level. 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 presence, 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, another 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 of 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 and purpose of stops on each tour are 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 proven behavioral reason for this structure, but it "balances" the model sensitivities between the two types of behavior described above. A similar approach is being used for Denver (and has been implemented in the initial Activity Generator for PSRC).

Number of network zones used: This and the next two 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 AB 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 be specified at a finer level. The Portland model uses such an approach for roughly 20,000 "block faces", while the Sacramento models use over 700,000 parcels. In both of these model systems, this fine level of disaggregation is used to define the destination choice alternatives and their attractiveness, to provide detailed mode choice information at the trip-ends related to accessibility of transit, and level-of-service for non-motorized modes and intra-zonal trips. Denver is going to use "utility hookup points" instead of parcels, but primarily for modeling mode choice. With these two-level systems, the importance of very small traffic assignment zones is lessened. But the size of zones still needs to be small enough to achieve homogeneity of travel times and costs for the motorized portion of inter-zonal trips, as well as the availability and nature of transit and highway access points.

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 sometimes 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, but has not been use since by any of the implemented model systems.

Modeled time periods and time-constrained scheduling: Most 4-step models only use two times of day—peak and off-peak, and use fixed time-of-day factors. All of the AB 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 highway 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. The effectiveness of modeling time at a detailed level is hampered by the use of no more than four or five time periods for traffic assignment, increasing the pressure to use more time periods for traffic assignment, and to move to dynamic traffic assignment. Denver is implementing 8 assignment time periods, and SFCTA and PSRC have implemented an assignment procedure that takes the equilibrated results of two 3-hour peak period assignments and generates differentiated level-of-service skims for each half hour within the peak.

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. SACOG models time of day below destination and mode. For DRCOG, the data support modeling tour TOD above mode choice for work and school tours, but below mode choice for other tour purposes.

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, origin and destination 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: The issue of how to include accessibility and land use effects in the upper level models is extremely important, because it determines the accuracy with which the models represent sensitivity of activity, tour and trip generation and patterns to transport level of service and the distribution of activity attractions. Calculation of full logsums across all possible nests of lower level alternatives is 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. Initially, the San Francisco model included 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. It has recently been enhanced to use logsum-based accessibility measures. 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. Both DRCOG and PSRC are using aggregate accessibility logsums similar to those used by SACOG.

Network model platform: One attractive aspect of the AB modeling framework used in the existing models is that it is fairly independent of the network modeling software platform. AB models have been implemented to work with Emme2, Cube/TP+ and TransCAD, although three of the four ongoing US AB model systems use Cube/TP+.

AB model software: As the number of AB model implementations has begun to increase, attention to the nature of the software implementation has begun to increase as well. The first implementations, in Portland, San Francisco and NYMTC all employed custom software developed primarily by the model developers themselves, in conjunction with the network modeling platform chosen by the client agency. The software was required for transforming survey data into a form usable for model estimation, and for running the AB components of the implemented model system. Subsequent implementations by those modelers have tended to use and enhance the software from the initial implementations. SACOG and the new PSRC activity generator use software called DaySim, written by Bradley and Bowman in Borland Delphi, a compiled language derived from Pascal that is similar to C++. The SFCTA application uses C++ and Pascal. NYMTC, MORPC, ARC and MTC use Java implementations derived by Parsons Brinckerhoff from their growing common modeling framework (CMF). DRCOG has chosen to develop their own database-oriented custom application with the assistance of a Cambridge Systematics programming team.

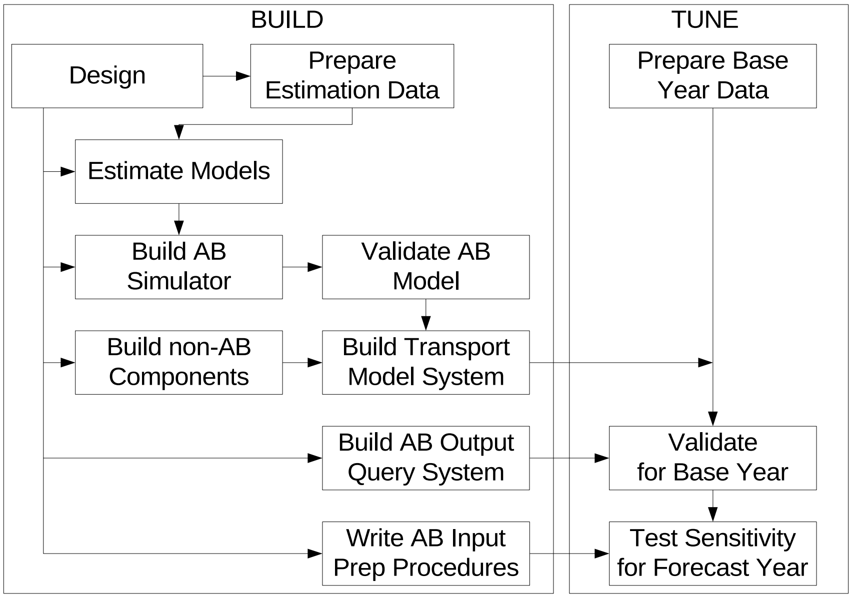
Table 1: Features of Various Activity-Based Model Systems

Feature	Portland Metro	San Francisco SFCTA	Sacramento SACOG	Denver DRCOG	Seattle PSRC with Activity Generator	New York NYMTC	Columbus MORPC	Atlanta ARC	Bay Area MTC	Dallas (CEMDAP)
Implementation Status	Discontinued	In use	In use	Development	Development	In Use	In Use	Development	Development	Laboratory
Controls / # categories for population synthesis	4 hh size 4 income 4 age	4 hh size 3 # workers 4 income age, children	4 hh size 4 # workers 4 income	4 hhsize, 3 # wkrs, 4 inc, #adlts, kids, 3 holder age	similar to ARC	5 hh size 4 # workers 4 income	5 hh size 4 # workers 4 income	100+ comb. of hh size, # wkrs, inc, age, children	4 hh size 4 # workers 4 income Age (?)	
Population synthesizer	Custom	ARC PopSyn	Custom	ARC PopSyn	ARC PopSyn (base year)	Custom	Custom	ARC PopSyn	ARC PopSyn	UT CEMDAP
"Usual" work & school locations at top level?	Yes	Yes	Yes	Yes	Work	No	No	Yes	Yes	Yes
Number of out-of-home activity purposes	8	3	7	7	7	4	7	8	7 or 8	11 for adults 3 for children
Number of in-home activity purposes	3	1	1	1	1	1	1	1	1 or 2	1
Day pattern type linked explicitly across HH?	No	No	No	No	No	No	Sequential	Simultaneous	Simultaneous	Sequential
Joint intra-HH activities?	No	No	No	No	No	No	Yes	Yes	Yes	Parent & child
Linked intra-HH "escort" trips?	No	No	No	No	No	No	No	No	No	Yes
Allocated HH activities?	No	No	No	No	No	No	Yes	Yes	No	Yes
Level where stop purpose and frequency are modeled	Person-day	Person-day	Person-day and tour	Person-day and tour	Person-day and tour	Tour	Tour	Tour	Person-day and tour	Person-day and tour
Network assignment zones	1,250	2,336	1,300	2,800	938	6,000	2,000	2,500	1,600	
Smaller spatial units used?	20K blocks	No	700K parcels	Points for mode choice transit access	No (parcels in next version of models)	No	No	No	No	
Mode and destination model estimation	Simultaneous	Sequential	Sequential	Sequential	Sequential trip-based	Sequential	Sequential	Sequential	Sequential	
Network time periods	5 per day	5 per day (12 ½hr peak subperiods)	4 per day	8 per day	5 per day (12 ½hr peak subperiods)	4 per day	5 per day	4 per day	5 per day	
Modeled time periods	5 per day	30 min (new)	30 min	30 min	30 min	4 per day	1 hour	1 hour	30 min	
Scheduling constrained by available time windows?	No	No	Yes	Yes	No	No	Yes	Yes	Yes	
Tour time of day relative to mode and destination	Above	Above	Below	Above for non-work	Below	Between	Between	Between	Between	
Departure time modeled separately at trip level?	No	Yes (auto peak trips)	Yes, lowest model	Yes, lowest model	Yes (auto peak trips)	No	No	Yes, lowest model	Yes, lowest model	
Accessibility measures in upper level models	Person-specific mode / dest logsums	Jobs reached by zone/ mode/time band,logsums	Mode & dest logsums by zone / segment	Mode & dest logsums by zone / segment	Mode & dest logsums by zone / segment	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	
Network model platform	Emme2	Cube/TP+	Cube/TP+	TransCAD	Emme2	TransCAD	Cube/TP+	Cube/TP+	Cube/TP+	TransCAD
AB model software	DaySim Precursor	Custom C++	DaySim	Custom C#	DaySim	PB CMF	PB CMF	PB CMF	PB CMF	UT CEMDAP

AB Model Development Tasks

Figure 5 identifies the tasks required to develop a new activity-based model, starting from an existing trip-based model. This section describes each of these tasks.

Figure 5: Activity-based model development tasks



Design, the most important task, comes first. If all subsequent tasks are done perfectly, but the design is flawed, then the forecasts will be unreliable, or if the design is limited, then the forecasts may not be useful, or if the design is inflexible, then the model system will be difficult to enhance.

Here we highlight several important aspects of the design. The model system includes several model components on each of several levels. At the top is the population synthesizer. It provides a synthetic population matching the best available estimate of the geo-demographic distribution of the region's population. After that come the models predicting outcomes that are longer term than a day. These are followed by models that predict the one-day pattern of activity for each person, and perhaps for each household. Tour models predict the attributes of each tour in the day, including destination, travel mode and time of day. Finally come the models that predict details of each trip and stop on the tour.

Including a lot of related model components makes it possible to simulate itineraries that appear to be realistic. But in order for them to **behave** realistically, they must be appropriately integrated. A basic integrating feature of the model systems is a model hierarchy. Models lower in the hierarchy need to treat the outcomes of higher models as given. For example, if a person is a worker with a non-home usual work location, then their day is likely to be a work day. If their day is a workday, then a work tour is likely to occur. If there is a work tour, then it is more likely to go to their usual workplace than elsewhere. Non-work tours cannot be made at times of the day that have already been claimed by a work tour, and Intermediate stops are likely to occur near their work destination. In order to properly achieve this downward integration, the design must specify how to do it in three subsequent tasks: in preparation of the survey data for model estimation, in model estimation itself, and in programming of the simulator software. And it must be done consistently in all three steps or the model system won't give reliable forecasts.

Models higher in the hierarchy need to be sensitive to conditions affecting lower level models. This is upward integration. For example, if transit service in a certain corridor is significantly improved for trips occurring during the peak period, then tours by transit at that time of day should increase among persons living and working in that corridor, and their tours in the corridor by other modes should probably decrease. More people will arrange their day to include a transit commute. But non-work auto tours might increase as workers run errands after work that they used to do by car on their work commute, and more non-workers have a car available during the day. Workers in the area may be more likely to buy a transit pass, and more households in the area should eventually own fewer cars. The preferred method of capturing these kinds of affects is through the use of logsums. These measure the expected utility across lower level alternatives, and are used as explanatory variables in the upper level models. A big problem is that Logsum calculations take a lot of time, so compromises are made in order to speed up run times, and creative techniques are developed to reduce the bad impact of such compromises. As with downward integration, the design must specify how to implement upward integration consistently: in preparation of the survey data for model estimation, in model estimation itself, and in programming of the simulator software. Because of the need for consistent models and software, the simulator should be designed during this task.

Finally, the design needs to cover the entire transport model system in which the activity-based simulator operates. Of special importance are two aspects, equilibration and performance. As with a trip-based model system, the demand and assignment must be iterated to achieve consistency in both parts of the system; this is equilibration. Since an activity-based model system can have more computation than its trip-based counterpart, the design needs to include distributed processing of both the demand and assignment portions of the model system.

The last aspect of design that we emphasize here is that it requires major decisions. Although the model systems in use today are strikingly similar, they also have some major differences. These differences need to be evaluated in light of the needs and available resources. Will you model spatial choices using parcels or similar small geography (as Sacramento did), or will you use a more traditional zone-based approach? Will you explicitly model joint decisions and actions of two or more household members (as pioneered in Columbus), or will you rely on implicit modeling of these interactions? And there are other differences among existing model systems that should be considered.

Beyond that, the microsimulation framework is well-suited to innovation. Most of the projects so far have introduced substantial innovations in order to achieve specific objectives, and there are still many advances to be made, in the areas of parking (for parking subsidies, park & ride, and park & walk), vehicles (with vehicle type modeling, vehicle choice for trips, and implementation of vehicle time and space constraints), pricing, and transit (such as pass-holding and subsidies). What specific needs do you have that may require innovative features? And do you want to risk the potential schedule delay and cost that come with innovation?

Having spent considerable time on the design, because of its importance, we will move quickly through the remaining eleven tasks in the development process, starting with the **preparation of data** for model estimation.

In most cases, the design is actually done AFTER the household survey has been designed and at least partially implemented. Nevertheless, after the design is complete, the household survey data must be prepared for model estimation, to represent tours and trips in accordance with the design, to identify the choice set and outcome for each choice, to properly restrict choice sets based on higher level outcomes, and to associate level of service and spatial attributes with each choice alternative. Level of service data will need to come in the form of skims from the existing model system. You may decide to make some enhancements so you can generate skims that are as consistent as possible with the new model design, with regard to the mode definitions and variations in level of service by time of day. For each zone, or parcel if you decide to go that way, you will need to have estimates of employment, school enrollment and housing units in various categories, as well as network attributes such as intersections, transit access, and parking.

Many books, courses and research papers have been devoted to the **estimation of models**, so we will not go into any detail here other than to say that, for each model that is estimated using local data, it must be specified, estimated and tested in an iterative process, using the household survey data.

Building the activity-based simulator is a major task. At the end of the development process, the simulator is the tangible result that takes the form of an executable program and embodies the activity-based portion of the model system. It creates the synthetic population, applies all the models, and constructs a detailed one-day itinerary for each person. In doing this it enforces the hierarchy, time & space constraints, and upward integration prescribed by the design. You can start building the simulator once you have your design. However, by coordinating this task with the prior tasks, the simulator can use code that was written for preparing the estimation data and estimating the models. Re-using this code has the big advantages of saving time and avoiding inconsistencies in the final product.

Once the simulator is complete, it can be used to do a **preliminary validation of the activity-based portion of the model system**. To do this, it is run for the year of the household survey, and its results are compared to aggregate statistics from the survey. It may be necessary to adjust calibration constants in component models, or perhaps to re-estimate one or more of the component models. And it is a natural time to discover and get rid of bugs in the simulator.

Since the activity-based simulator only forecasts trips of residents, it is necessary to **build the other components of the transport model system**. Most of these components probably already exist. During the design you will have decided whether to use them as they are, or to enhance them. However, a program will be needed to aggregate the trips from the activity-based simulator into matrices. And you will probably need to enhance your assignment and skimming procedures, although you might have done that already in order to prepare skims for model estimation.

Once you have a validated activity-based simulator and the other components of the model system, you're ready to put them together into the **integrated transport model system**. This involves installing hardware if necessary, assembling the scripts that run all the components, and testing and tuning the system. Here the focus of attention is convergence and performance. It may be necessary to adjust the iteration scheme and the distribution of processes among machines. In the end, you need to achieve two conflicting objectives simultaneously. Link flows and speeds coming from assignment must match those used to generate the final skims used as input in the demand model components, and the model must complete its job in an acceptable elapsed time. This is a challenge, and none of the existing model systems has fully achieved of these twin objectives.

The last two tasks on the build side of this diagram could be put off until you need them, but there are advantages to planning ahead for them because they're important, and useful for tuning the model system. The first one is a **reporting and query system**, used to extract information from the output of the activity-based simulator. If it is developed soon enough, it can be used to extract model results for base year model system validation. It extracts information from the output of the activity-based simulator and the convergent networks. The objective is to make it easy to aggregate the results as needed, and get them into standard reports, custom queries, and GIS-based display forms.

The last build task is to implement **procedures that generate input data** for forecast scenarios. These will be useful near the end of the project for model sensitivity tests. This task is especially important if you decide to use parcel information in the models. The Sacramento project has shown that generating forecast scenarios at the parcel level, while extremely valuable, can also be a lot of work, and automated procedures are needed to assist with that process.

Now we're ready to move to the model system tuning tasks. First is the **preparation of data for validating** the ability of the model to replicate base year conditions. If the base year for model forecasts is the same as the HH survey year, then some of this will have been done when data was prepared for model estimation. Otherwise, the base year network, spatial and population data must be assembled. And this must be supplemented by fairly standard validation data. One of the main differences may be that, in order to validate the temporal aspects of the model system, the validation data should be as detailed as possible by time of day.

With this data in hand, **base year validation** can occur. If the base year is different than the estimation year, then the activity-based model must be re-validated on the base year. Beyond that, base year validation is much like that for a trip-based model, but validation should be by time of day. If significant discrepancies occur, then calibration constants or re-estimation of models may be necessary.

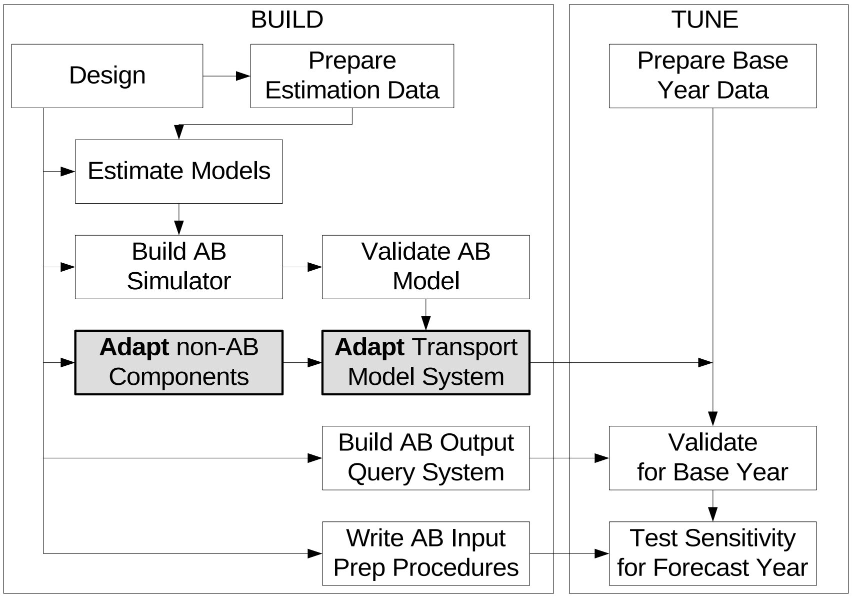
Finally, the model is ready for forecasting, and the final task is to do some **forecast sensitivity tests**. Elasticities can be checked for reasonableness. If problems are encountered, you might need to decide between working out the problems before releasing the model for prime time use, on the one hand, and releasing the model and queuing up a project to improve it, on the other. At this stage, the model can also be used to train users and familiarize clients with its capabilities.

Three Basic Build Approaches: Invent, Adapt, Adopt

In the previous section we described the development process if you are developing a model system from scratch. But there are important variations on this theme that we are calling basic build approaches. We have identified three basic approaches: invent, adapt and adopt. Another way to think about it could be to put invent on one end of a spectrum of approaches, adopt on the other end, with adapt occupying everything in between. Among all the US projects described above, we have classified only Metro, NYMTC, and MORPC in the Invent category (although, in reality, they all incorporated prior work of others.) In contrast, San Francisco, Sacramento, Denver, and now Seattle, are clear adaptations of the early work done in Portland. And later projects in this line have tended to adapt from the most recent adaptation. Atlanta and MTC are implementing adaptations of the MORPC model. And the Tahoe area model is the only one that can be classified as an adoption, taken straight from MORPC.

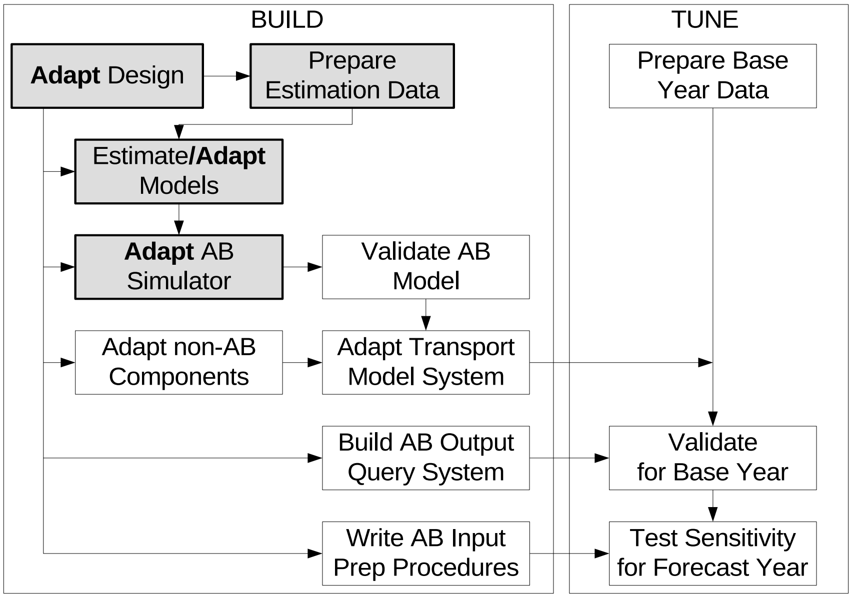
As we are using the term here, inventions carry out all aspects of the development process as described in the previous section. Some of them borrow significant aspects of design from earlier research and models developed in academic settings, but they are not based primarily on any one operational model system. On the other hand, even those I classify as inventions may adapt the non-activity-based components and the integrated model system from an existing trip-based model system, as noted in Figure 6.

Figure 6: Tasks in the Invent Approach



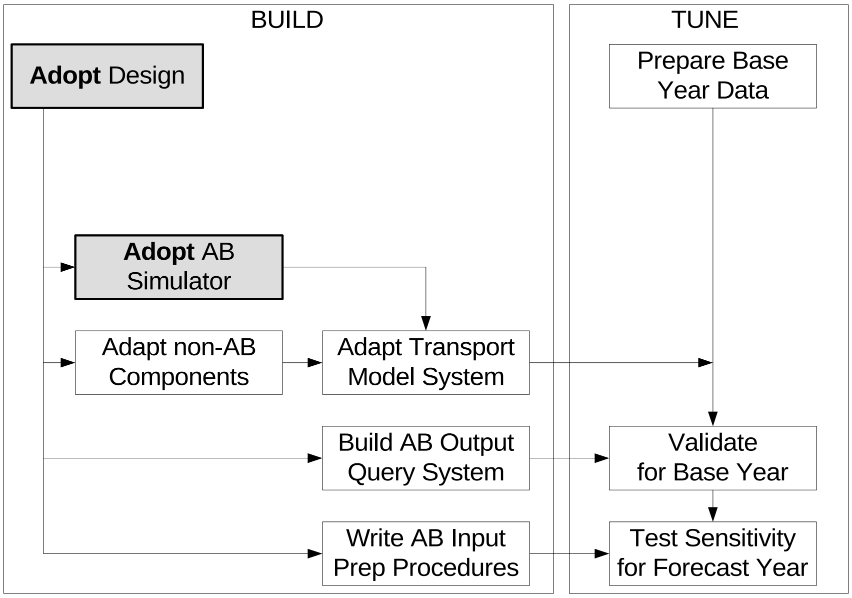
Adaptations (Figure 7) borrow heavily from other model systems for the development of the activity-based simulator. In most cases this includes not only the design, but also the programs used for preparing estimation data, the programs used to estimate the models, and the activity-based simulator program. Adaptations usually do most of all these steps, and usually introduce innovations to achieve important objectives, but they borrow heavily from a prior model system in the process.

Figure 7: Tasks in the Adapt Approach



An adoption, on the other hand (Figure 8), takes an existing simulator, incorporates it into a full model system, and validates it using local data. This bypasses the costly steps of preparing data, estimating models and programming a simulator. It also eliminates the need for a regional household survey.

Figure 8: Tasks in the Adopt Approach



Significantly, nearly all the model systems started in the last several years are adaptations or adoptions. Only Metro has decided to embark on a new design, and this work is just beginning.

Development Roles

In this section we look at the development tasks from the perspective of the roles assigned to various experts. Each activity-based model development project has been unique, including the way that roles have been assigned. However, there seems to be a pattern of assigning responsibilities to four primary roles, with each role tending to have more than one critical area of expertise. These roles include the activity-based developer, the trip-based model expert, the expert (or experts) in GIS, Database and Graphical User Interfaces, and the application expert.

The activity-based developer is responsible for the overall design, the development of the activity-based models, and the simulator. This requires knowledge of activity-based model system design, econometric skills, and programming skills. But because the activity-based model fits into a traditional model framework, an expert in the region's trip-based model system takes primary responsibility for the trip-based components, and, with advice from the activity-based developer, may also take primary responsibility for implementing and tuning the integrated transport model system. For the tasks of dealing with the inputs and outputs required by the model system, one or more persons with extensive skills in GIS, database and graphical user interfaces are needed. And finally, a person with extensive application experience should be heavily involved in model system validation and sensitivity testing. This leads naturally to their ongoing role of Model System Expert from the standpoint of internal users and external clients.

In all of the model systems now in use, and in most of the projects under way, an external consultant filled the role of the activity-based developer, and fulfilled other roles as needed. And in all the completed projects, the developer has had a repeated or ongoing role after implementation.

It is also important to note that DRCOG has deliberately chosen to share the activity-based developer role with the external consultant. DRCOG assisted heavily in the design, estimated some of the models, and augmented the survey data preparation as needed for model estimation. Perhaps most significantly, they chose to be the primary developer of the activity-based simulator and the model system software as a whole.

Insights from the Project Team—Keys to Success

We have reflected from time to time on the various projects that we have been directly involved with or observed from a distance, trying to assess what causes some projects to be more successful than others. From this we have identified three keys to success. Before describing them, we first provide a working definition of a successful project.

We consider the most essential criteria of success that the model system got fully implemented, was implemented consistently with a sound design, and continues to be used for its intended purpose. In addition, the most successful projects experience cost effective and timely development. Cost-effectiveness and timeliness tend to go hand in hand. Projects that string out over a long period of time tend to waste a lot of money. Because existing model systems have not yet tapped all the potential of the approach, successful projects also introduce useful innovation and provide a foundation for ongoing enhancement. Given this definition, we will identify three things that appear to be keys to success: a sound design, capable innovative developers, and sustained sponsorship.

For a sound design the basic framework needs to be workable. Nearly all of the projects started with a basic framework that had been earlier shown to work. Although we called the first Metro project an invention, it was actually based on a working prototype that had been built in the lab at MIT. A workable framework helps assure that a sound result is possible, provides a vision for what is possible, fosters confidence in the sponsor, and gives the developers something to build upon. Every project has some major design decisions to make, and they should be made up front. It is not a good idea to proceed into development without a clear design architecture. The design needs to be comprehensive and integrated. Each of the important component models should be identified, along with an understanding of its basic model form. The methods of integrating all the model components must be sound and clear. This includes the integration among the activity-based model components, and the integration of the activity-based simulator with all the other components of the entire transport model system, including the equilibration of demand models with traffic and transit assignment. And the design must be implemented consistently, in the preparation of data for model estimation, in the model estimation itself, and in the simulator software.

The second key to success is capable innovative developers. Here we're not just talking about the activity-based model developer, but people in all four of the main project roles. This enhances the technical soundness of the product. It also allows the innovations to occur. Although it is now possible to implement an activity-based model system just like one that exists, innovation is still the norm, rather than the exception, and can help achieve important objectives. Paying attention to the development roles of those who deal with the inputs and outputs, and those who run the models, can significantly improve the usability and usefulness of the model system, and assure that the necessary follow through occurs successfully. This follow through includes breaking in new features, seeing the need for and initiating ongoing improvements, and dealing effectively with model users and clients.

The first two keys refer to aspects that were discussed in prior sections. The third key is sustained sponsorship. It was not discussed above, but it is just as important for project success. Without a motivated sponsor, the necessary funding stream won't materialize, or it will dry up and kill the project just when it is needed most. Each successful project has had a motivated sponsor, and each sponsor had its own motivations. For example, SFCTA had specific unmet information needs, and NYMTC needed a model but the region was too complex for a 4-step model. At the beginning, the project needs an instigating advocate to secure sponsorship and get the ball rolling. Most of the projects can point to a specific person who took this role at the outset. Keith Lawton at Metro was the first of these advocates. Gordon Schultz served that role with NYMTC, ARC and MORPC. And as the project progresses, it needs an internal champion to make sure that momentum is maintained through implementation and beyond.

Let us look at two examples that cannot be declared successes, at least not yet. One of these models is no longer being used, and another has been in development for a very long time. First, "Why didn't Metro keep using their model?" The problem had to do with sponsorship. When the time came to calibrate and validate the model, the MPO was struggling financially, and substantial federal funds became available from TranSIMS, so the development money and staff resources were dedicated to the TranSIMS project. As a result, the activity-based model calibration and validation work was never funded. Second, "Why is it taking ARC so long?" Again, the answer is not technical, but has to do with sponsorship. In this case, a cautious ARC top management chose to invest at a very slow rate in the development effort. They also became pre-occupied with a major geographic expansion of their MPO region from 13 to 20 counties. Although the project began in 2001, ARC didn't commit adequate resources to development until 2008. So, among the projects that we are considering, although all three keys are essential, lack of sustained sponsorship has been the factor that has prevented success.

Insights from Agencies Using the Models

In order to identify lessons that the four agencies currently using AB models have learned, we submitted a questionnaire to each one. Here we provide a consolidated and edited list of comments that seem especially relevant to FDOT. The agency providing the insight is noted in parentheses.

If considering a parcel-based model system

1. Determine up front how each parcel-level data item used by the AB model will be forecasted. (SACOG)
2. Try to improve on the parcel-to-parcel proximity techniques used by SACOG, to deal better, for example, with geographic barriers. (SACOG)

Synthetic population input

3. Make sure the synthetic population accurately represents the number and location of university students. (SACOG)

Intra-household interaction

4. Intra-household interaction features are intuitively appealing, and can improve the realism of the itineraries across household members. As implemented by MORPC, they seem not to substantially increase run times. However, users are unable to cite evidence that they improve aggregate predictions. (MORPC)

Traffic and transit assignment

5. Improve highway assignment, and transit networks and assignment to take advantage of AB model capabilities. (SACOG, NYMTC)
    5.1 Use more than four highway assignment periods and two transit assignment periods in order to provide better information for the time-of-day models. (SACOG)
    5.2 If you use PA transit assignment, switch to OD assignment to fit better with the AB model. (SACOG)
    5.3 Design transit assignment and the interface so as to supply the AB model with fares by person-type, and use this information in the mode choice model so that it is sensitive to fare schemes that differ for different people. (SACOG)

Model design and calibration

6. Pay attention to how well the model system distributes trips to zones; this includes the AB model's location choice models as well as the handling of special generators. (MORPC)
7. It has been a challenge to deal with the diversity of the region and the particularities of Manhattan (NYMTC).

Hardware, software, run time and model operation

8. Model run time is a major issue. Use a software and hardware implementation that effectively distributes the AB model and the traffic assignment to multiple processors (The distributed processing in the MORPC AB model is clunky and prone to break down.) Plan for post-implementation work to improve model run times. (MORPC, NYMTC)
9. Extra software products used by the AB model software, such as Excel, ESRI Avenue, and even the assignment software, cause problems because the installation requires these products and because the version used will become out-of-date. (MORPC)

Using the model and model system outputs

10. Include in the development project the design and implementation of databases and data flows for generating desired performance measures that take advantage of the AB model and parcel level data/results. (SACOG)
11. Good model documentation has lagged the implementation (SFCTA, NYMTC).

12. Formal training has been required for stakeholders, including decisionmakers (1 day) and model users (3-5 days and more) (NYMTC).
13. There is a need to improve the user interface, online help and documentation for model users (NYMTC)

Implementation schedule

14. Don't implement during an MTP or other major statutory function. (SACOG)
15. Validate your trip-based model system because the AB model implementation will take longer than the original schedule. (MORPC)
16. Be sure to have specific milestones in the work plan, and a schedule for completion that all parties agree to. Establish a system of "point releases" of code that is workable and testable, which is separate from the constant stream of development updates which exist in an open-source collaborative project like UrbanSim. (SFCTA)
17. Expect it to take twice as long as envisioned (SFCTA).

Maintenance and enhancement

18. It is difficult to make big adjustments to the model without assistance from the consultant who established it. (MORPC)

REQUIREMENTS SPECIFIC TO THE FDOT AND THE TAMPA BAY REGION

The following list of specific requirements and assumptions has been developed pursuant to discussions with Myung-Hak Sung of Gannett-Fleming and Danny Lamb of FDOT. It serves as the foundation of all design and implementation planning work.

1. **Geographic scope.** The model will be implemented specifically for the Tampa Bay Region (FDOT District Seven—the counties of Hillsborough, Pinellas, Pasco, Hernando, and Citrus of Florida) plus a small portion of Manatee County. This is the territory spanned by the current model (TBRPM). However, the mode definitions should be implemented so that the model could be adapted easily to accommodate other regions in the future, if FDOT wants to do so.
2. **Software**. Citilabs CUBE will provide the software platform for the model system. It will be used for traffic and transit assignment, skimming, and trip demand other than activities and travel of residents within the region. An executable program based on the current SACOG household travel demand simulator (DaySim) will be called by CUBE to predict the activities and trips of residents.
3. **Input for activity-based demand models**. DaySim will use parcel-level information as a forecast input, coming either from the new DeltaSim land use model or from land use scenarios. It is desirable to be able to use inputs from the Citilabs DeltaSim land use model.
4. **Traffic and transit assignment.** Traffic and transit assignment will continue to be implemented at the zone level, assigning trip movements between zones. Highway and transit assignment procedures will be enhanced to assign for multiple periods during the day. This number of time periods will probably be four for highway assignment and two or three for transit assignment.
5. **Transit modes**. The model system needs to accommodate many possible transit modes, including local bus, express bus, commuter rail, heavy rail, light rail, bus rapid transit, people mover and jitney. Some of these modes need to be distinguished by mode of access (walk, park and ride, kiss and ride).
6. **Highway pricing**. The model system needs to be able to deal with highway pricing, including HOT lanes and managed lanes.
7. **Implementation approach**. We will plan for a two-stage implementation. The first stage, planned for 2009, will adopt existing activity-based models and software for predicting the activities and travel of residents. It will adapt all other components of the existing TBRPM model system to work with the adopted model. It will be calibrated to the base year 2006 using Tampa Bay regional data, most of which will have already been assembled and used to calibrate TBRPM. The second

stage will adapt and enhance the model system, estimating the activity-based model coefficients from Tampa Bay regional data (including the new NHTS data and other recent surveys), customizing the model system to Tampa Bay region requirements, and perhaps adding enhanced features, such as improved modeling of parking location choice.

RECOMMENDED IMPLEMENTATION APPROACH

Background

A preliminary analysis of the household survey data surfaced three issues that would make it difficult to estimate a full model system from it. One, the sample size (500 households) is quite small. Two, there is no information on car occupancy for car trips, so we can't separate shared ride from drive alone for mode choice. Three, there is no information on transit access model, so we can't separate walk to transit from park and ride.

This led to consideration and acceptance of the two-stage implementation approach listed as requirement seven of the above Requirements Specific to the FDOT and the Tampa Bay Region. In the first stage, we would adopt an existing activity-based model system during 2009, and in the second stage we would adapt and enhance it. The use of the terms 'adopt' and 'adapt' for stages one and two correspond directly to the adopt and adapt build approaches described in more detail in the above description of the State of the Practice. Reasons for choosing this approach include the desire to implement a model system in 2009, and the anticipated availability of NHTS household survey data near the end of 2009 that could greatly enhance the quality of the data available for model estimation.

Choosing an approach that adopts an existing model system and subsequently adapts it using local data requires the selection of an existing model system and the accompanying software in which it was implemented.

Recommendation

We recommend the adoption and subsequent adaptation of the DaySim activity-based models and simulation software that were developed for and are currently used by SACOG.

Rationale

Our recommendation is based on the following reasons:

1. DaySim is the only simulator that runs at the parcel level, enabling it to effectively integrate with the recently developed DeltaSim land use model software.
2. We believe that DaySim is far superior to the other existing model systems in terms of how it integrates the effect of level of service variables from bottom to top. Not only is it more consistent with discrete choice econometric theory, but also because of this it will yield predictions in which the sensitivity of trip generation to travel

conditions are less biased and more realistic. This is described above in the Comparative Summary of Features, in the section entitled "Accessibility measures in the upper level models", and also in a new paper by Bowman and Bradley, a draft of which is included in the Appendices.

3. In terms of "activity-based" versus "tour-based" approaches, DaySim is the only one that provides a balance – the SFCTA approach generates all activities in the day pattern model, and are not sensitive to the lower choices- particularly tour mode and destination. The MORPC approach is more completely "tour-based", and does not have any idea of how many additional stop activities need to be made/scheduled in the day when it does the tour level models. We know reality is always somewhere between the two, and we reflect that in a feasible way. This is described in the Comparative Summary of Features, in the section entitled "Level at which intermediate stop purpose and frequency are modeled".
4. The SACOG DaySim implementation has included extensive work in Cube with equilibration and other features, such as park and ride lot choice that is sensitive to lot capacities. These features, which would be available for the Tampa Bay region implementation, have not been replicated for any of the other model systems.
5. DaySim's modeling of time periods is superior to the other existing model systems. It is more fine-grained (half hour time periods), and its use of time windows to constrain modeled choices is consistently applied for all tours and trips, rather than only for work and school tours.
6. Denver and PSRC will soon be using models based on the DaySim framework. The models we have estimated for them show them to be quite transferable across regions.
7. Potentially desirable features that are not now in DaySim, including perhaps some of the household interaction models used by MORPC and some of the road-pricing evaluation features recently included by SFCTA, could be done as part of the Stage 2 implementation.
8. It would be a lot more expensive for us to implement the structure and code for any of the other model systems due to our lack of familiarity and access. The costs would be at least twice as great, and probably much more.

TECHNICAL MEMORANDUM 2—RECOMMENDED MODEL DESIGN

RECOMMENDED DESIGN FOR PHASE ONE IMPLEMENTATION

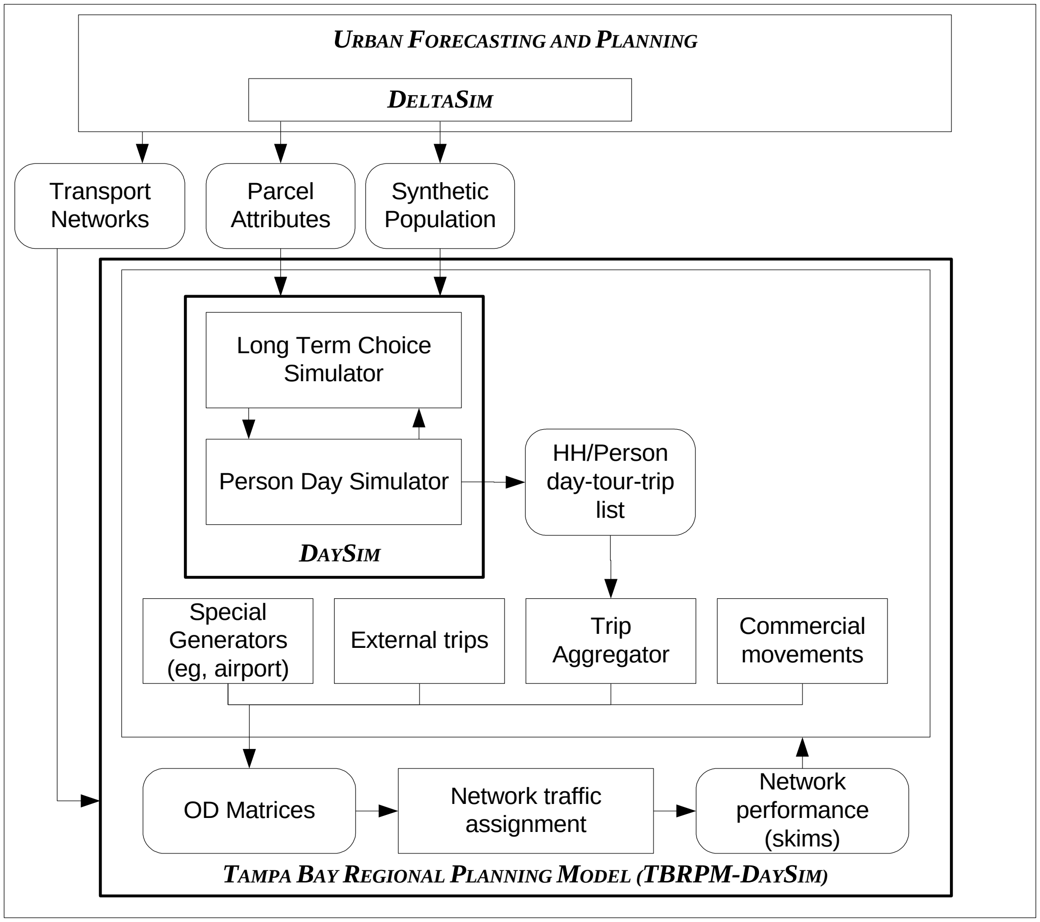
Overview of Design

Unless otherwise stated, this section deals with the design of the model system for the first stage implementation, in which the SACOG activity-based models and software (DaySim) will be adopted, and all other components of the existing model system will be adapted to work with DaySim.

Model System Description

Figure 9 shows the proposed enhanced Tampa Bay Regional Planning Model with an integrated activity-based travel simulator for activities and travel of the region's residents. The figure emphasizes its principal components, their interaction, and the inputs and outputs of that model system. The Long Term Choice Simulator receives a synthetic population from the new DeltaSim land use model, comprised of permanent resident and seasonal households drawn from the region's U.S. Census Public Use Microdata Sample (PUMS) that has been aged and allocated to parcels. It then simulates long-term choices (work location, school location and auto ownership) for all members of the population. The Person Day 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 two components are implemented in a single custom software program (DaySim). DaySim consists of logit and nested logit models connected by adherence to an assumed conditional hierarchy, and by the use of accessibility logsums. The trips predicted by DaySim are aggregated into trip matrices and combined with other trips predicted by the current TBRPM (for visitors, students, group quarters residents, external trips (EI, IE and EE), commercial trucks, taxis, and airport travelers) into time- and mode-specific trip matrices. The network traffic assignment models (highway and transit) load the trips onto the network. Assignment is iteratively equilibrated with DaySim and the other demand models. Since, except for DaySim, this model system is much like the existing TBRPM, for purposes of this report we call it TBRPM-DaySim.

Figure 9: Tampa Bay Regional Planning Model with activity-based simulation of residents' activities and travel (TBRPM-DaySim)



Comparison to the SACOG Model System

The proposed first implementation of TBRPM-DaySim will adopt the SACOG version of DaySim, so it shares most of the SACOG DaySim features:

1. The model uses a microsimulation structure, predicting outcomes for each household and person in order to produce activity/trip records comparable to those from a household survey.
2. The activity-based model works at 4 integrated levels—longer term person and household choices, single day-long activity pattern choices, tour-level choices, and trip-level choices.
3. The upper level models of longer terms decisions and activity/tour generation are sensitive to network accessibility and a variety of land use variables.
4. The model uses 7 purposes for both tours and intermediate stops (work, school, escort, shop, personal business, meal, social/recreation).

5. The model allows the specific work tour destination for the day to differ from the person's usual work location.
6. The model predicts locations down to the single parcel level.
7. The model predicts the time that each trip and activity starts and ends to the nearest 30 minutes, using an internally consistent scheduling structure that is also sensitive to differences in travel times across the day.
8. The accessibility variables used in the upper level models are approximations to a true expected utility structure, with single variables ("logsums") encapsulating differences across different modes, destinations and times of day. Some of these are person-specific, and some of them are "aggregate" (representing important categories of person and household types.)
9. The model accounts explicitly for time-space constraints in destination, mode and time-of-day choice to prevent the prediction of impossible combinations of choices.
10. The software implementation of the activity-based models (DaySim) allows the transport model to be used in several distinct ways, including (a) long-term, where all the long-term and within-day models are run, (b) short-term, where the long-term model outcomes are held fixed, and (c) FHWA New Starts mode, where the models down through tour destination choice are held fixed.

However, there are two aspects of DaySim that might be changed in order to satisfy Tampa Bay requirements:

11. Population synthesis. SACOG DaySim currently uses a custom-built population synthesizer that relies on the input of a joint distribution of households by household size, income and number of workers. Since the preliminary requirement is to use the synthetic population that comes from DeltaSim, the DaySim population synthesizer will probably not be used. However, since a preliminary review indicates that DeltaSim does not control the population for number of workers and income, which are important in DaySim, the recommended design includes implementing a Tampa version of the ARC population synthesizer to generate the base year permanent residents as input to DeltaSim, which DeltaSim would then evolve as it does now. The base year population would be controlled for household size, income, number of workers, presence of children, and age of householder (over/under 65).
12. Transit modes. The Tampa Bay model needs to accommodate more transit modes than does the SACOG model system. There are two possible ways to do this. In the first approach, DaySim would change very little. DaySim currently distinguishes transit submodes in the mode choice model only by mode of access (drive or walk), and produces for assignment purposes a single transit OD matrix for all transit trips for each time period. The Tampa Bay transit assignment procedures would be enhanced and relied on to handle transit submode choice via path selection. This is the recommended approach for the phase 1 implementation, to be kept for the phase 2 implementation if it is judged to be adequate. In the second approach,

DaySim would be enhanced to include many more transit submodes in its mode choice model. DaySim would produce a transit OD matrix for each transit submode, and each one would be assigned separately.

Comparison to the current TBRPM

TBRPM-DaySim retains most of the existing TBRPM features, but there are some significant changes.

1. The biggest difference from TBRPM is that DaySim will predict the trips of the region's year-round and seasonal residents, airport workers, and small trucks owned by households and driven for personal travel. The existing trip generation, distribution and mode split models will be retained, but only applied for trips other than those predicted by DaySim. These include visitors, college students, group quarters residents, commercial truck traffic, taxis, externals (EE, EI and IE) and airport travelers.
2. Highway assignment will be enhanced to assign trips for four time periods in the day (AM peak, midday, PM peak and evening). This is especially valuable because it will provide OD travel times and costs to DaySim that vary by time of day, providing information that enables the DaySim models to do a better job of predicting travel by time of day.
3. Mode choice and transit assignment. The TBRPM transit assignment procedures will be enhanced to assign transit trips for at least two time periods (AM peak and midday). A pre-assignment routine will be implemented that splits each drive-access transit trip into a drive trip and a transit trip, after determining the transfer point (park-and-ride lot or station). If DaySim outputs only one transit OD matrix per time period, then transit assignment would be enhanced as necessary so that it could be relied on to accommodate transit submode choice via path selection (See DaySim comparison item number 12 above). Otherwise, transit assignment would be enhanced to assign trips separately for any additional transit submodes for which DaySim provides OD matrices.
4. Skims. Skimming procedures will be adjusted to provide skims files and variables by time of day for each assignment time period, as required by DaySim.
5. CUBE user interface. The CUBE user interface needs to be modified to accommodate DaySim and the changes to the other components of TBRPM.

AB Models

Model Structure

Table 2 lists all the model components of DaySim, in order corresponding to a conditional hierarchy. Models lower in the hierarchy are conditioned by the outcomes of models higher in the hierarchy. Most of the models simulate multiple outcomes for a given household, and the simulation sequence proceeds as follows:

- All person-level outcomes are modeled one person at a time, in a fixed priority sequence
- Person-day outcomes are modeled one person at a time, in a fixed priority sequence
- All model outcomes above the tour level are simulated before any tour-level models are run.
- All tour-level and trip level outcomes are modeled one tour at a time, in a fixed priority sequence.
- Time window availability is restricted as tours and trips are simulated.

Within the above framework, the models simulate outcomes higher in the hierarchy first, so that their outcomes can be used to condition the models lower in the hierarchy.

Table 2. Component Models of DaySim

Model #	Model Name	Level	What is predicted
	Long-term models		
1.1	Usual Workplace Location	Worker	Workplace location zone and parcel
1.2	Usual School Location	Student	School location zone and parcel
1.3	Auto availability	Household	Number of vehicles available for use by household members
	Day-level models		
2.1	Day 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.
	Tour-level models		
3.1	Tour Primary Destination	(Sub)Tour	Primary destination zone and parcel (models are purpose-specific)
3.2	Work-Based Tours	Work Tour	Number and purpose of any subtours made during a work tour
3.3	Tour Main Mode	(Sub)Tour	Main tour mode (models are purpose-specific)
3.4	Tour Time of Day	(Sub)Tour	The time period arriving and the time period leaving primary destination (models are purpose-specific)
	Trip/stop-level models		
4.1	Stop frequency and purpose	Half Tour	Number and activity purpose of any intermediate stops made on the half tour, conditional on day pattern
4.2	Intermediate Stop Location	Trip	Destination zone and parcel of each intermediate stop, conditional on tour origin, destination, and location of any previous stops
4.3	Trip Mode	Trip	Trip mode, conditional on main tour mode
4.4	Trip Departure Time	Trip	Departure time within 30 min. periods, conditional on time windows remaining from previous choices

Long-term models

Usual workplace location (1.1) and Usual school location (1.2)

The dependent variable in the usual location models is the parcel address where the activity takes place. Since several hundred thousand parcels comprise the universal set of location choice alternatives, it is necessary to both estimate and apply the location choice models using a sample of alternatives. The sampling of alternatives is done using two-stage importance sampling with replacement; first a TAZ is drawn according to a probability determined by its size and impedance, and then a parcel is drawn within the TAZ, with a size-based probability.

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 university and grade school students who also work, the usual school location is known when usual work location is modeled; for other workers who also go to school, the work location is known when usual school location is modeled.

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. As a result, both of these models take the nested logit form, with all non-home locations nested together under the conditioning choice between home and non-home. In the estimation data, all workers have a usual work location and all students have a usual school location, so the model does not have an alternative called "no usual location".

Two important variables in the 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 some cases this is captured through a logarithmic form of distance. In other cases, where there is plenty of data to support a larger number of estimated parameters, a piecewise linear form is used to more accurately capture this nonlinear effect.

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 TAZ-level density variables affect location choice.

Auto availability (1.3)

This model is applied at the household level, and determines the number of vehicles available to the household drivers. Key variables are the numbers of working adults, non-working adults, students of driving age, children below driving age, income, auto and non-auto accessibilities to work and school locations, and more general pedestrian, transit and auto accessibility to retail and service locations.

Day-level models

Day activity pattern (2.1) and Exact number of tours (2.2)

This model jointly predicts 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 seven purposes are work, school, escort, personal business, shopping, meal and social/recreational. 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. The significant variables 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. An interesting result is that, 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.

Tour-level models

Within each tour, three main models are used, to first simulate the tour's destination, then the main mode used for the tour, and finally the beginning and ending period of the tour's primary activity. For work tours, the number of work-based subtours is also modeled, after destination choice, and before timing and travel mode.

Tour primary destination (3.1)

These models determine the primary destination parcel for home-based tours and work-based subtours. The tour destination choice model is very similar to the long-term usual location choice models described above. However, there are some important differences:

For the tour destinations, auto ownership levels are treated as given, and affect location choice.

For tour destinations, all usual locations are known.

Because a large majority of work tours go to the usual work location, the work tour destination model has this as a special alternative. Therefore, the model is nested, with all locations other than the usual location nested together under the conditioning binary choice between usual and non-usual. (Nearly all observed school tours go to the usual school location. Therefore, there is no school tour destination choice model.)

Since there are no modeled usual locations for activities other than work and school, the destination choice model of all remaining purposes is simply a multinomial logit model.

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 half-tour stop frequency and purpose model described below.

Tour main mode (3.3)

The tour mode choice model determines the main mode for each tour (a small percentage of tours are multi-modal), There are eight modes, although some of them are only available for specific purposes. They are listed below along with the availability rules, in the same priority order as is used to determine the main mode of a multi-mode tour:

(1) DT- Drive to Transit: Available only in the Home-based Work model, for tours with a valid drive to transit path in both the outbound and return observed tour

(2) WT- Walk to Transit: Available in all models except for Home-based Escort, for tours with a valid walk to transit path in both the outbound and return observed tour periods.

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

(4) S3- Shared Ride 3+: Available in all models, for all tours.

(5) S2- Shared Ride 2: Available in all models, for all tours.

(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.

In general, it was possible to obtain significant coefficients for out-of-vehicle times, but not for travel costs or in-vehicle times. This is a typical result for RP data sets, particularly when there are few transit observations. As a result, many of the coefficients for cost and in-vehicle time were constrained at values that met the following criteria: (1) the in-vehicle time coefficients meet the United States Federal Transit Administration (FTA) guidelines, (2) the imputed values of time are reasonable and meet FTA guidelines, and (3) the values were kept as close as possible to what the initial estimation indicated. The resulting values of time and out-of-vehicle/in-vehicle time ratios are shown in Table 3. The number of transfers was not found to be significant in any of the models, however transfer wait time is included in the out-of-vehicle time coefficients. Also, the higher the percentage of time in a Drive to Transit path that is spent in the car rather than on transit, the lower the probability of choosing it. This is a result often found in other cities as well, which serves to discourage park-and-ride choices that include long drives followed by short transit rides.

Table 3: 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

Two land use variables came out as significant in many of the models, increasing 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.

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.

A number of different nesting structures were tested. In the nesting structure that was selected there are three combined nests:

(1) Drive to Transit with Walk to Transit

(2) Shared Ride 2 with Shared Ride 3+

(3) Bike with Walk

These all gave logsum coefficients less than 1.0 but not significantly different from each other, so a single estimated nesting parameter applies to all 3 nests (as well as to the 2 additional "nests" that only have one alternative each: Drive Alone, and School Bus). The estimated logsum parameters are 0.51 for Work, 0.86 for School, and 0.73 for Other. For Work-Based tours, it was not possible to obtain a stable estimate, so a constrained value of 0.75 (similar to HBOther) was used. No nesting was used for the Escort model, as it contains only 3 alternatives and is a very simple model.

Tour time of day (3.4)

DaySim employs a method of modeling time of day developed by Vovsha and Bradley (2004). 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. It thereby also provides an approximate duration of the activity at each activity location. The model uses 48 half-hour periods in the day—3:00-3:29 AM, 3:30-3:59 AM, …, 2:30 AM-2:59 AM. Given the way that the activity diary data was collected, no tour begins before 3:00 AM or ends after 2:59 AM.

For each home-based or work-based tour, the model predicts the time that the person arrives at the tour primary destination, and the time that the person leaves that destination to begin the return half-tour. The tour model includes as alternatives every possible combination of the 48 alternatives, or 48 x 49 / 2 = 1,716 possible alternatives. The model is applied after the tour primary destination and main mode have already been predicted. 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. The time period of a work-based subtour is constrained to be within the time period of its parent tour.

A key concept in the time of day models is the "time window". A time window is a set of contiguous time periods that are available for scheduling tours and stops. When a tour or stop is scheduled, the portion of the window that it does not fill is left as two separate and smaller time windows. The time periods at either end of a scheduled sequence of activities on a tour are only partially filled, but the time periods in between are completely filled. It is possible to arrive at a tour or stop destination in a given time period if another tour ended in that period, and possible to leave a tour or stop destination if another tour began in that period, but it is not possible to arrive or depart in a time period that is already completely filled.

Another key aspect is the use of shift variables. These are dummy variables interacted with the arrival time and the duration of the alternative. If the arrival shift coefficient is negative, it means that activities tend to be made earlier (because the shift coefficient causes later arrival time alternatives to have lower utility), and if it is positive, it means that activities tend to be made later. If the duration shift coefficient is negative, it means that activities tend to be shorter (because the shift coefficient causes longer duration time alternatives to have lower utility), and if it is positive, activities tend to be longer. No departure shift coefficient is estimated because the departure shift is simply the sum of the arrival shift and the duration shift (e.g. if the arrival shift is an hour earlier and the duration shift is an hour longer, the departure shift is 0). In the models, shift variables interact extensively with other characteristics of the person, day activity pattern and tour, as well as time-dependent attributes of the network, such as travel times and measures of congestion, to effectively represent their influence on time-of-day choice.

The time of day models also use a variety of variables to represent scheduling pressure, conditional on what other activities have already been scheduled or remain to be scheduled for the day. The overall scheduling pressure is given by the number of tours remaining to be scheduled divided by the total empty window that would remain if an alternative is chosen. The negative effect indicates that people are less likely to choose schedule alternatives that would leave them with much to schedule and little time to schedule it in. A similar variable is the number of tours remaining divided by the maximum consecutive time window. This is also negative, meaning that people with more tours to schedule will tend to try to leave a large consecutive block of time rather than two or more smaller blocks.

Relative travel times across the day also influence time of day choice. The travel time for each period is based on the network travel times for the 4 periods of the day – AM peak, midday, PM peak, and off-peak. The variable is applied for both the outbound half tour (tour origin to tour destination) and the return half (tour destination to tour origin). For auto tours, the time is just the in-vehicle time, while transit time is in-vehicle time plus first wait time, transfer time, and drive access time. Walk access/egress time is not included, as that does not vary by time period. These variables are not applied for walk, bike or school bus tours. Significant travel time effects were found for Work and Other tours and for Intermediate Stops, but not for School or Work-based Tours.

Auto congestion may also cause time shifts within the AM peak and PM periods. For this purpose, the variable used was the extra time spent on links where the congested time is over 20% higher than the free flow time. This extra congested time was converted to shift variables by multiplying by the time difference between the period and the "peak of the peak":

1. AM shift earlier: If the period is 6 AM to 8 AM, multiply by (8 AM – time)
2. AM shift later: If the period is 8 AM to 10 AM, multiply by (time – 8 AM)
3. PM shift earlier: If the period is 3 PM to 5 PM, multiply by (5 PM – time)
4. PM shift later: If the period is 5 PM to 7 PM, multiply by (time – 5 PM)

With this formulation, the more positive the coefficient and the larger the congested time, the more that the peak demand is spread away from the peak of the peak.

For Work tours, in both the AM and PM, the estimation results show a tendency to move the work activity earlier as the time in very congested conditions increases. For School tours and Work-based subtours, no significant congestion effects were estimated. For Other tours, times in the PM peak were found to shift both earlier and later with high congestion.

Trip/stop level models

Although the presence of extra (intermediate) stops in the day pattern is determined in the pattern model, the exact number of stops for each purpose is a result of the stop level models. Within each half-tour, the stops are modeled one-by-one, first for stops before the tour destination, and then for stops after the tour destination.

Stops before the tour destination are modeled in reverse temporal sequence. First the number, purpose and sequence of stops is modeled. Then, for the last stop before the tour's destination, its location (4.2) and departing trip mode (4.3) are modeled. These results also determine the time period in which the trip from the stop location begins, since the trip mode and travel level of service are known. Next the time period of arrival at the stop location is modeled (4.4). This continues, constructing the trip chain from the tour primary destination to the tour origin in reverse chronological sequence until the trip from the tour origin to the first stop is modeled. The reason for modeling in reverse chronological sequence for the first half tour is the hypothesis that people aim to arrive at the primary destination at a particular time, and adjust their tour departure time so as to enable completion of the desired intermediate stops. After the trip chain for the first half-tour is modeled, the trip chain for the second half-tour back to the tour origin is similarly modeled, but this time in regular chronological order.

Half-tour stop frequency 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 intermediate stop activity purposes prescribed by the activity pattern model that have not yet been accomplished on other tours.

The estimation results for this model 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 we do 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.

Intermediate stop location (4.2)

For intermediate stop locations, the main mode used for the tour is already known, and so are the stop location immediately toward the tour destination (**stop origin),** and the **tour origin**. 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).

As with tour destinations, a sampling procedure is required for the stop location models, and a procedure is used that employs importance sampling with replacement. The exact procedure is different, however, because the sampling problem is more complex. For intermediate stops, the travel impedance affecting choice is a function of three locations instead of two: the intermediate stop location, as well as locations before it and after it in the half tour. This expands the number of relevant impedances geometrically. Therefore, a 3-stage importance sampling procedure is used. For each parcel to be drawn, first a stratum is drawn, then a TAZ within the stratum, and finally a parcel within the TAZ.

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. It combines all travel cost and time components according to assumptions about their relative values. Generalized time is used, instead of various separately estimated time and cost coefficients, because the intermediate stop data is not robust enough to support good estimates of the relative values. Generalized time is measured as the (generalized) time required to travel from stop origin to stop location and on to tour origin, minus the time required to travel directly from stop origin to tour origin. It is further modified by discounting it according to the distance between the stop origin and the tour origin. The discounting is based on the hypothesis that people are more willing to make longer detours for intermediate stops on long tours than they are on short tours.

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.

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. The most common occurrence of this is a Drive Alone trip that is made as part of a Shared Ride tour after the passenger has been picked up or dropped off. These cases are most common on Escort tours, where predicting the trip(s) that is Drive Alone is mainly a function of the half tour (away from home or towards home) and the time of day.

Trip 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 primary 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 alternatives (the same 30 minute periods that are used in the tour models). 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.

The use of time windows and variables for trip departure and arrival times is like those described for tour time of day above.

Accessibility variables

Accessibility measures are discussed separately in this memo for two reasons. First, they are very important because in a hierarchical model system, they capture the sensitivity of activity and travel decisions modeled in higher levels of the model hierarchy to the utility of opportunities associated with conditional (and hence undetermined) lower level model outcomes. In formal nested logit hierarchies the upward integrity comes from the logsum, the composite measure of expected utility across the lower level alternatives. For example, in a destination choice model, a logsum variable can capture the expected utility of the available travel mode alternatives. This is a very important aspect of model integration, and can be called upward vertical integration. Without it, the model system will not effectively capture sensitivity to travel conditions. Second, when there are very many alternatives (millions in the case of the entire day activity schedule model), the most preferred measure of accessibility, the expected utility logsum, requires an infeasibly large amount of computation.

DaySim uses two techniques to supplement the selective use of true logsums, in an effort to achieve upward vertical integrity with a feasible amount of computation. The basic idea of the first technique is to avoid the use of a time of day logsum when applying an upper level model by treating as given a yet un-modeled time of day. The assumed time of day is selected by a Monte Carlo draw using observed time-of-day distributions. Rather than making every simulated outcome sensitive to variability in level of service by time of day, sensitivity is achieved across the population through the variability of outcome in the Monte Carlo draws. In this way, the upper level choice models are sensitive to variations in transport level of service and spatial attributes across all possible combinations of time of day and mode, with the effects approximately weighted by the joint time of day and mode choice probabilities.

The basic idea of the second technique is to calculate an approximate, or aggregate, logsum. It is calculated in the same basic way as a true logsum, by calculating the utility of multiple alternatives, and then taking expectation across the alternatives by calculating the log of the sum of the exponentiated utilities. However, the amount of computation is reduced, either by ignoring some differences among decisionmakers, or by calculating utility for 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. The categories of decisionmakers and the aggregation of alternatives are chosen so that in all choice cases an approximate logsum is available that closely approximates the true logsum. In essence, this is a sophisticated ad hoc measure that is intended to achieve most of the realism of the true logsum at a small fraction of the cost. Two kinds of approximate logsums are used, an approximate tour mode-destination choice logsum and an approximate intermediate stop location choice logsum.

The approximate tour mode-destination choice logsum is used in situations where information is needed about accessibility to activity opportunities in all surrounding locations by all available transport modes at all times of day. Because of the large amount of computation required for calculating a true logsum for all feasible combinations in these three dimensions, an approximate logsum is used with several simplifications. First, it ignores socio-demographic characteristics, except for car availability and driving age, yielding four socioeconomic classifications (under age 16, household without a car, household with less cars than drivers, household with 1+ cars per driver). Second, it uses three aggregate distance bands for transit walk access (short access, long access, walk access unavailable). Third, sometimes it uses a logsum for a composite or most likely purpose instead of calculating it across a full set of specific purposes; distinct logsums are generated for the following categories: escort, personal business, shop, meal, social visit, recreation, work-based subtour, and composite nonwork purpose. Finally, instead of basing the logsum on the exact available time window of the choice situation, and calculating it across all of the available time period combinations within the window, it uses an assumed available time window size and time period combination that is most likely for the particular logsum being calculated. With these simplifications, it is possible to pre-calculate 96 logsums for each TAZ, and use them when needed at any point in the simulation of any person’s day activity schedule. These logsums are generated using simplified versions of the tour mode and destination choice model, estimated using only the explanatory variables that distinguish the 96 logsum categories.

The approximate intermediate stop location choice logsums are used in the activity pattern model, tour destination choice, and other models where the choice may be influenced by the convenience of access for intermediate stops on the way to or from the tour destination. Four logsums are calculated for each OD zone pair, distinguished by tour mode (transit or auto) and time of day (peak or offpeak). Each logsum is calculated across all possible intermediate stop zones, each stop's utility is a function of travel (detour) time and zonal attractiveness, and zonal attractiveness is a function of employment and school enrollment, taken from an estimated purpose-non-specific intermediate stop location choice model (simplified version of model 4.3).

Table 4 shows a list of the models in the DaySim model hierarchy, identifying the measures of accessibility used by each one (in addition to direct measures, such as travel time and cost), including true logsums, approximate logsums and simulated outcomes.

Table 4: Measurement of accessibility (impedance and spatial attribute effects) in the model hierarchy

	Model	**Tour mode choice logsum**	**Simulated conditional outcomes**	**Aggregate tour mode-destination choice logsum**	**Aggregate intermediate stop location choice logsum**
1.1	Usual Work Location	Yes.		At destination.	
1.2	School Location	Yes.		At destination.	
1.3	HH Auto Availability	To work. To school.		At home.	
2.1	Day Activity Pattern	For work & school.		At home.	Yes.
2.2	Number of Tours (by purpose)	For work and school tours.		At home.	
3.1	Tour Destination	Yes.	Primary activity periods	At destination.	
3.2	Number & Purpose of Work-Based Subtours				
3.3	Tour Mode		Primary activity periods		
3.4	Tour Destination Arrival and Departure Times				
4.1	Half-Tour Stop Frequency & Purpose				For auto-based tour modes.
4.2	Intermediate Stop Location				
4.3	Trip Mode				
4.4	Intermediate Stop Departure Time				

DaySim Outputs—Person-day, Tour and Trip Files

The main output of the activity-based model consists of three files: person-day, tour and trip. The person-day file includes one record for the simulated day of each person, consisting primarily of the number and purpose of tours and stops in the person's day. The tour file includes one record for each tour in the person's day, including work-based tours, and the trip file includes one record for each trip in each tour.

Table 5: Person-day file (from DaySim)

Label	**Definition**
SAMPN	Household ID (same as input SAMPNO)
PERSN	Person sequence number within HH (same as input PNUM)
HHTAZ	Residence zone (same as input HZONE)
HHCEL	Residence parcel ID (same as input HPARCEL)
HHSIZE	# persons in the household (same as input PERSONS)
HHCARS	# vehicles in the household – predicted
UWTAZ	Usual work zone – predicted
UWCEL	Usual work parcel – predicted
USTAZ	Usual school zone – predicted
USCEL	Usual school parcel – predicted
NTOURS1	Number of work tours – predicted
NTOURS2	Number of school tours – predicted
NTOURS3	Number of escort tours – predicted
NTOURS4	Number of personal business tours – predicted
NTOURS5	Number of shopping tours – predicted
NTOURS6	Number of meal tours – predicted
NTOURS7	Number of social/recreation tours – predicted
NSTOPS1	Number of work stops – predicted
NSTOPS2	Number of school stops – predicted
NSTOPS3	Number of escort stops – predicted
NSTOPS4	Number of personal business stops – predicted
NSTOPS5	Number of shopping stops – predicted
NSTOPS6	Number of meal stops – predicted
NSTOPS7	Number of social/recreation stops – predicted
WBTOURS	Number of work-based subtours – predicted
EXPFAC	Expansion factor (same as EXFAC x subsample rate)
WORKER	Worker dummy variable
PERSTYPE	Person type code: 1 = full time worker, 2 = part time worker, 3 = non-worker age 65+, 4 = other non-worker/non-student adult, 5 = university student, 6 = grade school student age 16+, 7 = child age 5-15, 8 = child age 0-4
HHINCOME	Household income ($)
HHWORKERS	Household # workers

Table 6: Tour file (from DaySim)

Label	**Definition**
SAMPN	Household ID (same as input SAMPNO)
PERSN	Person sequence number within HH (same as input PNUM)
TOURNO	Tour sequence number within person day
TOURPURP	Tour purpose (1 to 7): codes 1 'work' 2 'school' 3 'escort' 4 'personal bus' 5 'shopping' 6 'meal' 7 'social/recreation'
PRNTTOUR	Work-based subtour "parent" work tour ID (0 for home-based)
PDTAZ	Tour primary destination zone – predicted
PDCEL	Tour primary destination parcel – predicted
TIMARRPD	Tour primary destination arrival time (HHMM) – predicted
TIMDEPPD	Tour primary destination departure time (HHMM) – predicted
MAINMODE	Tour main mode – predicted: codes 1 'drive-transit-walk' 2 'walk-transit-drive (NA to tours)' 3 'walk- transit-walk' 4 'school bus' 5 'shared ride 3+' 6 'shared ride 2' 7 'drive alone' 8 'bike' 9 'walk'
TRIPSH1	Tour # of trips in first half tour – predicted
TRIPSH2	Tour # of trips in second half tour – predicted
SUBTOURS	Tour # of subtours – predicted
EXPFAC	Expansion factor (same as EXFAC x subsample rate)

Table 7: Trip file (from DaySim)

Label	**Definition**
SAMPN	Household ID (same as input SAMPNO)
PERSN	Person sequence number within HH (same as input PNUM)
TOURNO	Tour sequence number within person day
TOURHALF	Tour half (1=outbound, 2=return)
TRIPNO	Trip sequence number within half-tour
OTAZ	Trip origin zone – predicted
OCEL	Trip origin parcel – predicted
DTAZ	Trip destination zone – predicted
DCEL	Trip destination parcel – predicted
MODE	Trip mode – predicted: codes 1 'drive-transit-walk' 2 'walk-transit-drive (NA to tours)' 3 'walk- transit-walk' 4 'school bus' 5 'shared ride 3+' 6 'shared ride 2' 7 'drive alone' 8 'bike' 9 'walk'
OPURP	Trip origin activity purpose (1-7 as above, or 8=home)
DPURP	Trip destination activity purpose (1-7 as above, or 8=home)
DEPTIME	Trip departure time – predicted (HHMM)
ARRTIME	Trip arrival time – predicted (HHMM)
TRAVTIME	Trip door-to-door travel time (min)
TRAVDIST	Trip door-to-door travel distance (mi)
EXPFACT	Expansion factor (same as EXFAC x subsample rate)

Interface between DeltaSim and TBRPM-DaySim

The DeltaSim model generates base year conditions and evolves them on an annual basis from the base year through the entire forecast time horizon. Currently it is set up to provide input for TBRPM, and receive back information about travel conditions, every five years. The preferred design calls for TBRPM-DaySim to take the place of TBRPM in the DeltaSim evolution. As DeltaSim evolves annually, it provides a synthetic population and a file of parcel attributes to TBRPM-DaySim every five years, and receives back the needed information about travel conditions.

Synthetic Population from DeltaSim

DeltaSim needs to supply DaySim with a synthetic population, one record per person, including all permanent residents of households (size 1 or greater) and all seasonal residents. Table 8 lists the data items needed by DaySim, using the labels of the corresponding PUMS data, where applicable, from which DeltaSim draws the sample. The file should be ordered by SERIALNO and then by PNUM within household. The number of person records within any household must be equal to the value of PERSONS. The values of SERIALNO, PERSONS, UNITTYPE, HINC, EXFAC, VEHICL, HTAZ, and HPARCEL should be identical for all persons within a household. DBase IV file format is preferred.

Table 8: Included synthetic sample variables

Label	Definition	Issues
SERIALNO	Household or GQ ID (called SAMPNO in DaySim)	
PNUM	Person sequence number within HH	
PERSONS	# persons in the household	
UNITTYPE	type of unit (1-housing, 2-institutional GQ, 3-noninstitutional GQ)	
NPF	number of persons in family	DeltaSim does not evolve this in its synthetic population. DaySim uses this only to determine whether household is a family household or a non-family household.
HINC	Household annual income ($2000)	DeltaSim does not evolve this in its synthetic population. DaySim currently uses income defined in year 2000 dollars, in 16 categories with thresholds at 5000, 10000, 15000, 20000, 25000, 30000, 35000, 40000, 45000, 50000, 60000, 65000, 75000, 100000, and 150000.
SEX	Gender (1=male, 2=female)	
AGE	Age (years)	
STUDENT	Person is a student (0-no/1-yes)	DeltaSim does not evolve this in its synthetic population.
GRADE	Grade/level in school (PUMS coding)	DeltaSim does not evolve this in its synthetic population. DaySim uses this simply to distinguish college students from grade school students.
WORKER	Person is a worker (0-no/1-yes)	DeltaSim does not evolve this in its synthetic population.
HOURS	Hours worked per week	DeltaSim does not evolve this in its synthetic population. DaySim uses this simply to distinguish part-time workers from full-time (>-32 hrs/wk)
EXFAC	Expansion factor	
VEHICL	PUMS # of vehicles in the household	DaySim does not use this, but it can be retained and used for testing purposes.
HTAZ	Residence zone	
HPARCEL	Residence parcel ID	
SEASONAL	Household is a seasonal resident (0-no/1-yes)	

Parcel Attributes from DeltaSim

The following table shows the complete list of parcel variables input by the SACOG DaySim model that would come from DeltaSim. Most of the variables occur in three versions: the count within the parcel, and the count within two buffer zones surrounding the parcel centroid (quarter-mile and half-mile). DaySim does not actually use the variables in italicized light print, although many of them were considered for use during model estimation, and might be important for the Tampa region when the DaySim models are estimated using local survey data in the second implementation.

Table 9: Parcel Attributes used in DaySim Model (to be supplied by DeltaSim)

Label	Definition	Issues
PARCELID	Parcel ID number	
X_COORD	X coordinate – state plane feet	
Y_COORD	Y coordinate – state plane feet	
AREA_SQF	*Area – square feet*	

TAZ	TAZ number	
HOUSESP	Housing units – parcel (x 100)	
HOUSESQ	*Housing units – quarter mile radius (x 100)*	
HOUSESH	Housing units – half mile radius (x 100)	
STUDK12P	Students K-12- parcel (x 100)	
STUDK12Q	Students K-12– quarter mile radius (x 100)	
STUDK12H	*Students K-12– half mile radius (x 100)*	
STUDUNIP	Students University– parcel (x 100)	
STUDUNIQ	*Students University – quart. mile radius (x 100)*	
STUDUNIH	*Students University – half mile radius (x 100)*	
NODES1Q	1 link nodes– quarter mile radius	The nodes should be counted using a road network that has all streets, rather than only those used for traffic assignment.
NODES1H	1 link nodes– half mile radius	
NODES3Q	3 link nodes– quarter mile radius	
NODES3H	3 link nodes– half mile radius	
NODES4Q	4+ link nodes– quarter mile radius	
NODES4H	4+ link nodes– half mile radius	
DIST_LRT	Distance to nearest LRT stop (miles x 100 -1 if none)	Should be nearest rail stop of any type (LRT or otherwise)
DIST_BUS	Distance to nearest bus stop (miles x 100, -1 if none)	
PARKDY_P	Daily paid parking spaces- parcel	NOT AVAILABLE. Probably eliminate this variable from the models that use it.
PARKDY_Q	*Daily paid parking spaces- quarter mile radius*	NOT AVAILABLE
PARKDY_H	*Daily paid parking spaces- half mile radius*	NOT AVAILABLE
PPRICDYP	*Avg price daily parking- parcel (cts)*	NOT AVALIABLE. Use avg. daily parking cost in the parcel' s zone for workers, available exogenously.
PPRICDYQ	Avg.price daily parking- quarter mile (cts)	NOT AVAILABLE. Use avg. daily parking cost in the parcel' s zone for workers, available exogenously.
PPRICDYH	*Avg.price daily parking- half mile (cts)*	NOT AVAILABLE. Use avg. daily parking cost in the parcel' s zone for workers, available exogenously.
PARKHR_P	Hourly paid parking spaces- parcel	NOT AVAILABLE. Probably eliminate this variable from the models that use it.
PARKHR_Q	*Hourly paid parking spaces- quarter mile radius*	NOT AVAILABLE
PARKHR_H	*Hourly paid parking spaces- half mile radius*	NOT AVAILABLE
PPRICHRP	Avg price hourly parking- parcel (cts)	NOT AVAILABLE. Use avg. hourly parking cost in the parcel' s zone for shopping, available exogenously.
PPRICHRQ	Avg.price hourly parking- quarter mile (cts)	NOT AVAILABLE. Use avg. hourly parking cost in the parcel' s zone for shopping, available exogenously.
PPRICHRH	*Avg.price hourly parking- half mile (cts)*	NOT AVAILABLE. Use avg. hourly parking cost in the parcel' s zone for shopping, available exogenously.
EMPEDU_P	Education jobs – parcel (x 100)	
EMPFOODP	Food service jobs – parcel (x 100)	
EMPGOV_P	Government jobs – parcel (x 100)	
EMPOFC_P	Office jobs – parcel (x 100)	
EMPOTH_P	Other jobs – parcel (x 100)	
EMPRET_P	Retail jobs – parcel (x 100)	
EMPSVC_P	Service jobs – parcel (x 100)	
EMPMED_P	Medical jobs – parcel (x 100)	
EMPIND_P	Industrial jobs – parcel (x 100)	
EMPTOT_P	Total jobs – parcel (x 100)	
EMPEDU_Q	*Education jobs – quarter mile radius (x 100)*	
EMPFOODQ	Food service jobs – quarter mile radius (x 100)	
EMPGOV_Q	*Government jobs – quarter mile radius (x 100)*	

EMPOFC_Q	*Office jobs – quarter mile radius (x 100)*	
EMPOTH_Q	*Other jobs – quarter mile radius (x 100)*	
EMPRET_Q	Retail jobs – quarter mile radius (x 100)	
EMPSVC_Q	Service jobs – quarter mile radius (x 100)	
EMPMED_Q	Medical jobs – quarter mile radius (x 100)	
EMPIND_Q	*Industrial jobs – quarter mile radius (x 100)*	
EMPTOT_Q	*Total jobs – quarter mile radius (x 100)*	
EMPEDU_H	*Education jobs – half mile radius (x 100)*	
EMPFOODH	Food service jobs – half mile radius (x 100)	
EMPGOV_H	Government jobs – half mile radius (x 100)	
EMPOFC_H	*Office jobs – half mile radius (x 100)*	
EMPOTH_H	*Other jobs – half mile radius (x 100)*	
EMPRET_H	Retail jobs – half mile radius (x 100)	
EMPSVC_H	Service jobs – half mile radius (x 100)	
EMPMED_H	*Medical jobs – half mile radius (x 100)*	
EMPIND_H	*Industrial jobs – half mile radius (x 100)*	
EMPTOT_H	Total jobs – half mile radius (x 100)	

The following table lists the employment categories used in the above parcel variable table, accompanied by a more detailed list of subcategories comprising each category.

Table 10: AB Model Employment Categories

DaySim Category	Subcategories
Education	Primary Education
	other education
Food	Restaurants
Government	State and Local Govt
	Govt Non-utility enterprises
Industrial	Ag and Mining
	Construction
	Manufacturing
	Commun. and Utilities
	Wholesale trade
Medical	Health services
Office	Finance, Insurance, Legal
	Real Estate
	Business Services
	"Membership and Non-profit" orgs
	Prof. Services
Retail	Retail trade
Services	Hotels
	Transp. Services
	Automotive services
	Amusement services
	Personal Services
Other	Military
	Other

Accessibility Variables from DaySim

DaySim currently generates several kinds of disaggregate and aggregate logsums, as described in the previous section of this report on Accessibility variables. These can be used by DeltaSim instead of the variables currently supplied by the TBRPM.

Needed Enhancements

Based on preliminary discussions with Wade White of Citilabs, we have identified several aspects of DeltaSim and DaySim that need to be enhanced in order for DeltaSim and DaySim to interface effectively:

Enhancements Needed for Synthetic Population.

As indicated in Table 8, DaySim relies on several household and person variables that are available in PUMS, but which DeltaSim does not use and evolve. These include number of persons in family, household income, student status, student grade, worker status and worker hours. At a minimum, DeltaSim should be enhanced to provide the PUMS variables needed by DaySim. In order for DaySim to make reliable predictions, DeltaSim should be enhanced to evolve these variables correctly. In some cases, it may be easier to evolve simplified definitions of the variables (e.g., instead of evolving the grade variable, it would be acceptable to distinguish college students from grade school students, based on the DaySim requirements identified in Table 8.) If simplified versions are provided, then DaySim would need to be adjusted in order to accommodate the change in input variable definition.

In generating its base year population, DeltaSim does not control for household income, size and number of workers, which are important variables in DaySim. Therefore, in order to properly evolve the population for purposes of DaySim, DeltaSim needs to start with a base year population that is well-controlled by income, size and number of workers. We recommend that DeltaSim be enhanced to incorporate a version of the Atlanta Regional Commission population synthesizer for purposes of generating the base year population. The ARC population synthesizer could be used to generate a year 2000 population directly from decennial census data for permanent resident households and, if desired, non-institutionalized group quarters. It could also be used to update the population to the 2006 base year using control information supplied by the Tampa region, perhaps from ACS data, or DeltaSim could evolve the population from the year 2000. DeltaSim would continue to be relied on exclusively to generate seasonal households, hotel/motel occupants and perhaps group quarters residents.

Enhancements needed for Parcel Data

DeltaSim needs to be enhanced to provide the parcel attributes listed in Table 9. A few important issues need to be addressed in order to achieve definitions that are as consistent as possible with the DaySim definitions. Most significantly, the DaySim employment categories do not match the employment categories used by DeltaSim. In order to provide parcel employment in these categories, DeltaSim would need to be enhanced as follows: (1) Using formulas from the existing trip generation model, calculate area type for all zones in the region; (2) Implement a new set of rules that uses employment SIC codes of employment in the parcel and the area type of the parcel's zone to allocate the parcel's employment among the DaySim employment categories. Other issues are identified in the right-hand column of Table 9.

DaySim works best when developed parcels are relatively small geographically, and in terms of the number of workers employed or students enrolled on a parcel. Some cases, such as universities on single parcels, or very large undeveloped parcels that get subdivided and intensively developed, can cause problems. The quality of the overall model system would be improved if a method were developed in DeltaSim to subdivide these parcels, allocate the employment and population among the subdivided sections, and treat these subparcels as parcels when generating the parcel input to DaySim.

Enhancements needed for Accessibility variables

DeltaSim might need adjustments to substitute DaySim accessibility variables for the existing TBRPM accessibility variables currently used. It may be desirable to expand or adjust the way that accessibility variables are used instead of simply making a straight substitution. DaySim might need a minor enhancement to supply the needed accessibility variables efficiently to DeltaSim.

Interface between DaySIm and the other TBRPM-DaySim Components

The design of TBRPM-DaySim calls for the retention of many of the current TBRPM components. The main change is to remove from it the generation, distribution and mode split of trips for permanent households and seasonal residents, and to make other adjustments that are needed in order to supply DaySim with the needed skim inputs and to merge DaySim trip outputs with TBRPM trip outputs for purposes of assignment.

Needed Enhancements

Non-DaySim trips. DaySim will predict the trips of the region's year-round households and seasonal residents, including their trips to and from the airport for work, and their trips in small trucks for personal activities. The existing trip generation, distribution and mode split models will be retained, but only applied for trips other than those predicted by DaySim. These include visitors, college students, group quarters residents, commercial truck traffic, taxis, externals (EE, EI and IE) and airport travelers. It is expected that some of the existing procedures may need to be adjusted in order to accommodate this change. The resulting trip matrices will be for four separate time periods for highway assignment, and for at least two time periods for transit assignment (see below.)

Highway assignment. Highway assignment will be enhanced to assign trips for four time periods in the day (AM peak, midday, PM peak and evening). This is especially valuable because it will provide OD travel times and costs to DaySim that vary by time of day, providing information that enables the DaySim models to do a better job of predicting travel by time of day.

Mode choice and transit assignment. The TBRPM transit assignment procedures will be enhanced to assign transit trips for at least two time periods (AM peak and midday). A pre-assignment routine (the existing SACOG routine) will be implemented that splits each drive-access transit trip into a drive trip and a transit trip, after determining the transfer point (park-and-ride lot or station). If DaySim outputs only one transit OD matrix per time period, then transit assignment would be enhanced as necessary so that it could be relied on to accommodate transit submode choice via path selection (See DaySim comparison item number 12 above). Otherwise, transit assignment would be enhanced to assign trips separately for any additional transit submodes for which DaySim provides OD matrices.

Skims. Skimming procedures will be adjusted to provide skims files and variables by time of day for each assignment time period, as required by DaySim, and shown below in Tables 11 through 15.

Trip matrices. A procedure will be implemented that aggregates the DaySim trip outputs and merges them with the other trips into the matrices used for highway and transit assignment.

CUBE user interface. The CUBE user interface needs to be modified to accommodate DaySim and the changes to the other components of TBRPM.

For many of the above changes, the existing Cube/Voyager/TP+ code used in the SACOG model system can be used or serve as a guide for the required modification of the TBRPM scripts.

Skim data from Assignment Models

Tables 11 through 15 describe the skim data used by DaySim. DaySim reads all level of service files as space delimited ASCII files with no header record. All values are integer values, with no decimal.

Table 11: File 1—Walk skim file

Label	Definition
ORIG	Origin zone
DEST	Destination zone
WALKDIST	Walk distance (miles x 100)

Table 12: Files 2&3—AM peak and PM peak highway skim files

Label	Definition
ORIG	Origin zone
DEST	Destination zone
D1TIME	SOV time (minutes x 100)
D1DIST	SOV distance (miles x 100)
D1EXTT	SOV congested time 1 (minutes x 100)
D1EXTT2	SOV congested time 2 (minutes x 100)
D1TOLL	SOV toll (cents)

D2TIME	HOV time (minutes x 100)
D2DIST	HOV distance (miles x 100)
D2EXTT	HOV congested time 1 (minutes x 100)
D2EXTT2	HOV congested time 2 (minutes x 100)
D2TOLL	HOV toll (cents)

Notes:
--Congested time 1 and congested time 2 in these and other skim files are excess travel time due to congestion beyond 1.2 x free-flow time and 1.5 x free-flow time, respectively.
--SACOG does not distinguish skims for HOV2 from those for HOV3+. However, if Tampa Bay region needs distinct HOV 2 and HOV 3+ skims then DaySim can accommodate them, using an additional five skim variables (D3TIME, etc).

Table 13: Files 4&5—Midday and evening highway skim files

Label	**Definition**
ORIG	Origin zone
DEST	Destination zone
D1TIME	SOV time (minutes x 100)
D1DIST	SOV distance (miles x 100)
D1EXTT	SOV congested time 1 (minutes x 100)
D1EXTT2	SOV congested time 2 (minutes x 100)
D1TOLL	SOV toll (cents)

Note:
--Only the peak periods have separate HOV skims. HOV 2 is set to equal SOV for the off-peak periods. HOV 3+ is set to equal HOV 2 for all periods. However, if Tampa Bay region needs distinct HOV 2 skims for midday and evening, DaySim can accommodate them.

Table 14: Files 6-8: AM peak, midday and evening walk to transit skims (only OD pairs with valid transit paths have records in the file)

Label	Definition
ORIG	Origin zone
DEST	Destination zone
XFNUMW	Number of transfers
XFTIMW	Transfer time (min. x 100)
FWTIMW	First wait time (min. x 100)
FAREW	Fare (cents)
TRDISW	In-vehicle distance (miles x 100) *
WATIMW	Walk time (min x 100) *
TRTIMW	In-vehicle time (min x 100)

* not used in models

Notes:

--The reverse directions of the AM peak paths are used for the PM peak. If Tampa Bay region has separate PM peak skims, DaySim can instead accommodate them.

--Walk time is not used in the models because we have parcel-specific walk distances to transit.

Table 15: Files 9&10—Peak and off-peak drive to transit skims (only OD pairs with valid transit paths have records in the file)

Label	Definition
ORIG	Origin zone (Drive end)
DEST	Destination zone (Walk end)
PKTAZD	Park and ride lot zone number
XFTIMD	Transfer time (min x 100)
FWTIMD	First wait time (min x 100)
DRTIMD	Drive access time (min x 100)
FARED	Fare (cents)
DRDISD	Drive access distance (miles x 100)
TRDISD	In-vehicle distance (miles x 100) *
WATIMD	Walk egress time (min x 100) *
XFNUMD	Number of transfers
TRTIMD	In-vehicle time (min x 100)

* not used in models

Note:

--The reverse directions of the AM peak paths are used for the PM peak, but the drive portion is assumed to be egress in the PM peak. If Tampa Bay region has separate PM peak skims, DaySim can instead accommodate them.

--Walk time is not used in the models because we have parcel-specific walk distances to transit.

DATA REQUIREMENTS FOR PHASE ONE IMPLEMENTATION

Data for Population Synthesis

If it is decided to use the ARC population synthesizer to provide a starting population of permanent and perhaps also non-institutionalized group quarters residents, then the following input data would be required by the population synthesizer.

- 2000 PUMS data for all PUMAs in the region.
- 2000 SF1, SF3 and CTPP tables for the entire region.
- Control tables for 2006 and all forecast years to be used for updating the synthetic population for base (2006) and forecast years. For 2006 it might be possible to use ACS data for this purpose.
- Conversion tables that relate various geographic partitions of the region (block, block group, tract, county, PUMA and TAZ).

Data for Model Calibration and Validation

Long-term models

- Usual work location: Calibration and validation targets will be derived from the household survey and the year 2000 Census data (adjusted to reflected changes in employment and work flows between 2000 and 2006). These measures will include subarea employment, subarea-to-subarea home-to-work flows, home-to-work distance distributions, and work-at-home estimates.
- Usual school location: Calibration and validation targets will be derived from Tampa Region estimates of school enrollment by TAZ and grade level (preschool, elementary, middle, high and university), and any available information about home-to-school distance distribution for each of the grade levels.
- Auto availability: Calibration and validation targets will be derived from the household survey and the year 2000 Census Transportation Planning Package (adjusted to reflected changes in regional households between 2000 and 2006). These measures will include numbers and shares of households by vehicle availability class and district. No additional data collection is required.

Day-Level Models

- Person day pattern: The household survey will be the primary source for model calibration and validation targets. These targets will include tour and stop frequency shares by person type, and may be adjusted in order to validate the entire model system. No additional data collection is required.

- Exact number of tours by purpose: The household survey will also be the primary source for model calibration and validation targets. These targets will include shares and totals of tours by person type, household income segment, auto sufficiency, and district, and may be adjusted in order to validate the entire model system. No additional data collection is required.

Tour-Level Models

- Tour destination: The household survey, Census data, and employment totals at various levels of geographic aggregation will be the primary sources of calibration and validation targets. Tour destination measures will include average tour distances, tour length and tour duration frequency distributions, and district-to-district flows by tour purpose. These targets may be adjusted to account for discrepancies amongst data sources or to validate the entire model system.
- Work-based subtour generation (workers only) – number and purpose: The household survey will be the primary source for model calibration and validation targets. These targets may be adjusted in order to validate the entire model system. No additional data collection is required.
- Tour main mode: The household survey, Census data, and any other appropriate mode choice information potentially derived from transit on-board surveys and transit agency data will be the primary sources for calibration / validation targets. Mode shares by purpose will be the primary targets, and may be adjusted to account for discrepancies amongst data sources or to validate the entire model system.
- Tour time-of-day: The household survey, traffic and transit counts, and other appropriate data will be the primary sources for model calibration and validation targets. Targets will include shares of tours by purpose and time-period-combination and duration. The household survey can provide information on time-period combinations for tours, while detailed traffic and transit counts (at a minimum resolution of 1-hour intervals) can provide information on overall levels of tripmaking by time-of-day. New data collection or processing may be required to develop or extract the required time-of-day information. These targets may be adjusted in order to validate the entire model system.

Trip/Stop-Level Models

- Intermediate Stop Generation (exact number) – Predict number purpose and sequence of stops: The household survey will be the primary source for model calibration and validation targets. These targets will include shares and totals of stops by person type, household income segment, auto sufficiency, and district as well as shares of tour/stop combinations, and may be adjusted in order to validate the entire model system. No additional data collection is required.

- Stop location: The household survey will be the primary source for model calibration and validation targets. Primary measures will include average trip distances (including "extra distance" on the tour), trip length and duration frequency distributions, and district-level summaries. These targets may be adjusted in order to validate the entire model system. No additional data collection is required.
- Trip mode choice: The household survey will be the primary source for model calibration and validation targets. Mode shares by purpose will be the primary targets, though these shares will be evaluated in the context of observed roadway and transit counts and may be adjusted to account for discrepancies amongst data sources or to validate the entire model system. Trip mode choice will probably not be validated separately from tour mode choice.
- Trip arrival and departure time: The household survey, traffic and transit counts, and other appropriate data will be the primary sources for model calibration and validation targets. Targets will include shares of trips by time period. As with tour time-of-day, he household survey can provide detailed information on tripmmaking by time period and purpose, while traffic and transit counts can provide information on overall levels of tripmaking by time-of-day. New data collection or processing may be necessary to develop or extract the required time-of-day information. These targets may be adjusted in order to validate the entire model system.

Network Assignment

The primary means of validating the entire travel model system will be by comparing forecasted to observed roadway and transit volumes and speeds. The calibration data requirements are much the same as for the existing TBRPM model system, so the 2006 TBRPM calibration data should serve as an excellent starting point.

In order to ensure that the component models and the overall model system are performing reasonably well, a comprehensive count and speed dataset should be developed for roadways, and a transit volume dataset should be developed for transit. This dataset should be based around the year 2006, though it will likely be necessary to include counts and speeds from other adjacent years in order to achieve more robust calibration and validation information.

For roadways, this database would ideally include traffic counts by disaggregate time period such as the half hour time periods used in the tour time of day and stop arrival and departure models. This detail would provide the greatest flexibility in analyzing and defining the more aggregate time periods that are anticipated to be used in the initial roadway assignment models and would also support the analysis of phenomena such as peak spreading. This temporal detail can also support calibration of the tour and stop timing models. In addition, the roadway validation database should include traffic counts that provide a reasonable coverage of the entire region and at critical external boundaries, but that are also focused on critical cordons and screenlines within the region. In the roadway assignment calibration / validation process estimated and observed traffic volumes will be evaluated by facility type, area type, volume class, district, time-of-day, and at critical cordons.

Roadway travel times or speeds by segment can provide important information to the roadway assignment model (as well as overall model system) calibration and validation efforts. Travel time and speed data should be focused on the most congested facilities during periods of peak congestion, and ideally would incorporate both highway and surface street segments.

The transit volume dataset should at a minimum include transit boardings by route and time period, for all transit operators and services in the region. Ideally, this regional transit database would also include additional information such as transit volumes at critical cordons and screenlines are also essential to validating the tour and trip mode choice and transit assignment models. In addition to providing detailed route-level transit validation data, Automated Passenger Count (APC) and Automated Vehicle Location (AVL) data can also provide essential information about transit passenger flows and the relationship between roadway and transit travel times and speeds. This information can be used to refine roadway and transit network assignment parameters and ensure the reasonableness of mode choice model outputs. Transit on-board survey data can also provide critical insights into the reasonableness of forecast transit travel by providing observed information on transit origin-destination patterns, transfer rates, and access modes.

PHASE TWO MODEL IMPLEMENTATION

The recommended Stage Two model implementation has two primary focuses. The first focus is to estimate the existing DaySim models using new household survey data collected from residents of the Tampa Bay region as part of the current 2008 National Household Travel Survey (NHTS), which will be available sometime during 2009. While doing this, new variables may be developed that better capture things like walk accessibility. Also, the model specifications will probably be enhanced somewhat, in order to embody any distinctive aspects of the region's residents, as well as to improve the models based on insights that have been gained since the DaySim models were estimated for SACOG. For example, the accessibility variables used to capture the effects of travel conditions on the generation of tours may be improved so that they are more realistically sensitive to congestion pricing that varies by time of day. Any region-specific issues that arise in the Stage One model implementation can also be addressed in Stage Two.

A second possibility for Stage Two is to incorporate additional model components that would improve the model's usefulness for the kinds of projects and policy analysis that are important for the region. For example, if it is important to improve the effects of fuel prices and vehicle types on vehicle usage and mobile emissions, then vehicle type choice models could be added.

A potential third focus of Stage Two might be to expand the scope of persons handled by DaySim to include non-institutional group quarters residents and visitors (hotel/motel). This would require survey data to be collected from these segments of the population similar to the 24-hour diaries collected from resident households.

Potential Model Structure

Table 16 identifies models that could be included in an enhanced Phase Two model structure. The Phase One models are unshaded, and the potential Phase Two additions are shaded.

Table 16. Possible Component Models of DaySim in Phase Two

Model	Model Name	Level	What is predicted
	Long-term models		
1.1	Usual Workplace Location	Worker	Workplace location zone and parcel
1.2	Usual School Location	Student	School location zone and parcel
1.3	Auto availability	Household	Number of vehicles available for use by household members
1.4	Auto type (optional)	Vehicle	Type of vehicle
	Day-level models		
2.0.1	Household day pattern (optional)	HH-day	Whether pattern is (1) work or school on tour, (2) other on tour, or (3) at home all day for all persons in household
2.0.2	Household joint tour generation (optional)	HH-day	Number and purpose of joint tours in the household
2.0.3	Joint tour participation (optional)	HH-day	Persons on each joint tour
2.1	Day 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.
	Tour-level models		
3.1	Tour Primary Destination	(Sub)Tour	Primary destination zone and parcel (models are purpose-specific)
3.2	Work-Based Tours	Work Tour	Number and purpose of any subtours made during a work tour
3.3	Tour Main Mode	(Sub)Tour	Main tour mode (models are purpose-specific)
3.3.4	Tour vehicle (optional)	Auto Tour	Vehicle used for tour
3.4	Tour Time of Day	(Sub)Tour	The time period arriving and the time period leaving primary destination (models are purpose-specific)
	Trip/stop-level models		
4.1	Stop frequency and purpose	Half Tour	Number and activity purpose of any intermediate stops made on the half tour, conditional on day pattern
4.1.1	Linked Escort Trips (optional)	Half-tour	Linkage of escort trips with trips of the escorted
4.2	Intermediate Stop Location	Trip	Destination zone and parcel of each intermediate stop, conditional on tour origin, destination, and location of any previous stops
4.3	Trip Mode	Trip	Trip mode, conditional on main tour mode
4.4	Trip Departure Time	Trip	Departure time within 30 min. periods, conditional on time windows remaining from previous choices

Descriptions of Optional Model Components

The model components that we recommend FDOT consider as options for the Phase Two implementation are grouped into two categories below.

The first three optional model components explicitly predict intra-household interactions during the day:

Household day pattern (2.0.1)—Optional

This model determines the basic pattern type simultaneously for all members of the household. As estimated for Atlanta Regional Council (ARC) by Vovsha and Bradley, it defines three basic pattern type alternatives for each person: (1) out of home work or school activities, (2) other out of home activities, or (3) at home all day. An alternative definition would include work-at-home patterns in category (1). In the same way that the person-day pattern model ties together the tours of one person within a day, this model ties together the day patterns of members of the household. It can thus naturally extend the day pattern approach to encompass the entire household. Starting at the top, the basic within-day hierarchy becomes household-day > person-day > tour > trip. The most tangible advantage of integrating the person-day models in this way is that it yields more realistic household day patterns, capturing tendencies for persons in a household to coordinate their schedules. For example, in two-worker households without children, workers might be inclined to work on the same days, whereas in two-worker households with children they might be inclined to not work on the same days, or to stay home when a child stays home.

Joint tour generation and participation (2.0.2-2.0.3)—Optional

The joint tour generation model identifies the number and purpose of tours that are conducted jointly by two or more members of the household (meaning that all travel and activities on the tour are done together), and the participation model identifies the household members who participate in each joint tour. An advantage of including this model is that the conditional models of destination, mode and timing can be constrained to have the same outcome for all household members participating in a joint tour. This type of model has been implemented for the Columbus MPO (MORPC), and will also be implemented for the Atlanta (ARC) models. This model and the previous (household day pattern) model are likely to make the model more responsive to regional changes in demographics. There is as yet no clear evidence as to whether or not they significantly change the models' sensitivity to level of service changes.

Linked escort trips (4.1.1)—Optional

Another important 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). The main benefit of explicitly linking these trips is that it makes the timing, mode and destination of the escort consistent with those of the person being escorted. The best specification of this model is not yet clear, given that such a model has not yet been implemented in practice.

Most existing models, however, including the SACOG model, include a separate "escort" purpose, so that the most important special characteristics can be captured—particularly the fact that the mode is nearly always auto, with the exception of infrequent cases of walk escort. Also, children's 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. So, the most substantial benefits can already be achieved by using implicit linkages, without resorting to the explicit linkages provided by this option.

The second group of optional model components includes two models that determine the type of vehicle and fuel usage for the autos used by the household. As fuel prices and vehicle technologies may change significantly in the future, models of this type are likely to become important. The most difficult aspect of such models, however, is that the vehicle technology situation that will be in place 20 or 30 years from now is likely to be much different from now, and it is not yet exactly clear how consumers will respond to such new technologies. For that reason, any models of this type will need to be made flexible enough to run in a scenario-based mode, borrowing input data and perhaps even model coefficients from external studies and forecasts.

Auto type (1.4)—Optional

This model would assign a vehicle type to each vehicle determined by the auto availability model. The definition of vehicle type can be devised so as to be useful for policy analysis, while at the same time distinguishing types that represent realistic differences that matter when households acquire vehicles. Possible determinants of the vehicle type definition include size class (e.g. compact, midsize, full size, SUV/pickup), fuel technology (e.g. gasoline, (bio)diesel, hybrid, electric) and price class (e.g. standard vs. luxury) . The data source for model estimation is the household survey, with the vehicle type variable derived from the vehicle make, model and year information collected for each household vehicle. For forecast years, however, use of such a model will require that the user provide a future scenario of vehicle types available on the market, in terms of vehicle technology characteristics and prices. Consumer preferences for new/future vehicle technologies can be borrowed from models estimated elsewhere from Stated Preference data

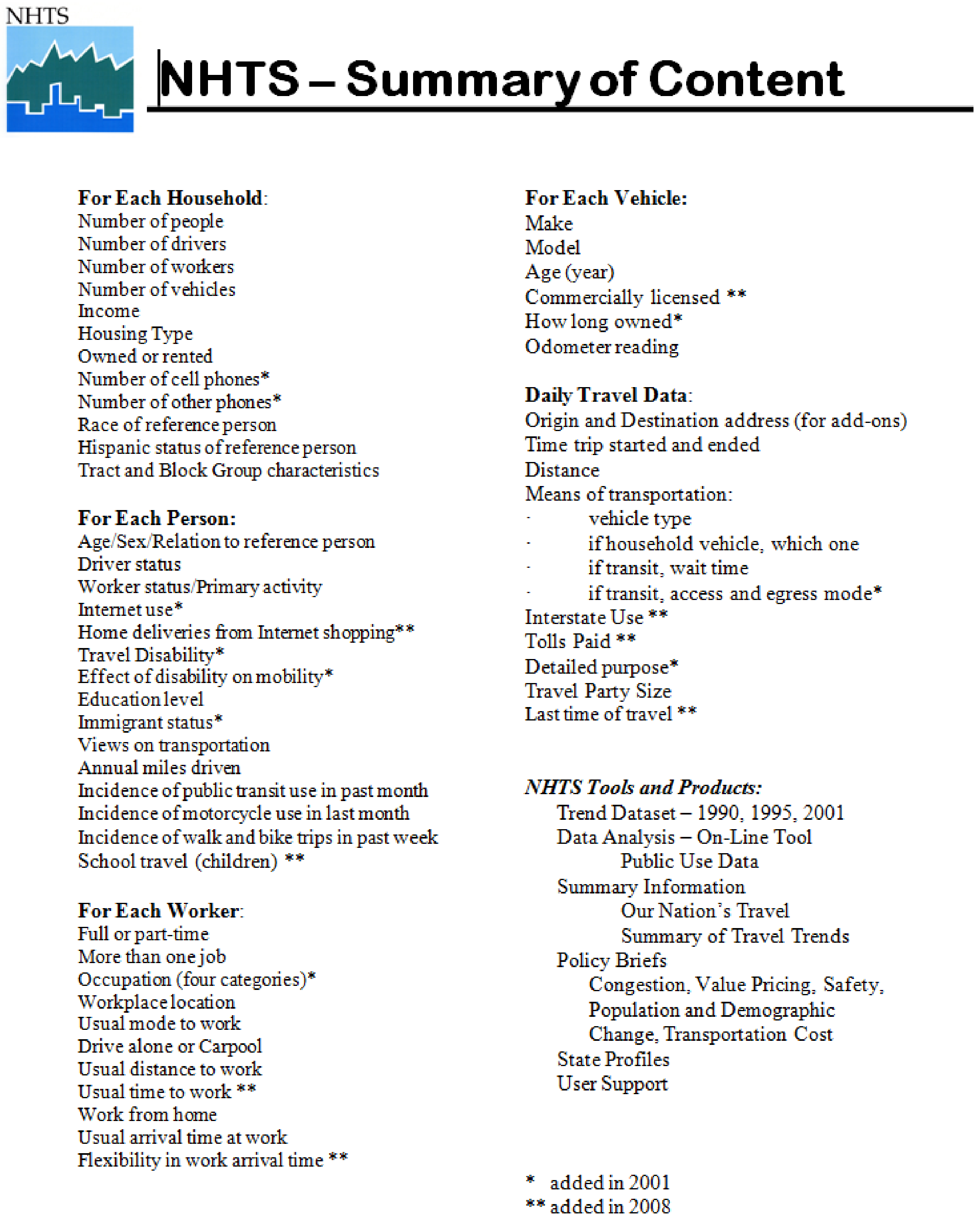
Tour vehicle (3.3.4)—Optional

If the tour is by auto and the household has more than one vehicle, then the tour vehicle model determines which vehicle to use for the tour. Important variables explaining the vehicle selection include the vehicle type, the tour purpose, the number of passengers, and the time window availability of the household's autos when not needed for other tours. It would be possible to model tour vehicle choice without strictly accounting for the available time windows for each vehicle of the household's vehicles (in the same way that each person's available time windows are accounted for), in the same way that mode choice can (and usually is) modeled without such strict accounting. However, including strict vehicle fleet accounting would enhance the quality of the vehicle type and mode choice predictions.

Household Survey Data for Phase Two Model Estimation

As part of the National Household Travel Survey (NHTS) of 2008, FDOT has commissioned an "add on" sample of approximately 14,000 households in Florida. The Tampa Region has a sample size of approximately 2,500 households. All addresses in the trip diary data for those households will be geocoded. The data items in the survey are listed in the table on the following page. The survey appears to have all data items necessary to estimate all of the models currently included in DaySim.

Note that as part of the Phase Two Model Estimation, FDOT would have the option of gaining some efficiency by estimating the same models for more than one region at the same time. As an example, using the NHTS data, models could be estimated for both the Tampa and the Jacksonville regions at the same time, with separate alternative-specific calibration constants for the regions, but most of the behavioral parameters constrained to be the same.



NHTS

NHTS – Summary of Content

For Each Household:
Number of people
Number of drivers
Number of workers
Number of vehicles
Income
Housing Type
Owned or rented
Number of cell phones*
Number of other phones*
Race of reference person
Hispanic status of reference person
Tract and Block Group characteristics

For Each Person:
Age/Sex/Relation to reference person
Driver status
Worker status/Primary activity
Internet use*
Home deliveries from Internet shopping**
Travel Disability*
Effect of disability on mobility*
Education level
Immigrant status*
Views on transportation
Annual miles driven
Incidence of public transit use in past month
Incidence of motorcycle use in last month
Incidence of walk and bike trips in past week
School travel (children) **

For Each Worker:
Full or part-time
More than one job
Occupation (four categories)*
Workplace location
Usual mode to work
Drive alone or Carpool
Usual distance to work
Usual time to work **
Work from home
Usual arrival time at work
Flexibility in work arrival time **

For Each Vehicle:
Make
Model
Age (year)
Commercially licensed **
How long owned*
Odometer reading

Daily Travel Data:
Origin and Destination address (for add-ons)
Time trip started and ended
Distance
Means of transportation:
- vehicle type
- if household vehicle, which one
- if transit, wait time
- if transit, access and egress mode*

Interstate Use **
Tolls Paid **
Detailed purpose*
Travel Party Size
Last time of travel **

NHTS Tools and Products:
- Trend Dataset – 1990, 1995, 2001
- Data Analysis – On-Line Tool
    - Public Use Data
- Summary Information
    - Our Nation's Travel
    - Summary of Travel Trends
- Policy Briefs
    - Congestion, Value Pricing, Safety, Population and Demographic Change, Transportation Cost
- State Profiles
- User Support

* added in 2001
** added in 2008

TECHNICAL MEMORANDUM 3—IMPLEMENTATION PLAN

OVERVIEW OF RECOMMENDED APPROACH

In Technical Memorandum 1, we outlined three alternative approaches to developing an activity-based model system—(1) **invent** a new model system design and computer code, (2) **adapt** an existing model design and code to local data and conditions, and (3) **adopt** an existing model system design, coefficients and code, calibrating to local conditions as necessary. We then recommended a two phase approach. In Phase 1, we will **adopt** the DaySim activity-based model code and coefficients implemented in Sacramento, and implement it inside the TBRPM system, linking the microsimulation of residents' travel with the existing models of non-resident and good travel, the existing network assignment procedures, and the DeltaSim land use microsimulation model. The only major changes to DaySim in this first phase will be the incorporation of the TBRPM mode choice model modes and structure.

In Phase 2, we will **adapt** the DaySim models, estimating new model coefficients using local NHTS household survey data that will become available during 2009. During the design stage of Phase 2, we can determine whether any additional model components or other local adaptations of the DaySim models and structure are desirable to FDOT.

The main purpose of pursuing the two-phase approach is so that we can work to get the activity-based model structure and components up and running and get all users progressing along the learning curve, without having to wait for the new data to become available. Then, by the time that the data is available, the initial implementation will have provided FDOT and other model users with a clearer idea of what types of local enhancements to the models will be most useful, ensuring that local data can be used most efficiently and effectively.

PROJECT SCOPE—PHASE 1 TASKS

In this section, we lay out the individual tasks required for the Phase 1 implementation. For each task, we describe separately the work that will need to be carried out.

Task 1: Provide base year input data

This task involves preparation of two categories of data:

- **Parcel data, zonal data, and skim data for the base year (2006)**: This includes all required input data for the TBRPM models, with the exception of the synthetic population, which is created in Task 4. With the exception of the parcel data, these will be the same data files that are needed to run the current TBRPM, so only the parcel data will need to be provided especially for this project.
- **Calibration and validation target data for the base year (2006)**: This includes typical data for model validation such as survey-based choice shares, Census-based data distributions from CTPP, traffic count and speed data, etc. If the trip-based TBRPM has been/is being re-calibrated to a new base year, then essentially the same calibration and validation targets can be used for both model systems.

Primary responsibility: Gannett Fleming

Deliverables:

1. All required input data files
2. A technical memorandum documenting file names, content, and formats.

Task 2: Adapt DaySim components

Make any necessary changes to the latest version of the DaySim program code, as implemented in Sacramento. Anticipated changes include:

- Change the dimensionality and data structures to match the TRBPM input data (number of zones, number of parcels, number of auto and transit sub-modes, number and definition of time periods for skim matrices, etc.)
- Change the code for applying the tour- and trip-level mode and location choice models to be consistent with the current TRBPM mode choice model, in terms of the (sub)modes included, the mode nesting structure, and the values of in-vehicle and out-of-vehicle time.

- Adapt the SACSIM model for simulating the choice of park and ride lot for park and ride trips. This includes the skim procedures, the choice model implementation, and the procedure for splitting park and ride trips into separate auto and transit trips.
- Set up and test procedures to use the ARC Population Synthesizer to generate a base year synthetic population. This sub-task will involve:
  - Using existing data sources to calculate sampling control targets for each specified demographic category in each zone.
  - Setting up the ARC Population Synthesizer software to use the specified demographic categorization, number of zones, correspondence of zones to PUMAS, local PUMS data records, etc.
  - Allocating the households to individual parcels within zones.
  - Generating the base year synthetic population, and checking it for reasonableness against exogenous data.
  - Putting the output household and person records in proper format for input to DaySim.

Primary responsibility: BB&C

Deliverables:

1. Revised DaySim source code and executable program
2. Base year synthetic population file(s)
3. A technical memorandum documenting revisions to DaySim
4. A technical memorandum documenting the synthetic population

Task 3: Adapt existing non-DaySim TBRPM components

Make any needed changes to the TBRPM components that generate trip matrices for all types of trips other than those made by permanent and seasonal residents and residents of non-institutional group quarters. These will include models of visitor trips, commercial trips, internal-external trips, and external-external trips. It will also include an "aggregator" module which reads in DaySim person-trip records and collapses them to O/D matrices to use in assignment. It is not yet completely clear what changes, if any, will be necessary for each of those components. Particular attention will be given to accounting for residents who work outside of the region, as well as non-residents who work inside the region.

Primary responsibility: Gannett Fleming

Deliverables:

1. CUBE scripts and any other data/programs needed to run non-DaySim TBRPM components.
2. A technical memorandum documenting these components and any changes made.

Task 4: Implement integrated TBRPM-DaySim software

First, procure and install any new custom hardware for running the integrated model system. It would be worthwhile to invest some resources in a multi-processor machine with very fast hard disk and I/O capabilities, anticipating the level of typical high-end workstations that will be running 2 or 3 years from now.

Then, put the results of tasks 1, 2 and 3 together, setting up and testing runstreams/scripts to run DaySim, CUBE assignment and skims, and all non-DaySim TBRPM components together in an integrated, automated fashion. Test this implementation for the base year, running it on the full synthetic population. At this point, testing and tuning will be done for various operational aspects of running the model system:

- Testing strategies for reaching system equilibrium, including combinations of equilibration within assignment and equilibration between assignment and the demand simulator.
- Testing equilibration of the shadow-pricing method for equilibrating predicted workplaces with the number of jobs at each location (the microsimulation approach for implementing a doubly-constrained model).
- Testing distribution of households and assignment periods across different CPU's and/or computers. This can be implemented either with or without the Citilabs CUBE Cluster software.

Primary responsibility: BB&C

Deliverables:

1. Runstreams/scripts for running all base year components together
2. A technical memorandum documenting results of tuning and testing
3. A draft User's Guide for the integrated model system

Task 5: Build DaySim query capabilities

The output from DaySim is essentially a synthesized household travel diary "survey" for the entire regional population. It is possible to perform virtually any type of analysis on the output that can be performed on travel survey data. For example, when comparing the forecasts from two different alternatives, an analysis can be done to determine who are the "winners" and the "losers" with respect to the difference between alternatives. This analysis could be in the form of tables—for instance, by income or household type—or else in the form of maps plotted in a GIS. Standard scripts or programs can be created to generate standard reports or custom queries of this type, in whatever database and/or GIS framework most suits the client.

Primary responsibility: Gannett Fleming

Deliverables:

1. Scripts and/or programs for doing queries of DaySim output in database and/or GIS environments.
2. A draft User's Guide for the query facilities

Task 6: Perform base year calibration and validation

First, calibrate the DaySim models to local survey and Census data, to the extent possible. This procedure will involve adjusting the alternative-specific constants in many of the choice models. Because those models were originally estimated using Sacramento data, this first level of calibration will be important to ensure realistic forecasts for the local population. Calibration checks will include commute patterns and auto ownership, as well as trip rates, tour rates, trip distance distributions, time of day distributions and mode shares by tour and trip purpose. Typically, the disaggregate nature of the DaySim models allows more flexibility and realism when making calibration adjustments than is possible with aggregate trip-based models.

Then, when running the full integrated TBRPM-DaySim system, perform validation checks on the model outputs, including comparing predicted link flows against traffic counts, and predicted transit boardings against on-board survey data. At this level, the validation process is no different than for a trip-based model system.

Primary responsibility: Calibration- BB&C; Validation- Gannett Fleming

Deliverables:

1. Updated model coefficients and, if necessary, updated base year input data
2. A technical memo describing calibration and validation procedures and results.

Task 7: Adapt DeltaSim for integrated use with DaySim

For forecast years, new types of input data that need to be generated for TBRPM-DaySim are (a) parcel-level land use data for the forecast years, and (b) forecast year synthetic populations, allocated to individual parcels. It is the intention that both of those inputs be provided by DeltaSim, run in an iterative fashion with the travel demand model.

Most of the work in this task will be done by Whitehouse Group. It is envisioned that the future year parcel and population input data will be provided from DeltaSim in exactly the same format as the base year input data, so that no adjustment to DaySim will be necessary to run different years. Once those formats are specified, DeltaSim will need to be adapted to:

- Read in and use accessibility measures produced by DaySim.
- Output future year parcel-level demographic and employment variables, and parcel-level land use buffer density variables.
- Output future year synthetic population files, located at the parcel level.

This task will also involve setting up and testing runstreams/scripts to run TBRPM-DaySim and DeltaSim over time in an iterative fashion, starting in the base year and evolving the land use and population in one year intervals, and running the transportation model in 5 or 10 year intervals.

Primary responsibility: Whitehouse Group

Deliverables:

1. DeltaSim source code and executable programs and required input files
2. Runstreams/scripts for iterating DeltaSim with TBRPM-DaySim across forecast years
3. A draft User's Guide for the full integrated land use-transportation model system

Task 8: Run the integrated model system for future years

Once all forecast year input data has been provided (the input zonal data and transit and highway network data will be the same for the activity-based model system as for the trip-based model system), this task will involve running the fully integrated land use-transportation model system over time. Forecast year outputs can be checked for reasonableness, both against the base year forecast and against corresponding forecast year output from the current trip-based model. In this stage, we will also examine the future year land use and population inputs generated by DeltaSim, to determine whether further adjustments are needed to the land use model.

As a final exercise in Phase 1, we will also perform a series of sensitivity tests to changes in key input variables, including fuel prices, parking costs, highway capacities, transit fares, and transit service levels. This exercise can provide information as to areas that need focused attention during Phase 2 model estimation.

Primary responsibility: Gannett Fleming

Deliverables:

1. Any future year input data files that are not generated by DeltaSim.

2. A technical memo describing results of initial future year runs and sensitivity tests.
3. A final draft of the TBRPM-DaySim report and user's guide, incorporating material from previous tasks.

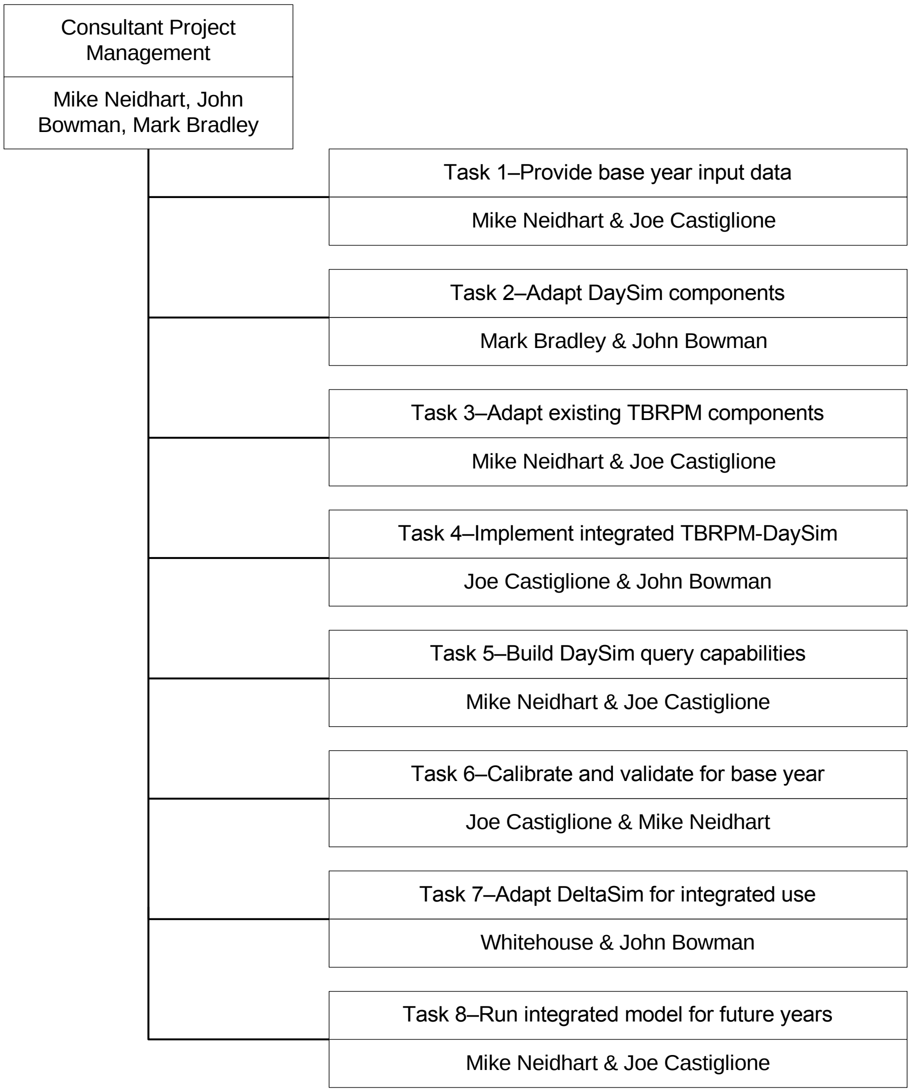
Summary of Tasks and Responsibilities

Table 17 below summarizes the primary responsibility and assisting roles for each of the eight tasks, while Table 18 breaks the tasks down into more detailed sub-tasks. Figure 10 lists key personnel by task, as well as overall project management responsibility. For each task, the person listed first would carry primary responsibility for task completion.

Table 17: Responsibilities by general tasks	**BB&C**	**Gannett Fleming**	**Whitehouse Group**
Task 1- Provide base year input data	Assist	**Primary**	
Task 2- Adapt DaySim components	**Primary**	Assist	
Task 3- Adapt existing TBRPM components	Assist	**Primary**	
Task 4- Implement integrated TBRPM-DaySim	**Primary**	Assist	
Task 5- Build DaySim query capabilities	Assist	**Primary**	
Task 6- Calibrate and validate for base year	**Primary**	Assist	
Task 7- Adapt DeltaSim for integrated use	Assist	Assist	**Primary**
Task 8- Run integrated model for future years	Assist	**Primary**	Assist

Table 18: Responsibilities by detailed tasks	**BB&C**	**Gannett Fleming**	**Whitehouse Group**
Task 1- Provide base year input data			
Management	**Primary**	Assist	
Specify data requirements	**Primary**	Assist	
Prepare parcel data	Assist	**Primary**	
Prepare skim data and zonal data	Assist	**Primary**	
Prepare validation target data	Assist	**Primary**	
Technical memo on data sets and formats	Assist	**Primary**	
Task 2- Adapt DaySim components			
Management	**Primary**	Assist	
Modify existing DaySim model code	**Primary**	Assist	
Implement park and ride lot choice model	**Primary**	Assist	
Modify population synthesis code	**Primary**	Assist	
Create and validate base year population	**Primary**	Assist	
Technical memo on DaySim changes	**Primary**	Assist	
Technical memo on base year population	**Primary**	Assist	
Task 3- Adapt existing TBRPM components			
Management	Assist	**Primary**	
Adjust non-DaySim model components	Assist	**Primary**	
Technical memo on TBRPM components	Assist	**Primary**	
Task 4- Implement integrated TBRPM-DaySim			
Management	**Primary**	Assist	
Procure and install custom hardware configuration	Assist	**Primary**	
Set up runstreams, implement, and debug	**Primary**	Assist	
Design and test equilibration procedures	**Primary**	Assist	
Preprare model system user's guide	**Primary**	Assist	
Task 5- Build DaySim query capabilities			
Management	Assist	**Primary**	
Build DaySim query capabilities	Assist	**Primary**	
Prepare user's guide for query system	Assist	**Primary**	
Task 6- Calibrate and validate for base year			
Management	**Primary**	Assist	
Calibrate the DaySim models to observed behavior	**Primary**	Assist	
Validate the full model against observed counts	Assist	**Primary**	
Technical memo on model calibration and validation	**Primary**	Assist	
Task 7- Adapt DeltaSim for integrated use			
Management	**Primary**	Assist	Assist
Specify data exchanged between DaySim & DeltaSim	**Primary**	Assist	Assist
Adapt DeltaSim to use DaySim accessibilty measures	Assist	Assist	**Primary**
Adapt DeltaSim to provide parcel/population input data	Assist	Assist	**Primary**
Create runstreams for integrated model runs	Assist	Assist	**Primary**
Update model user's guide to include DeltaSim	Assist	Assist	**Primary**
Task 8- Run integrated model for future years			
Management	Assist	**Primary**	
Provide future year network input data	Assist	**Primary**	
Specify and conduct future year runs/sensitivity tests	Assist	**Primary**	
Examine forecast year populations and parcel data	**Primary**	Assist	Assist
Technical memo on future year runs/sensitivity tests	Assist	**Primary**	
Compile final TBRPM-DaySim project report	Assist	**Primary**	

Figure 10: Key Personnel by Task



PHASE 1 BUDGET AND TIMETABLE

Budget Estimates by Task

In Table 19, we provide an estimate of the budget required by BB&C to complete each of the tasks described in the previous section. The total required budget is $240K, exclusive of travel costs. With travel, the estimated budget is approximately $250K.

Table 19: Estimated BB&C Budget by Task	**Cost ($ x 1000)**
Task 1- Provide base year input data	9
Task 2- Adapt DaySim components	99
Task 3- Adapt existing TBRPM components	15
Task 4- Implement integrated TBRPM-DaySim	36
Task 5- Build DaySim query capabilities	4
Task 6- Calibrate and validate for base year	38
Task 7- Adapt DeltaSim for integrated use	13
Task 8- Run integrated model for future years	27
Total	**240**

Implementation Schedule

Figure 11 below shows the anticipated project schedule. It is important that all key model input data for the base year, including network skims, zonal data, and parcel data, be ready as early as possible—ideally by month 4 of the project—to allow reliable testing of the various model components that need to be adapted and tested. Any significant delays in providing input data would likely cause delays in the other project tasks.

Our goal is to have the model system fully implemented and validated for the base year by the end of 2009. Then, further sensitivity testing and test runs for future years can proceed into 2010, at the same time as we begin work on the Phase 2 tasks, which are described in the following section.

Figure 11: Proposed Implementation Schedule	**Month**	**Month**	**< 2009 >**				**< 2010 >**				**< 2011 >**			
	Start	**Finish**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
PHASE 1 TASKS														
Task 1- Provide base year input data	1	3	■											
Task 2- Adapt DaySim components	1	6	■	■										
Task 3- Adapt existing TBRPM components	1	6	■	■										
Task 4- Implement integrated TBRPM-DaySim	4	9		■	■									
Task 5- Build DaySim query capabilities	4	9		■	■									
Task 6- Calibrate and validate for base year	7	12			■	■								
Task 7- Adapt DeltaSim for integrated use	7	12			■	■								
Task 8- Run integrated model for future years	13	18					■	■						
PHASE 2 TASKS														
Task 1- Prepare the NTHS survey data	13	18					■	■						
Task 2- Design the enhanced model system	16	21						■	■					
Task 3- Estimate, implement and test new models	22	36								■	■	■	■	■

PHASE 2—PRELIMINARY TASK DESCRIPTION

While Figure 11 above shows a preliminary schedule for Phase 2, we have not yet prepared a scope or budget for Phase 2 in nearly the detail that we have provided for Phase 1. It is likely that our ideas and the client's ideas about what features of the enhanced model will need the most attention will evolve as Phase 1 proceeds. Therefore, we propose to prepare a detailed scope and budget for Phase 2 during the last quarter of 2009, with the objective that Phase 2 work could begin in early 2010. Below, we provide a preliminary sketch of the main Phase 2 tasks and possible budget.

Task 1: Prepare the NHTS survey data for model estimation

This task will involve a number of subtasks, including:

- Processing the data into various units of trips, tours, person-day activity patterns, and household-day activity patterns.
- Analyzing and reporting important behavioral distributions in the data, and assessing how well the data will support the estimation of various types of possible models.
- Processing the data to prepare model estimation data files for each of the model components that is to be estimated.

Primary responsibility: BB&C

Deliverables:

1. Cleaned and processed survey data files.
2. A technical memo describing results of preliminary data analysis
3. Data files for model estimation.

Task 2: Design the enhanced model system

Here, we will weigh the experience of the Phase 1 implementation, the priorities of the client, and the suitability of the survey data, as analyzed in Task 1, in order to arrive at a detailed model design and estimation plan. The plan will deal with:

- An approach and model specifications for re-estimating all of the model components that are part of the current DaySim.
- If agreed upon with the client, an approach for enhancing the model system to include some of the optional model components described above in Technical Memorandum 2.

- An approach for enhancing the model sensitivity and capabilities for analyzing pricing policies, including using distributed Values of Time, toll versus non-toll choice in the mode choice models, and other possible enhancements.

Primary responsibility: BB&C

Deliverables:

1. A technical memo with detailed specification of each model component to be estimated, along with an overall estimation and implementation plan.

Task 3: Estimate, implement, and test the enhanced model system

As the task name indicates, this task will contain a number of subtasks:

- Estimate each of the model components included in the design from Task 2, using the data sets prepared from local NHTS data in Task 1.
- Make any necessary changes to the DaySim code in order to apply the enhanced models.
- Make any necessary changes to the other TBRPM-DaySim components and/or the input data files.
- Test the new enhanced model system, including re-calibration and validation for the base year, and new sensitivity tests for future years.

Primary responsibility: BB&C for estimation, coding; Gannett Fleming for calibration and validation

Deliverables:

1. New DaySim code and model coefficients.
2. Updated scripts and model system User's Guide.
3. A project final report documenting all new models.

Preliminary budget estimate

Our preliminary estimates indicate that Phase 2 work will cost in the range of $240K to $480K for BB&C hours, depending on which new features and capabilities are added to the model system during the estimation and implementation stage. The low end of the budget range would be a scenario where few changes are made to the current DaySim model system structure.

APPENDICES

APPENDIX 1-- ACTIVITY-BASED MODELS: APPROACHES USED TO ACHIEVE INTEGRATION AMONG TRIPS AND TOURS THROUGHOUT THE DAY

ACTIVITY-BASED MODELS: APPROACHES USED TO ACHIEVE INTEGRATION AMONG TRIPS AND TOURS THROUGHOUT THE DAY

Presented at the 2008 European Transport Conference, Leeuwenhorst, The Netherlands, October, 2008

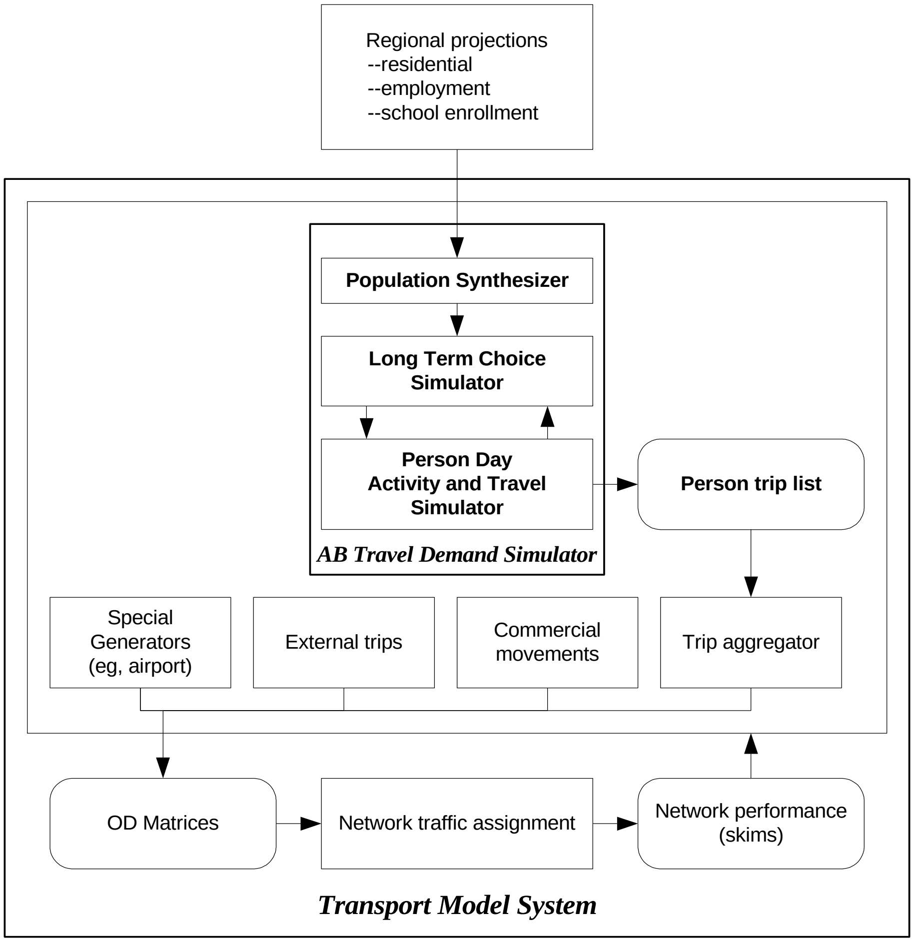
John L. Bowman, Ph. D.
email: John_L_Bowman@alum.mit.edu, website: http://JBowman.net
Mark A Bradley
mark_bradley@cox.net

introduction

This paper examines the so-called activity-based (AB) regional travel forecasting models implemented to date in the United States, explaining and discussing the various techniques that have been used to achieve model integration.

These models simulate 24-hour activity and travel itineraries for each synthetic resident of a region. The resulting trips are aggregated into trip matrices, combined with commercial trips and trips of non-residents, and assigned to transit and road networks (see Figure 1). In simulating the itineraries of one person, many dimensions of choice are modeled, including activity participation, timing and location, as well as the mode of associated travel. It is necessary to address the issue of integrating multiple model components because, on the one hand, the outcomes are related to such an extent that it seems appropriate to treat them as a single complex outcome, modeling all dimensions simultaneously and, on the other hand, the outcome is so complex that it is impractical to capture all the needed detail in a single model. Therefore, the models have been implemented as a large number of carefully integrated components. The objective of these models, and hence the objective that guides the selection of integration techniques, is to realistically model travel behavior that can be affected by changes in activity opportunities and travel conditions, especially those that are affected by public policies and programs.

Figure 1: Typical Activity-Based Regional Travel Forecasting Model System



This paper does three things. First, in Section 2, it discusses principles that have guided the development of AB models now in use for travel forecasting in the United States, focusing on principles related to the integration of systems composed of many models. Second, in Section 3, using the terminology established in the discussion of principles, it describes the specific integration techniques employed by existing AB model systems in the United States. Third, it provides a reasoned (though not empirically supported) discussion of the strengths and weaknesses of the techniques, making some preliminary judgments about their effectiveness. This critique occurs throughout Sections 2 and 3, and is summarized in Section 4 with a prioritized list of integration features. Although some model systems fare better than others in the critique, they all have substantial integration weaknesses. The purpose is not to select winners and losers. Nor is the purpose to give definitive answers regarding the best integration techniques. Rather, it is to increase the awareness of this important topic, focus the issues, and stimulate further thought, discussion and research that may lead to the development of improved integration techniques in AB models.

Integration Principles and Terminology

We start by considering two different phenomena that affect the activities and travel a person ends up completing in any given day. The first one is the passage of time. Every new action emerges from the situation in the present created by events of the past. It is tempting to thus organize an AB model in the same way, with outcomes modeled in strict temporal sequence. Indeed, many models of activity and travel have been developed according to this organizing principle. In such a framework, newly modeled activities can take as given all outcomes that occurred earlier, and must treat as unknown all outcomes that may (or may not) occur later. This does not necessarily preclude taking into consideration the various possibilities for the future; but to do so requires some complex techniques to quantify those future possibilities so that they can be considered in the model's prediction of the present activity.

A second phenomenon is purposeful human planning. People often think ahead, schedule important future activities, and arrange other activities around them in order to achieve their objectives more effectively. It is also tempting to organize an AB model in this second way, with outcomes also modeled sequentially, but according to a plan in which more important activities are modeled first, and less important activities fill the remaining time. This does not preclude taking into consideration, at the time that a more important activity is modeled, the various possibilities for the less important activities. But to do so requires complex techniques to quantify the possibilities so that they can be considered in the model's prediction of the more important activity.

In reality, the implemented days of most people are probably the result of both planning and the passage of time. Furthermore, most data that are available for developing these models provide an observed itinerary, but little or no information about how planning and the passage of time combined to cause it. Because of this, it is tempting to organize an AB model in the same way, representing a person's (or even an entire household's) day as a single complex outcome, simultaneously representing many components that are highly correlated because of the forces of planning and passage of time that together shaped the person's day. Indeed, because of the practical data limitation, and the importance of both planning and the passage of time, it seems better to formulate a simultaneous model than to formulate a sequential model organized only on planning or only on the passage of time.

However, a person's day (and even more so an entire household's day) is so complex that a simultaneous representation is not feasible. It is too complex to understand all at once, put into a mathematical form and carry out the needed computation. Model developers have been forced to break the outcome into pieces that can be implemented sequentially, specify a model of each of the components, specify important relationships among the components, and integrate them in an attempt to preserve the important relationships.

Without explicitly explaining the need to implement the models in a sequence, which is what gives rise to the concern about adequate integration, Vovsha et al (2005) have used the terms "downward integrity" and "upward integrity" to describe effective integration of sequentially applied models:

> "**Downward integrity** means that all lower-level decisions in the choice model hierarchy are properly conditional upon the upper-level decisions and take into account a gradually narrowed scope of lower-level choice alternatives as the upper-level choices progress....Downward integrity is ensured by properly sequencing the models, tracking the important variables from choice to choice that accurately describe the feasible scope left for each subsequent choice, and preventing conflicting choices for the same individual.
>
> "**Upward integrity** means that when modeling upper-level choices the composite measure of quality of the lower-level choices available for each upper-level alternative is properly taken into account."

These terms associate a direction of movement for the model sequence, with the beginning of the sequence at the 'top level' and the end of the sequence at the 'bottom level'. In this paper we adopt this commonly used—and entirely arbitrary—directional reference. Accordingly, we call downward and upward integrity two aspects of **vertical integration**.

However, in some cases, it may be important and feasible to retain the simultaneous modeling approach for a portion of the overall outcome consisting of two or more components. This would be the case where there is important complex correlation among component outcomes that can be correctly represented by a known and practical model structure. We call this **horizontal integration**. Defined in this way, horizontal integration is superior to vertical integration. A sequence of vertically integrated model components is a second best approach, to be used only when horizontal integration is infeasible. The efforts of most academic researchers in this field are often limited to the domain of horizontal integration. In the United States, vertical integration has primarily been the work of consultants carrying out projects with the clear objective of implementing complete model systems that can be used by a public agency for travel forecasting and policy analysis.

Since reality is so complex that full horizontal integration is infeasible, and it is necessary to use a sequential approach with vertical integration, then what organizing principle(s) might be used in choosing the sequence? In the context of a one-day itinerary, where choices are made for only one day, but are often heavily dominated by prior choices with longer term consequences—such as where to live and work, and whether to have a car for every driver in the household—the **time horizon** of decisions can serve as an organizing principle, with longer term choices coming before shorter term choices in the sequence. The principles of **temporal sequence** and **human planning**, described above, are also both good candidates. All three principles have been used in all or most of the existing US AB model systems. Increasingly, modelers have recognized the impact of long-term choices and habits on within-day behavior, specifying more of these models and placing them first in the sequence. For within-day choices, the existing model systems rely primarily on a human planning sequence, with temporal sequence used in some cases for what can be viewed as minor decisions.

Unfortunately, it has not even been feasible for modelers to capture, via vertical integration, all the apparent correlations among the components of a person's one-day itinerary. Nor has it been feasible to test enough variously specified vertically integrated models to state with confidence that the most important correlations have been correctly captured by the selected specifications. As a result, a great deal of modeler judgment undergirds the existing model systems, and will probably continue to do so for the foreseeable future. Judgment guides the modeler's choice of the components to include (and exclude), the components to keep together via horizontal integration, the specific sequence to use for the separate components, and the techniques of vertical integration to employ.

Examples From Model Systems Now in Use

This section examines specific integration techniques and features of the AB models now in use in the United States. The model systems considered include those of:

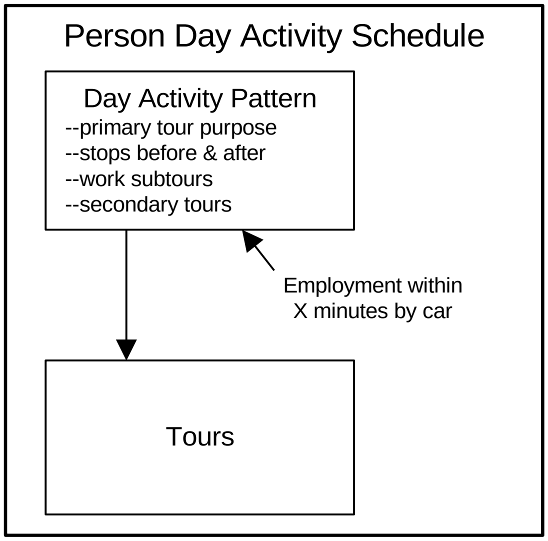
- San Francisco County Transportation Authority (SFCTA), (San Francisco County Transportation Authority and Cambridge Systematics, Inc. 2002a);
- New York MetropolitanTransportation Commission (NYMTC), (Parsons Brinckerhoff Quade & Douglas, Inc., 2005);
- Mid-Ohio (Columbus) Regional Planning Commission (MORPC), (PB Consult Inc. and Parsons Brinckerhoff, 2005); and
- Sacramento Area Council of Governments (SACOG), (Bradley, Bowman and Griesenbeck, 2006).

This list reflects the chronological order in which the model systems were developed. The subsequent discussion examines them one at a time, in the same order.

SFCTA

SFCTA has a day activity pattern model (San Francisco County Transportation Authority and Cambridge Systematics, Inc. 2002a), based on the Bowman and Ben-Akiva prototype (Bowman, 1995; Bowman and Ben-Akiva, 2001), spanning an entire day of activities and travel for one person. As shown in Figure 2, in one (horizontally integrated) model it identifies the most important on-tour activity purpose of the day, whether one or more stops is made before, during and/or after that activity on the same tour, and the presence of one or more additional tours for maintenance and/or discretionary activities during the day. Thus, this single model provides information about (a) the purpose and structure of the primary tour of the day, and (b) participation in additional tours. This enables the model to represent the total amount of tour-making during the day, and to capture trade-offs between trip-chaining on the primary tour and conducting additional tours. However, it provides no information about activity purpose, other than the purpose of the primary on-tour activity of the day.

Figure 2: The SFCTA model includes a horizontally integrated day activity pattern that encompasses tours and trip-chaining. Its upward integration is limited because it uses simple zonal measures instead of logsums that would account for differences among persons and their available travel destinations, modes and times of day.

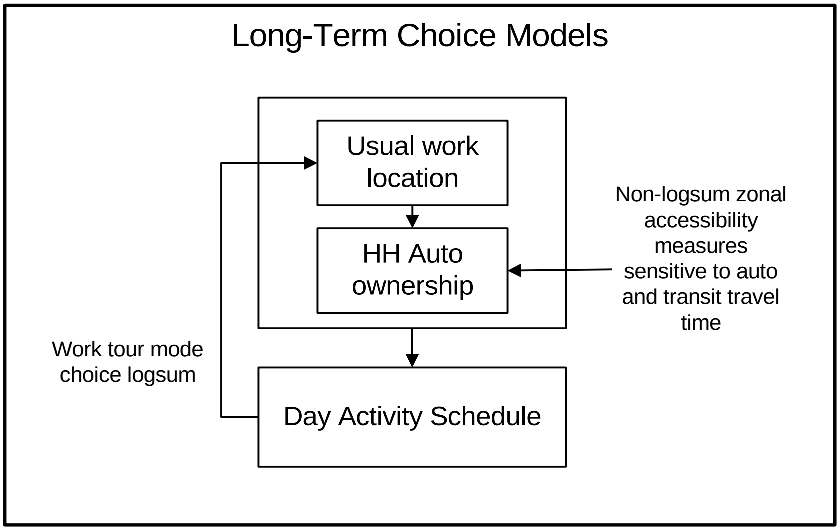


The subsequently simulated tour and stop models are conditioned by the activity pattern, providing downward integration that makes the tour models consistent with the modeled aspects of the day activity pattern. However, the model system distinguishes only five time periods during the day, so the downward integration isn't able to effectively deal with the time dimension.

The horizontal and downward integration are unable to capture the sensitivity of activity participation and tour-making (i.e. the pattern model outcome) to travel conditions such as travel time and cost. However, because the pattern model spans an entire day, it would be easy to implement upward integration that captures the effect simultaneously across all activities and tours in the day. This was a primary emphasis of Bowman and Ben-Akiva, who used logsums from the tour models to do this when they introduced their prototype. However, rather than using logsums from the tour and intermediate stop models, which would provide the best known way of capturing the accessibility effects across multiple travel modes and times, the original implementation of the SFCTA model used ad hoc accessibility measures, primarily retail and service employment within 15 minutes by car at certain times of the day. This renders the pattern model insensitive to changes in transit level of service, and to accessibility changes that affect tours and intermediate stops unequally.

SFCTA includes long-term choices of usual work location for each worker and a downwardly integrated household vehicle ownership model (see Figure 3). These are downwardly integrated with the day activity schedule of all household members. Also, for each person the usual work location model is upwardly integrated via a tour mode choice logsum, and the auto ownership model includes non-logsum zonal accessibility measures that are sensitive to auto and transit travel times.

Figure 3: SFCTA integrates long-term choices of usual work location and household vehicle ownership with the day activity schedule. The usual work location model is upwardly integrated via a tour mode choice logsum, and the auto ownership model includes non-logsum zonal accessibility measures that are sensitive to auto and transit travel times.



NYMTC

NYMTC has no day activity pattern that horizontally integrates the representation of a person’s day. Instead, as shown in Figure 4, it uses a sequence (or cascade) of thirteen tour generation models, each with a distinct combination of person type (children, non-workers and workers) and purpose (school, university, work, at-work, maintenance, and discretionary). Each subsequent model can take into consideration the outcome of the prior tour generation models. Thus, tour generation for later person-type-purpose combinations is affected by variables indicating generation of tours for earlier-simulated person-type-purpose combinations. Because there is no horizontally integrated activity pattern, there is no good way for earlier person-type-purposes to be affected by the results of those later in the sequence. For the same reason, the overall activity and travel agenda of the day cannot adjust—via upward integrity mechanisms—to changes in travel conditions (Figure 5). Furthermore, none of the thirteen individual NYMTC tour generation models is sensitive to changes in auto travel conditions, and only discretionary tour generation of non-workers is sensitive to the transit travel conditions. Even if the individual models were more sensitive to travel conditions, the cascade approach provides no good way of providing upward integrity that enables the earlier models to compensate for changes that would come indirectly through the lower level tour decisions.

Figure 4: NYMTC uses a sequence of downwardly integrated tour generation models to capture interactions among household members in tour generation. There is no horizontal or upward integration to capture non-hierarchical correlations.

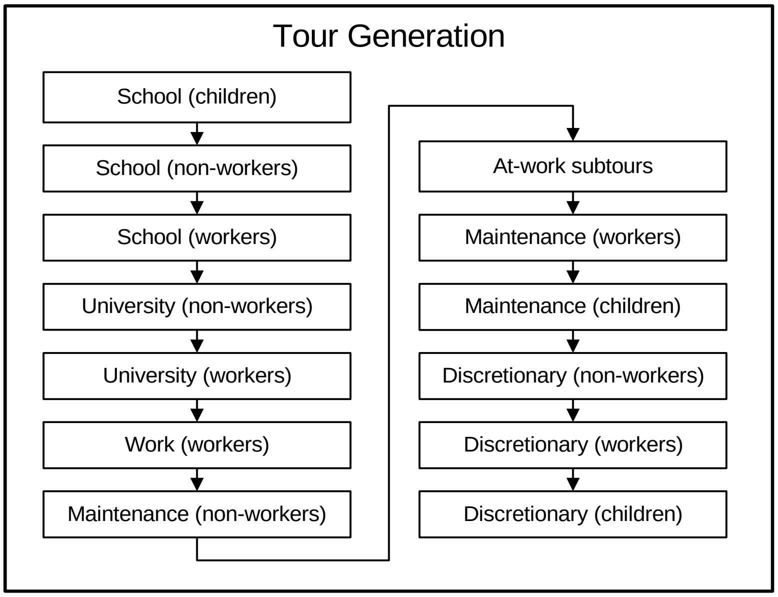
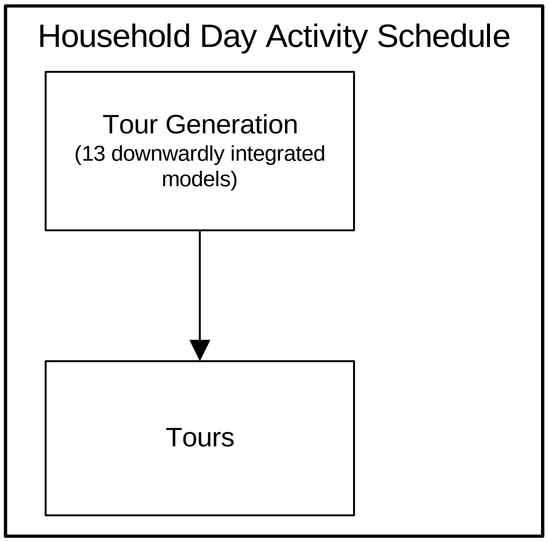


Figure 5: NYMTC lacks integration to make tour generation sensitive to travel conditions



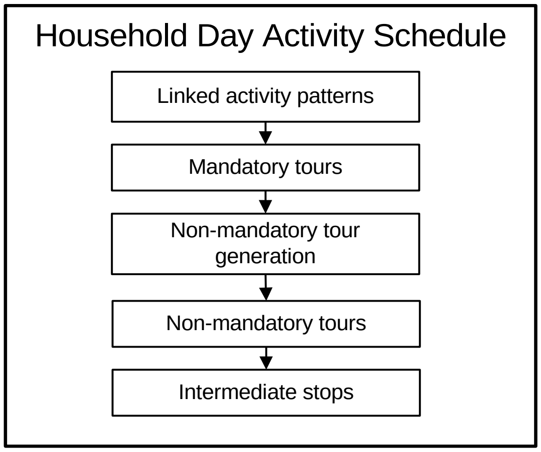
NYMTC includes a long-term model for auto ownership that is downwardly integrated with the rest of the model system, and includes zonal accessibility measures that are sensitive to auto, transit and walk travel times.

The NYMTC model and the SFCTA model employ two contrasting approaches that highlight a major trade-off faced by AB modelers. The trade-off is between emphasizing upward integration and emphasizing integration among household members; it is difficult to achieve both simultaneously. Sacrificing upward integrity weakens the ability of the model to be accurately sensitive to changes in transport conditions, especially at the highest levels of the model system, where tour generation is modeled. Sacrificing intra-household integration weakens the ability of the model to accurately represent the joint behavior of household members. SFCTA favors upward integration, whereas NYMTC favors household integration.

MORPC

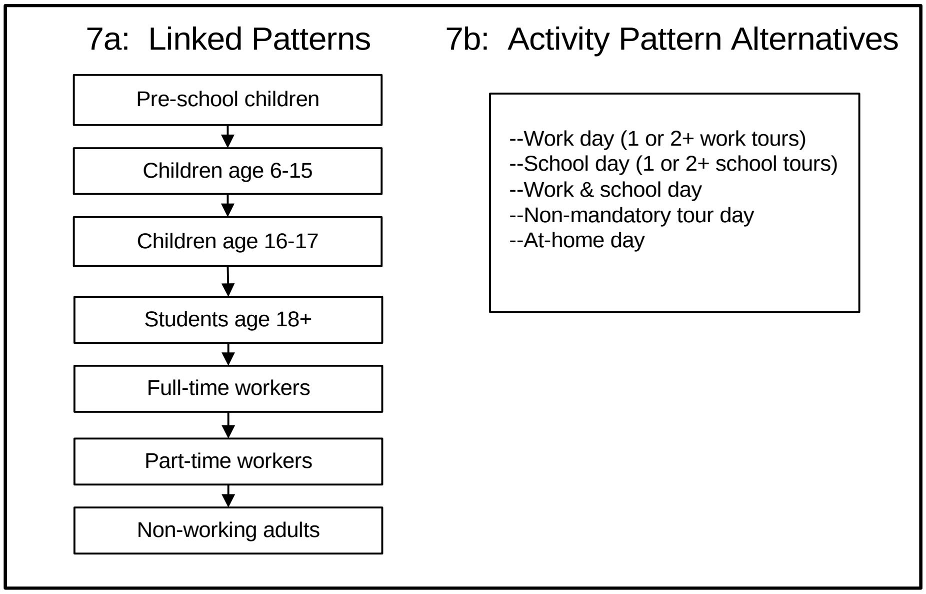
Like NYMTC, MORPC relies heavily on downward integration of a cascade of models to represent the tours and trips of persons in a household. Compared to NYMTC, it has a substantially more complex sequence of models that attempts to more realistically represent interactions among household members. The focus of attention is on downward integration. Horizontal and vertical integration are not emphasized. The model sequence is broken into several subsequences including, in order, those for 'linked activity patterns', 'mandatory' tour details, non-mandatory tour generation (including joint tour, household maintenance tours and individual discretionary tours), non-mandatory tour details, and intermediate stop generation and details (see Figure 6).

Figure 6: MORPC uses a sequence of downwardly integrated tour generation and tour model subsequences that is more complex than NYMTC's in order to capture more interactions among household members in tour generation, including linked activity patterns, joint tours and household maintenance tours. Downward integration among tours prevents conflicting time-of-day results for tours, within and across the schedules of household members. There is no horizontal or upward integration among the subsequences to capture non-hierarchical correlations among them.



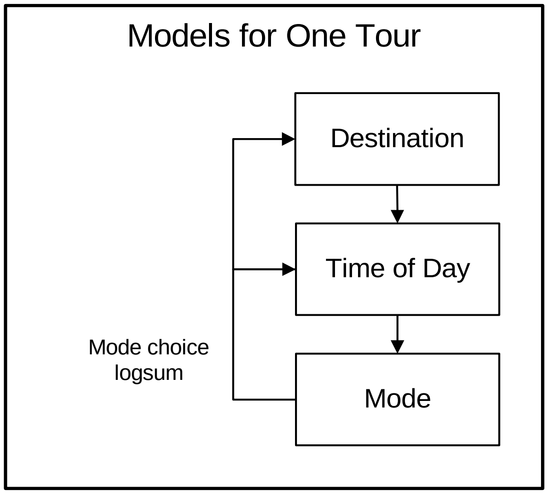
As shown in Figure 7, each model in the 'linked activity pattern' sequence (Figure 7a) predicts one of a few basic pattern types (Figure 7b) for one person in the household: whether they travel for work or school, for only other purposes, or stay at home all day, given the same type of prediction for household members earlier in the sequence (PB Consult Inc. and Parsons Brinckerhoff, 2002). The downward integration among these models allows them to partially capture strong observed correlations in primary activity purpose among household members. In particular, it captures the tendency for multiple people in the household on any given day to together stay at home, or to not go to work or school. The sequential nature of the model makes it infeasible to enable early models in the sequence to compensate for changes that affect the likelihood that lower level persons in the sequence will stay home or not go to work or school. Also, for each person in the household, this isolates the modeling of their primary activity purpose from the modeling of their other (subsequently modeled) activities of the day. This makes it impossible to use a horizontally integrated model to (simultaneously) capture the effect of travel conditions on the content and structure of their entire day. It would not preclude using upward integration—via logsums—to capture the effect of travel conditions on each individual's choice of primary activity purpose. However, the MORPC model relies instead on zonal accessibility indices that are mode- and time-of-day specific. This makes them insensitive to significant differences across households, such as income and car ownership. It also makes each accessibility index sensitive only to policies that affect a single mode and time-of-day. Because of high correlation across modes and times-of-day, it is usually possible to include only one index in a model. As a result, the model is sensitive only to one mode and time of day. In this sequence of the MORPC models, the upward integration makes the models sensitive only to walk access to jobs (for work-tour patterns) and walk access to retail (for non-work patterns). Since walk access is a function of distance, these measures (and hence the linked activity pattern) are sensitive to the distribution of employment, but insensitive to transport conditions.

Figure 7: MORPC's linked activity patterns consist of a sequence of downwardly integrated activity pattern type models, one per person in the household, that capture the tendency for multiple people to stay home together on the same day (Figure 7a). Each person's activity pattern (Figure 7b) represents the purpose of the main tour(s) of the day, but not all tour purposes or any intermediate stops. There is no upward integration from subsequent tour or stop models, nor is there upward integration that makes them sensitive to transit or auto travel conditions.



The 'mandatory' tour sequence simulates the destination, time of day and travel mode (in that order) for any work, university or school tours (called 'mandatory' tours by the developers) prescribed above for household members (see Figure 8). Through downward integration to the subsequent models, this allows the models of tour generation for maintenance and discretionary tours to be conditioned by the amount of time persons devote to their mandatory tours. There is no upward integration from the subsequent models to these detailed models of mandatory tours.

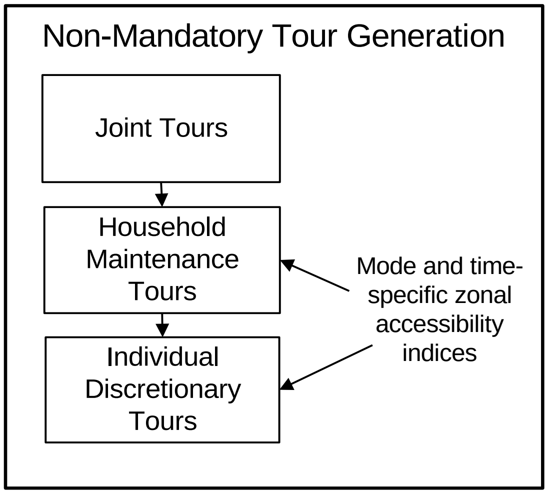
Figure 8: For each tour, MORPC uses a vertically integrated sequence of destination, time of day (TOD), and mode choice. Mode choice logsums upwardly integrate mode choice with TOD and destination choice, but destination choice uses an assumed TOD rather than a TOD logsum. The sensitivity of destination choice is thus limited to effects that occur during the assumed TOD.



Within the sequence of destination, time and mode models, there is downward integration that prevents unrealistic times to be chosen for a predicted destination, and prevents unrealistic mode for a predicted destination and time. There is also upward integration from mode to time, and from mode to destination. Upward integration is not implemented from time-of-day to destination. Rather, the destination choice model has mode choice logsums that are specific to a particular time period combination (for tour beginning and end); the time period used for the logsum depends on the purpose and other already modeled aspects of the day. This makes the destination choice model sensitive to changes in transport conditions that vary by time of day, but in a biased way, capturing sensitivity for the typical time of day, but not for the other times of day.

The model sequence that generates joint, maintenance and discretionary tours is depicted in Figure 9. The joint tour generation subsequence generates tours taken jointly by two or more household members, and then identifies the household members who participate in each joint tour. The horizontal integration of this choice among household members is consistent with the fact that some tours are indeed joint tours reflecting a mutual decision to conduct a tour together for the same purpose. This group of models has downward integration from the mandatory models, and internally, taking into consideration the time required of the various household members for their mandatory tours. There is no upward integration within this model subsequence, from subsequent models, or from the assignment models; therefore, generation of joint tours is insensitive to transport level of service.

Figure 9: MORPC generates joint tours using a model that horizontally integrates this outcome for all members of the household and subsequently assigns household members to the tours. A similar approach is used for household maintenance tours. Then discretionary tours are generated sequentially for each household member. There is no upward integration to joint tour generation, so it is not sensitive to transport conditions. For maintenance and discretionary tour generation, the upward integration comes from zonal accessibility indices that are mode and time-specific; most models are sensitive to at most one mode (auto, transit or walk) and time of day, which biases their policy sensitivity.



The maintenance tour generation and allocation subsequence has a tour generation model that horizontally integrates the generation of joint tours among all members of the household. The horizontal integration of maintenance tour generation among household members is consistent with the hypothesis that the number of maintenance tours is primarily a household decision for the purpose of achieving household objectives. This is followed by downwardly integrated models that assign the maintenance tours to individual household members. The generation and assignment of maintenance tours to household members is conditioned by the work, university and school tour obligations modeled above. There is no upward integration to capture the effect, on maintenance tour generation and assignment, of an individual's propensity for discretionary tours. There is upward integration to this generation model using zonal accessibility indices that are mode- and time-of-day specific as described above for linked activity patterns. Whether auto, transit or walk index is used varies from purpose to purpose.

Next comes a subsequence of discretionary tour generation models for each person in the household. For each person, this is conditioned by their prior modeled obligations for work, school, university, joint and maintenance tours, as well as the at-home-all-day status of other household members. Upward integration is handled like it is for maintenance tour generation.

After the generation of all joint, maintenance and discretionary tours, a sequence of models simulates their destinations, times-of-day and modes, one tour at a time, just like the sequence for mandatory tours. Downward integration limits destination and time-of-day choices according to the time already taken for previously simulated tours. This is accomplished by maintaining a set of 19 one-hour time blocks for each person, marking as unavailable those that are occupied by a newly modeled tour, and checking time slot availability to determine the choice set whenever a tour time-of-day is modeled.

The modeling of participation, location and trip mode of intermediate stops occurs on a tour-by-tour basis, but only after the generation, destination, mode and timing of all tours by all household members. The stop models are restricted by time commitments for all modeled tours, and for prior-modeled stops. There is no upward integration from the stop models to the tour models. Therefore, tour models are unable to capture trade-offs between conducting additional tours for maintenance and discretionary activities, on the one hand, and conducting those activities as intermediate stops, on the other hand.

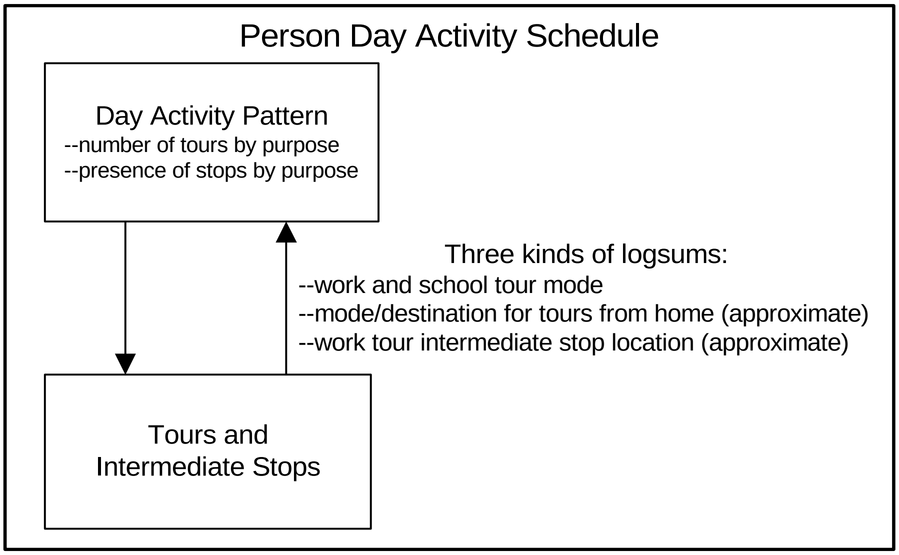
MORPC includes a long-term model for auto ownership that is downwardly integrated with the rest of the model system. It includes zonal accessibility logsums for walk and transit, making the model sensitive to transit and walk travel times, but insensitive to auto travel times.

In summary, MORPC emphasizes aspects of integration that achieve two major improvements over SFCTA and MORPC. The first emphasis is horizontal and downward integration that improves the realism of joint household outcomes and the consistency of individual schedules among household members. The second emphasis is on more detailed time of day modeling, with accompanying downward integration, that improves the consistency of time-of-day outcomes within an individual's day and across household members. This comes at the expense of upward integration which makes MORPC weaker in modeling sensitivity of the upper level models, especially tour generation and trip-chaining, to travel conditions.

SACOG

Like SFCTA, SACOG has a day activity pattern model spanning an entire day (Bowman and Bradley, 2006). It identifies the participation in one or more tours for each of seven purposes, and the participation in one or more additional stops (throughout the day) for each of the same seven purposes. This single model provides a complete inventory of the activity purposes in a day and, for each purpose, whether it is the primary objective of a tour and/or a supplemental objective on a tour. This enables the model to realistically capture the mix of tours and stops, by purpose, throughout the day. The utility of a tour or stop for one purpose directly affects its probability, as well as the probability of tours and stops for all other purposes. However, the model does not provide information on the exact number and purpose of the stops on a specific tour, or the positions of the stops on the tour; this is left for models later in the sequence.

Figure 10: The SACOG model includes a horizontally integrated day activity pattern that encompasses tours and trip-chaining for seven purposes. Its upward integration uses three kinds of logsums that account for differences among persons and their available travel destinations and modes, but not times of day.



As with the Bowman and Ben-Akiva prototype, the subsequently simulated tour and stop models are conditioned by the activity pattern, providing downward integration that makes the tour models consistent with the modeled aspects of the day activity pattern.

Also as with the Bowman and Ben-Akiva prototype, it is easy to implement upward integration that captures the effect of tour accessibility simultaneously on the overall pattern of activities and tours in the day. In contrast to the SFCTA model, the SACOG model does this with logsum accessibility variables for tours and intermediate stops, capturing several types of composite accessibility effects on the pattern, including mode choice logsums for tours to the usual work and school locations, approximate mode/destination logsums for tours from home, and approximate location choice logsums for intermediate stops on work tours.

The approximate, or aggregate, logsum is calculated in the same basic way as a true logsum, by calculating the utility of multiple alternatives, and then taking expectation across the alternatives by calculating the log of the sum of the exponentiated utilities. However, the amount of computation is reduced, either by ignoring some differences among decisionmakers, or by calculating utility for 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. This makes it computationally feasible to use logsums at the upper levels of the model system. The categories of decisionmakers and the aggregation of alternatives are chosen so that in all choice cases an approximate logsum is available that closely approximates the true logsum. In essence, this is a sophisticated ad hoc measure that is intended to achieve most of the realism of the true logsum at a small fraction of the cost.

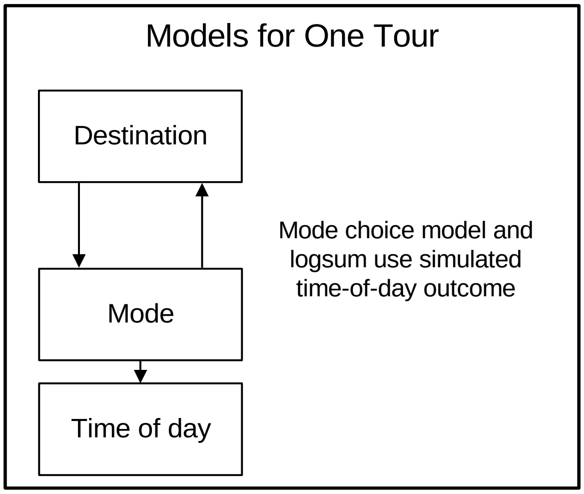
The approximate tour mode-destination choice logsum is used in situations where information is needed about accessibility to activity opportunities in all surrounding locations by all available transport modes at all times of day. Because of the large amount of computation required for calculating a true logsum for all feasible combinations in these three dimensions, an approximate logsum is used with several simplifications. First, it ignores socio-demographic characteristics, except for car availability. Second, it uses aggregate distance bands for transit walk access. Third, sometimes it uses a logsum for a composite or most likely purpose instead of calculating it across a full set of specific purposes. Finally, instead of basing the logsum on the exact available time window of the choice situation, and calculating it across all of the available time period combinations within the window, it uses a particular available time window size and time period combination. With these simplifications, it is possible to pre-calculate a relatively small number of logsums for each zone, and use them when needed at any point in the simulation of any person's day activity schedule.

The approximate intermediate stop location choice logsum is used in the activity pattern models, where accessibility for making intermediate stops affects whether the pattern will include intermediate stops on tours, and how many. Four logsums are calculated for each OD zone pair, distinguished by tour mode (transit or auto) and time of day (peak or offpeak). Each logsum is calculated across all possible intermediate stop zones, each stop's utility is a function of travel time and zonal attractiveness, and zonal attractiveness is a function of employment and school enrollment, taken from an estimated purpose-non-specific location choice model.

Although the upward integration of the activity pattern is better than SFCTA's, it is still limited. The approximate logsums are limited as described above. In addition, most of the non-work purposes do not have mode/dest logsums, so some of the benefit of the purpose-specific specification is lost in the upward integration.

For each tour there is a vertically integrated sequence of destination, mode and time-of-day models (see Figure 11). Downward integration prevents unrealistic destination, mode and time-of-day combinations. There is also upward integration from mode to destination, but it does not use time-of-day logsums. Rather, a simulation technique was implemented to make the mode and destination choice models sensitive to policy effects that may vary by time of day. The basic idea is to avoid the use of a logsum (and its associated computational costs) when applying an upper level model by treating as given a conditional outcome that is not known, and would otherwise require the calculation of a logsum from all possible conditional outcomes. In this case the assumed conditional outcome is the tour time-of-day. It is selected by a Monte Carlo draw using approximate probabilities for the conditional outcome. Rather than making every simulated outcome sensitive to variability in the conditional outcome, sensitivity is achieved across the population through the variability of outcome in the Monte Carlo draws. In this way, the mode and destination choice models are sensitive to variations in transport level of service and spatial attributes across all possible combinations of time-of-day, with the affects approximately weighted by the time-of-day choice probabilities.

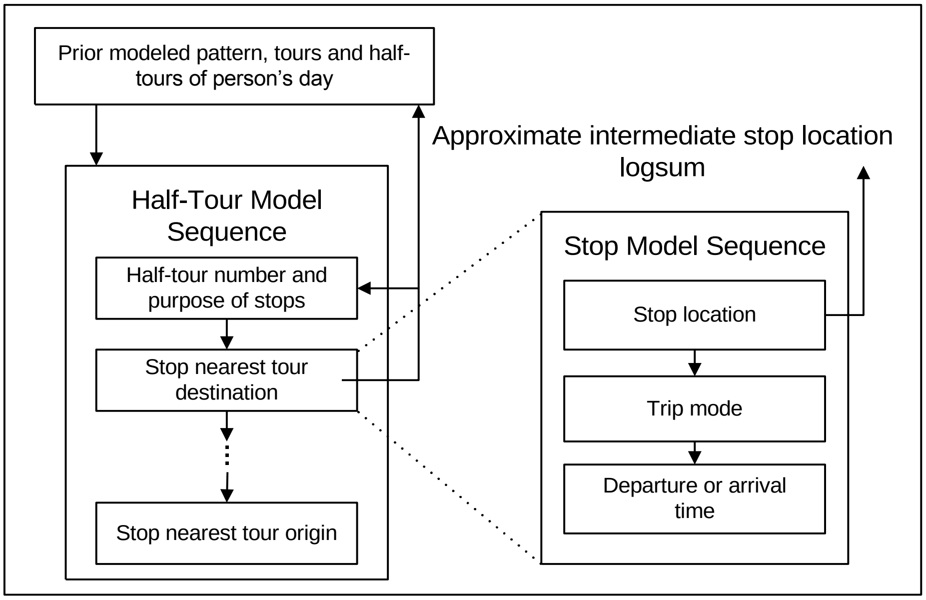
Figure 11: For each tour, SACOG uses a vertically integrated sequence of destination, mode choice, and time of day (TOD). Mode choice upwardly integrates with destination choice; it uses a simulated TOD that makes the destination and mode choice models sensitive to changes in transport conditions that vary by time of day.



SACOG conditions half-tour stop participation and purpose upon the modeled aspects of the day activity pattern, prior modeled tours, current tour, and—in the case of the second half-tour of the tour—the first half-tour (Figure 12). This enables consistency of stop participation by purpose among models at all levels, and complements the horizontal and upward integrity of stop purpose enabled by the SACOG specification of the activity pattern model.

The SACOG intermediate stop models of location, mode and timing are conditioned by the same models as the half-tour model. In addition, each one is conditioned by the half-tour model that predicts the number and purpose of stops on the half-tour, as well as all prior-modeled stop outcomes on the half-tour. The intermediate stops are simulated in sequence emanating from the tour's primary destination, in reverse chronological order for stops before the tour destination, and in chronological order for stops after the tour destination. This is based on the assumed importance of arriving at and departing from the primary destination at a previously modeled time. Accordingly the timing of each intermediate stop is conditioned by the timing of all prior-modeled stops. Similarly, the intermediate stop location choice is conditioned by all previously modeled locations, and the trip mode is conditioned by the mode set allowed by the modeled tour mode and the modes used on all prior modeled trips on the tour. This downward integration allows a consistent and feasible representation of each tour's entire travel itinerary, with regard to timing, mode and location. This is complemented by upward integrity provided by the use of intermediate stop location choice logsums in the half-tour and activity pattern models. These capture the effect of accessibility on the stop participation choices modeled in those two models.

Figure 12: Downward integration of half-tour and intermediate stop models conditioned by the pattern, tour, half-tour and prior-modeled stop models, with accompanying upward integration via approximate stop location logsums.



SACOG includes long-term choices of usual work location for each worker and usual school location for each student. For young student workers, the usual school location model conditions the usual work location model, and for old worker students the sequence is reversed. These condition a downwardly integrated household vehicle ownership model, and all of them are downwardly integrated with the day activity schedule of all household members. All three long-term models are upwardly integrated via work and/or school tour mode choice logsums, as well as approximate mode-destination logsums. For the usual location models the mode/dest logsums measure accessibility for tours from the usual location. For auto ownership, they measure accessibility for tours from home.

In summary, SACOG emphasizes and implements techniques for achieving upward integration that improves the model system's ability to accurately capture sensitivity to travel conditions at all levels of the model system, and especially at the upper levels. It also implements more complete downward integrity of the model system with regard to the modeling and accounting of participation, time-of-day, mode and location of intermediate (chained) stops on all tours of an individual in a day. The big weakness of the SACOG model system, relative to MORPC and NYMTC, is that it excludes explicit integration of the day activity schedule model components across household members.

Discussion

In this section we shift from examining the integration features of particular model systems to identify a set of integration features that spans the features seen in the reviewed models and also includes a few features that none of the models has yet incorporated. Table 1 lists detailed features within seven major categories, identifying for each feature the model systems that include it. Although all the categories are important, some seem more crucial than others, so we list them in a suggested priority order.

1. **Integration among destination, mode and time of a tour**. Effectively integrating the models of mode and destination choice has long been an important objective in trip-based models. One of the great desired strengths of the activity-based framework is the realistic incorporation of time-of-day modeling into the model system. This expands the focus of tour model integration to include time-of-day choice with mode and destination. It is important to move beyond the use of fixed assumed times in the calculation of mode and destination choice models, and in the calculation of logsums that are used as accessibility measures in higher level models. Doing so will enable the models to more realistically capture how time-specific policies affect choices other than time-of-day, such as mode, destination, and tour generation. In summary, this **is important because it helps the model capture the effects of time-specific policies on all dimensions of choice**.
2. **A horizontally integrated person-day activity pattern model with purpose-specific information about the tours and intermediate stops in the day.** This horizontal integration helps the model system to realistically represent the total amount and mix of activity and travel carried out by a person in one day. Importantly, it is a pre-requisite for additional features that improve the realism of these model predictions. In particular, it allows effective upward integration from purpose-specific tour and stop models. This **allows changes in transport conditions to affect overall tour and stop generation, as well as trade-offs between tour and stop generation**. Including activity purpose information in the pattern model enables it to capture the impact of purpose-specific changes in accessibility on pattern choice. For example, improved work accessibility can have different impact than improved accessibility for shopping.

Table 1: Integration Features of Existing Activity-Based Model Systems (a 'y' indicates that a model system has a particular feature)

Integration Feature	Priority	SFCTA	NYMTC	MORPC	SACOG
Integration among destination, mode and time of a tour					
—downward	1	y	y	y	y
—upward using assumed times		y	y	y	y
—upward accounting for available times	1				y
Horizontal integration of tour and stop generation in a day **for a person**					
—downward integration among tours			y	y	y
—downward among tours and stops					y
—horizontal among tours	2	y			y
—horizontal among tours and stops simultaneously	2	y			y
—horizontal purpose-specific among tours and stops	2				y
Upward integration: Transport conditions influence **tour generation**	3				
—upward integration in very few cases				y	
—upward integration in most or all cases	3	y			y
—accounting for differences among persons, available destinations and available modes	3				y
—accounting for available times	3				
Upward integration: Transport conditions influence **generation of trip chains**	3	y			y
—via intermediate stop logsums	3				y
—accounting for differences among persons and available destinations	3				y
—accounting for available modes and times	3				
Downward integration of tour and stop details in a day	4	y	y	y	y
—accounting for time used on mandatory tours	4			y	y
—accounting for time used on all trips and tours	4				y
—accounting for stop purposes	4				y
Integration between long-term and within-day models	5				
—downward from long-term to	5	y	y	y	y

within-day models					
—included auto ownership model	5	y	y	y	y
—included usual work location model	5	y			y
—upward to long-term models via mode choice logsums	5	y			y
—included usual school location model	5				y
Downward integration related to the use of vehicles from auto ownership model through tour and stop models	6				
—accounting for the use of each household vehicle	6				
—with vehicle type in long-term and mode choice models	6				
Integration of tours and stops in a day **for a household**	7				
—upward integration: transport conditions influence all modeled tour and stop generation	7				
—horizontal among persons for joint tours	7			y	
—horizontal among persons for day pattern	7				
—horizontal among persons for maintenance stops	7			y	
—among persons for escorting and shared trips	7				
—downward among persons for staying at home all day	7		y	y	
—downward among persons for additional aspects	7				

3. **Upward integration to the activity pattern model from both tour and intermediate stop models via logsums that account for important differences among persons (especially income, car availability and driving age) and that also account for available modes, destinations and times of day.** Accounting in logsums for available modes, destinations and times of day makes the pattern model sensitive to policies that apply differently across these three dimensions. For example, with this type of upward integration, a peak period toll on a major commute corridor will have an effect on pattern choice, as will a peak period transit improvement or a highway improvement that affects all time periods; the nature and magnitude of the effects will be governed by the probability proportions of the models used to calculate the logsums. Accounting for important

differences among persons enables the model to capture changing aggregate effects when population demographics and long term choices change, and helps the model to more accurately capture differences in how policies affect various population segments. For example, highway improvements help households with cars more than they help households without cars. Accounting for tour and intermediate stop accessibility enables the pattern model to better capture trade-offs between tour and intermediate stop generation. For example, improvements along a person's commute route should increase the likelihood that they will make intermediate stops on their way to or from work, whereas improvements away from their commute route should instead increase their likelihood of making additional tours. In summary, this **makes the sensitivity of pattern level changes to transport conditions and demographic changes more realistic**.

4. **Extensive and consistent downward integration of the within-day models: from activity pattern to tour models and then to all intermediate stops within each tour, including participation, purpose, location, travel mode and timing of all travel.** We think that a model sequence proceeding from day to tour to intermediate stop provides a reasonably realistic modeling sequence based on purposeful human planning. For example, it seems realistic to assume that the exact number and location of intermediate stops on a work tour should depend directly on the work location and the main mode used to get to work, even if the stops occur on the way to work. Careful downward integration helps assure that all modeled aspects of the entire modeled outcome are mutually consistent. For example, the use of time window accounting for each person can prevent the model from scheduling two of their tours to occur at the same time. Similarly, time window accounting for each household vehicle could be used to prevent the model from having two different household members use auto drive mode for separate tours if the household has only one vehicle. Extending the downward integration to include all intermediate stops on all tours would also yield travel itineraries that might eventually integrate effectively with traffic simulation models because of their realism. Finally, downward integration of purpose-specific choices can enable the model system to capture the correlation between activity choice and travel conditions, when combined with purpose-specific upper model levels and upward integration of purpose-specific tour and trip effects into the upper levels of the model. Such features can move the model system closer to being truly activity-based. This type of accounting does not directly improve the policy responsiveness of the model system. However, it **imposes constraints that should improve the accuracy of the policy responsive aspects of the model system, in particular sensitivity to transport conditions** that cause people to travel more or less, or to change travel modes, destinations and/or times of day.
5. **Integration that conditions within-day choices upon long-term choices (downward), and incorporates the effects of short-term opportunities**

and conditions on the long-term choices (upward). The long-term choices include residential location, work location, school location, auto ownership, and possibly also vehicle type, usual mode to work, usual mode to school, and transit pass ownership. An important benefit of conditioning within-day choices upon long-term choices is that the day activity pattern and other within-day models can directly use information related to the long-term outcomes. For example, the person activity pattern model can be influenced by the mode choice logsum associated with a tour to the usual work place. As another example, the location choice model of a household adult with an escort tour or stop can assign higher probability to the usual work and school locations of other household members. Thus, the use of long-term models with downward integration to the within-day models can make the within-day models more realistically policy responsive. In other words, the short-term elasticities of the model should be more realistic. There is a danger in this, however. Conditioning the day activity schedule on long-term outcomes without making the long-term outcomes correctly responsive to policy changes could severely bias the model predictions. This is because some long-term decisions, such as usual mode to work, can be much more elastic to travel conditions than their within-day choice of how to get to work today. Thus, any modeling of long-term choices should be accompanied not only by downward integration, but also by upward integration that captures the effect of travel conditions on the long-term choice itself, yielding accurate long-term elasticities. Doing this **should enable the model system to more realistically distinguish between long-term and short-term responses to policies**. To represent short-term responses to a policy scenario, the long-term model outcomes can be held fixed, and to represent long-term responses, they can be allowed to change in response to the policy.

6. **Downward integration related to the use of vehicles, from auto ownership model down through tour and stop models.** This feature, which is not included in any of the reviewed model systems, **should enhance the ability to forecast air quality impacts and fuel consumption of alternative future scenarios**. To achieve this benefit, the vehicle type would need to be modeled along with vehicle ownership. Vehicle types would need to be defined so as to be useful for policy analysis, while at the same time distinguishing types that represent realistic differences that matter when households acquire vehicles. The model system would also need to include the choice of tour vehicle for each auto driver tour and a full accounting of household vehicle use by time of day would need to occur, enforcing time-space constraints on every vehicle in the fleet. Doing this would enable a specific vehicle to be assigned to each vehicle trip, significantly improving the ability to provide information for air quality analysis.
7. **Integration of activities, tours and stops in a day for a household.** This category includes horizontal and downward integration for joint and correlated outcomes, such as coordinated day patterns, joint tours, household maintenance

tours, and shared trips such as escort trips. Importantly, it also includes vertical integration, so that the benefits of upward vertical integration identified in features one through five above are preserved.

A horizontally integrated model of household joint tour generation would capture the tendency of persons in a household to conduct activities and associated travel together. A potential advantage is that the conditional models of destination, mode and timing for joint tours might differ from those for individual tours, and the generation of joint tours might be more or less sensitive to transport conditions than the generation of individual tours. In order to achieve these benefits, the joint tour generation model would need to be effectively downwardly integrated with the person models, so that each person's day activity pattern would include the joint tour and their individual tours would not conflict with the joint tours. Just as importantly, the joint tour generation model would need to be upwardly integrated via logsums from the tour mode, destination and timing models, so that it would be realistically sensitive to transport conditions.

A horizontally integrated household day pattern would simultaneously represent the major activity and travel choices of the day for all members of a household, especially whether they traveled at all during the day, whether they traveled to work or school, and perhaps whether they worked at home. It could thus naturally extend the day pattern approach to encompass the entire household. Starting at the top, the basic within-day hierarchy would become household-day > person-day > tour > trip. The most tangible advantage of integrating the person-day models in this way is that it would yield more realistic household day patterns, capturing tendencies for persons in a household to coordinate their schedules. For example, in two-worker households without children, workers might be inclined to work on the same days, whereas in two-worker households with children they might be inclined to not work on the same days, or to stay home when a child stays home. However, to our knowledge, there is no current evidence that modeling the household day pattern would make the model system more accurately sensitive to transport prices and policies. Furthermore, using the household day pattern prevents all the major aspects of pattern choice for a given person from being horizontally integrated in the person day activity pattern. For example, the choice of whether to conduct a work tour is separated from the choices related to tours for other purposes. This has the potential of reducing the realism of the activity pattern's response to changes in accessibility. Therefore it is not clear whether the household day pattern would improve the overall model performance. In order to maximize the benefits of the household-day horizontal integration and minimize the problems caused by breaking the person-day horizontal integration it would be important to include upward integration to the household day pattern model from tour models via logsums that account for important differences among persons (especially income, car availability and driving age) and that also account for available modes, destinations and times of day (see discussion above related to upward integration to the person-day activity pattern model). This would help the household day pattern model more realistically respond to transport conditions.

Downward integration that captures correlations in activity and travel decisions among household members would involve conditioning day activity pattern, tour and trip choices of a person in a household upon earlier modeled outcomes of other persons in the household. It would depend heavily on the choice of run order of the model components among household members. The benefit is that it would increase the consistency of the modeled outcomes among the household members, capturing natural intra-household correlations in those outcomes. However, it is not clear that this would improve the accuracy of the model system's policy responsiveness, because the coordination of activity and travel choices (such as staying home with a sick child) may not be sensitive to transport conditions. Furthermore, vertically integrating among household members may make it more difficult to accurately model policy responsiveness of each person's activity and travel choices.

Integrating households in the models has been strongly advocated by academics because of the significant influence of the household on individual behavior. A clear benefit of this type of integration is that the predicted itineraries of household members would be much more realistic when viewed from the household perspective. For example, without this type of integration, a household with young children could easily be predicted to send all the adults to work and leave a small child at home alone. Or in a household with two adults and one car, one of them might be predicted to drive alone to work, and the other to ride as a passenger to work. Other potential benefits of explicit household integration in the models would occur if the integration caused the model to behave differently, and more accurately, to changes in demographics, land use or travel conditions. This would be the case if, for example, increasing real estate prices and declining incomes lead to more large non-family households, and modeling the activity and travel of such a household as a unit yields more or less travel than modeling the activities of its members separately.

However, we place the explicit integration of household behavior at the bottom of the priority list for several reasons. First, we think that the benefit of more realistic sensitivity to travel conditions brought by effective upward integration is of utmost importance. Second, introducing explicit household interactions makes it more difficult to implement the needed upward integration. Third, even if such upward integration could be implemented in a model with extensive household integration, it is not clear that doing so would substantially improve the quality over a well-integrated individual model that includes household effects indirectly through the use of household characteristics in the model equations. In summary, we think that the implementation of household integration **is important because of its ability to improve the realism of predicted household schedules and its potential to improve the model's responsiveness to changes in demographics, land use or travel conditions, but it should only be implemented without sacrificing the effective integration of features we identify as higher priorities.**

In this paper we have discussed integration principles that have guided the development of AB models now in use in the United States, described and critiqued the integration techniques employed by these models, and listed the techniques that we think are important, in order of importance, based on our experience and judgment about how the techniques would work in real-world model systems.

We reiterate that our purpose is not to select winners and losers among the reviewed models. Each one has made important innovative contributions to the state of the knowledge and practice in travel demand modeling. Rather, we hope that this discussion increases the awareness of the important topic of model system integration, focuses the issues, and stimulates further thought, discussion and research that may lead to the development of improved integration techniques in AB models.

Bibliography

Bowman, John. L. (1995) Activity based travel demand model system with daily activity schedules, Master of Science Thesis in Transportation, Massachusetts Institute of Technology, 92 pages. Available at http://jbowman.net.

Bowman, J. L., Ben-Akiva, M. (2001) Activity-based disaggregate travel demand model system with activity schedules, *Transportation Research Part A*, **35** 1-28. Available at http://jbowman.net.

Bowman, John L. and Mark Bradley (2006), Activity-Based Travel Forecasting Model for SACOG: Technical Memo Number 6: Day Pattern Activity Generation Models, July 31, 2006 – Draft 2, 21 pages. Available at http://jbowman.net.

Bradley, Mark A., John L. Bowman and Bruce Griesenbeck (2006) Development and application of the SACSIM activity-based model system, presented at the 11th World Conference on Transport Research, June, 2007, Berkeley, California, USA. Available at http://jbowman.net.

Parsons Brinckerhoff Quade & Douglas, Inc. (2005) New York Metropolitan Transportation Council, The Metropolitan Planning Organization, Transportation Models and Data Initiative, General Final Report: New York Best Practice Model (NYBPM), 541 pages.

PB Consult Inc. and Parsons Brinckerhoff (2002) Task 3.1A Modeling Individual Daily Activity Pattern: Statistical Evidence and Approach for MORPC (project task deliverable report), 42 pages.

PB Consult Inc. and Parsons Brinckerhoff (2005) The MORPC Travel Demand Model: Validation and Final Report, 237 pages.

San Francisco County Transportation Authority and Cambridge Systematics, Inc. (2002a) San Francisco Travel Demand Forecasting Model Development Final Report: Executive Summary, 39 pages. Available at http://www.sfcta.org.

San Francisco County Transportation Authority and Cambridge Systematics, Inc. (2002b) San Francisco Travel Demand Forecasting Model Development Final Report: Tour Generation & Time of Day Models, 44 pages. Available at http://www.sfcta.org.

Vovsha, P., Bradley, M. A., Bowman, J. L. (2005) Activity-based travel forecasting models in the United States: Progress since 1995 and Prospects for the Future, in Progress in Activity-Based Analysis, H Timmermans, ed., Elsevier, pp 389-414. Pre-publication draft available at http://jbowman.net.