COMPREHENSIVE VALIDATION OF AN ACTIVITY-BASED MODEL – EXPERIENCES
FROM HOUSTON-GALVESTON AREA COUNCIL’S ACTIVITY-BASED MODEL
DEVELOPMENT

Marty Milkovits, Corresponding Author
Cambridge Systematics, Inc.
100 CambridgePark Drive, Suite 400
Cambridge, MA 02140
Phone: (617) 354-0167
Fax: (617) 354-1542
E-mail: mmilkovits@camsys.com

Arun Kuppam
Cambridge Systematics, Inc.
10415 Morado Circle, Building II, Suite 340
Austin, TX  78759
Phone: (512) 691-8503
Fax: (512) 691-3289
E-mail: akuppam@camsys.com

David Kurth
Cambridge Systematics, Inc.
999 18th Street, Suite 3000
Denver, CO 80202
Phone: (303) 357-4661
Fax: (303) 446-9111
E-mail: dkurth@camsys.com

Thomas Rossi
Cambridge Systematics, Inc.
100 CambridgePark Drive, Suite 400
Cambridge, MA 02140
Phone: (617) 354-0167
Fax: (617) 354-1542
E-mail: trossi@camsys.com

Word Count = 6112 (words) + 6 * 250 (tables and figures) = 7612

Prepared for presentation and possible publication
Transportation Research Board
94th Annual Meeting
Washington, D.C.
Submitted: August 1, 2014
ABSTRACT

Activity-based models (ABM) are very sophisticated and consist of a number of interrelated, discrete choice modeling components. That is, the upstream model components (long-term choice models) provide inputs to the person-level models (daily activity, intra-household interactions) which, in turn, provide inputs to tour-level models (destination, mode, time of day) and, finally, trip-level models. Due to the number of model components, the complexity of the various models, and the amount of information passed between models, there is a high chance of error propagation if all of the components are not properly calibrated and validated. A rigorous calibration process is necessary for various levels – geographic (regional, sub-regional, county, or district), market segments (income, trip purpose), and household structures and person-types (gender, full time worker, etc).

This paper discusses the need for a detailed validation plan and lays out priorities for different tests to be performed. This paper focuses on a comprehensive methodology that was laid out for the Houston-Galveston Area Council (H-GAC) ABM development effort. This paper also presents example calibration results from the H-GAC ABM validation effort for several ABM components.

Keywords: Activity-Based Models, Validation, Travel Demand Forecasting
INTRODUCTION

The Houston-Galveston Area Council (H-GAC) is the Metropolitan Planning Organization for the Houston region. Since the 1980s, H-GAC has maintained trip-based travel demand models for the Houston region in partnership with the Texas Department of Transportation and the Metropolitan Transit Authority of Harris County (Metro). In late 2010, the H-GAC selected Cambridge Systematics to develop an activity-based model (ABM) for the region along with the specialized software required to implement the models.

The H-GAC’s desired features and capabilities for the new ABM included goals such as explicit consideration of relevant individual and household characteristics and travel behavior, the effect of intra-household decisions on traveler behavior, sensitivity to both short and long-term travel choices and household characteristics, explicit consideration of time of day in the determination of the existence and nature of tours/stops and modal choices, and the capability to explicitly account for and evaluate the impacts of transportation system characteristics, improvements, and/or policies on specific population subgroups.

The H-GAC ABM includes several enhancements derived from some of the ABMs implemented recently in other metropolitan areas. H-GAC’s ABM simulates the daily activity patterns of individuals in a “synthetic population” (representing all persons residing in Metropolitan Houston); determination of each worker’s usual workplace and each student’s school location in relation to the home location; modeling daily activity patterns and converting activity locations into tours; modeling times of day, destinations, and modes of tours and trips; modeling intra-household interactions and joint travel; and the use of conventional static highway and transit assignment procedures. The model design is described in detail in the Model Design Plan document prepared by Cambridge Systematics, Inc. for H-GAC at the beginning of this project (1).

The ABM is designed to produce accurate results for project analyses. It is also designed to provide improved sensitivity to transportation system changes, policy testing, and traveler characteristics over that of a conventional model since it models interactions among a wide range of variables and traveler characteristics. The important planning needs that H-GAC and METRO often deal with include regional transportation plans, transit-related ridership forecasting and alternative analyses, FTA New/Small starts, air quality and conformity analyses and corridor and subarea planning studies. All these can be analyzed with the existing trip-based model, but the new ABM can also provide more information about the effects of more sophisticated policies such as reversible HOT lanes, environmental justice and Title 6 analysis, land use impacts and sustainable livable centers, as well as transportation policy questions, on a variety of population segments.

While the provision of forecasting capabilities such as those outlined above is, in itself, a worthwhile goal, H-GAC recognized the importance of demonstrating that the added model sensitivities are, in fact, reasonable and valid. Simply providing the added sensitivities does not preclude the potential for compensating modeling errors or the need for unnecessarily large adjustment factors in some model components due to the cascading impacts of model specification or calibration errors in previous model components. This paper describes the calibration/validation process used to explicitly test the reasonableness of the added model sensitivities and calibrate/validate all of the specific model components to minimize compensating modeling errors and model adjustments due to cascading impacts of model specification or calibration errors.

This paper first presents a literature review of validation practices for ABMs. Next, the structure and operation of H-GAC’s ABM is described. The paper then discusses the overall H-GAC ABM validation process and provides details and examples from the single-pass validation step. Finally, some of the challenges to validating ABMs are discussed.

LITERATURE REVIEW

ABMs require much more validation than traditional trip-based models owing to their sheer number of model components feeding data into one another. This poses a steep challenge of dealing with and minimizing instances of compensating errors that are often undiscoverable in aggregate models, but more obvious in disaggregate models such as ABMs (2). This gives rise to an opportunity to conduct more
rigorous model calibration and validation as well as examine the sensitivity of an ABM in response to a whole host of explanatory variables.

The ABM validation process is faced with a unique hurdle – the lack of well-established validation procedures or guidelines. The 2010 Travel Model Improvement Program Travel Model Validation and Reasonableness Testing Manual – Second Edition (3) includes discussions of possible ABM validation and reasonableness tests. However, many of the guidelines regarding possible ABM tests are adaptations of typical tests performed for aggregate, four-step travel models. While the development and use of ABMs is becoming more prevalent for urban areas, there is still a relative lack of experience with the models. There is no real guidance on how close the modeled results should be to observed values. The use of common trip-based model validation standards, such as acceptable error ranges in aggregate model statistics (e.g., vehicle miles traveled), is appropriate for ABM system validation, but such tests have little meaning if the disaggregate models feeding the final assignments are not well calibrated and validated. In addition, the validation burden for individual model components is increased if improved sensitivity to system characteristics, policy variables, and traveler characteristics is to be demonstrated. Some examples of validation procedures used for ABMs are presented below.

The DRCOG ABM validation included a wide range of tests including: (a) checks to ensure that each model component was producing correct results (i.e., verification of computations); (b) disaggregate validation of all model components estimated through the application of the model component against the estimation data set and comparison of the model outputs to the estimation data; (c) testing of each model’s sensitivity to variables through controlled modification of those input variables; (d) comparisons of each model component outputs to summaries from expanded survey data; and (e) comparisons, where data were available, of base year outputs from each model component to independent observed data (e.g., comparisons of mode choice model outputs to linked trips estimated from transit boarding counts) (4).

As part of SACOG’s ABM development, each of the SACSIM components were calibrated at different geographic levels (e.g., TAZs, districts) depending upon the prominence and importance of each component (5). The calibration targets and thresholds were set by SACOG staff, taking into account several factors, including the importance of the variable or metric, the level of confidence and certainty in the data source, and internal consistency of the calibration data. In addition to using the household travel survey data for calibration, ACS data were also used for the calibration of some SACSIM components such as the auto availability model.

In a research paper presented at the 2006 Innovations in Travel Modeling (ITM) Conference, Pendyala and Bhat noted that the quality of a travel demand model system, in the context of an ABM, is better judged based on an assessment of its ability to respond to a range of scenarios and policies of interest (6). The paper presented a list of scenarios that included measuring changes in land use, socio-economic and demographic characteristics, multimodal network characteristics, implementation of transportation polices (e.g., parking pricing, congestion pricing, telecommuting), impacts of technologies and their interactions with travel behavior, changes to spatial and temporal resolutions, and accommodation of emerging behavioral paradigms (i.e., household inter-dependencies, constraints and flexibility, positive utility of travel, alternate behavioral processes and decision rules).

The estimation of an ABM’s discrete choice model components require information on the alternatives available to individual travelers and the choices those travelers make in response to the characteristics of the alternatives available. The estimation dataset relies heavily on auto and transit skim information. This skim information should be validated thoroughly before it is used for model estimation. The authors have previously pioneered the use of prediction-success tables of modeled versus reported travel attributes, especially for ABMs (7).

For H-GAC’s ABM, CS conducted disaggregate validation tests of modeled skim data that involved a comparison of modeled versus reported skim distributions for peak and off-peak periods as well as for auto and transit modes. This resulted in improving the modeled skim data, eliminating reported travel data (from the estimation datasets) with unresolved, illogical differences between modeled and reported travel impedances, and quantifying errors in modeled data use for model estimation. Similar tests
have been performed for other ABM development efforts for the Minneapolis-St. Paul and Baltimore regions.

Another key input to the ABM system is the disaggregate socio-economic information from population synthesizers. These synthesizers also need to be validated thoroughly to ensure accurate representation of the underlying population. The Chicago Metropolitan Agency for Planning validated its population synthesizer to marginal distributions across four household dimensions – size, number of workers, income group and age of household head (8).

A few MPOs have also compared results from their new ABMs to results from their existing trip-based models. The San Francisco County Transportation Authority’s SF-CHAMP ABM was compared with the regional trip-based model and, as expected, the differences in the base year model were minimal since they were both validated to the same observed survey dataset (9). However, while the trip generation characteristics in the forecast year (2030) were comparable, there were some notable differences in trip distributions and modes of travel.

**H-GAC ABM MODEL STRUCTURE**

Figure 1 shows the H-GAC ABM disaggregate demand model sequence as implemented in TourCast. TourCast takes inputs from a population synthesizer and skimming procedures and produces aggregate trip tables for assignment. This paper is focused on the single pass validation of the demand models, which are discussed in more detail below.

**FIGURE 1 H-GAC ABM Model Structure**

**Long-Term Choices**

Auto Ownership, Work Location, School Location

**Tour Generation**

Daily Activity Pattern (Including Work/School Travel)

School Escorting Model

Fully Joint Travel

Individual Nonmandatory Travel

**Tour-Level Choices**

Mandatory Tour Destination & Time of Day

Joint Tour Destination & Time of Day

Individual Nonmandatory Tour Destination & Time of Day

All Tour Stop Generation & Mode Choice

**Stop/Trip-Level Choices**

Stop (Trip) Level Destination, Time of Day, and Mode Choice

**Long-Term Choices**

Long term models include the vehicle availability, usual work location choice, and school location choice models. These choices are modeled and applied first in the ABM sequence.

**Vehicle Availability Model**

The vehicle availability model predicts the number of motorized vehicles owned, leased, or belonging to the fleet of vehicles possessed by each household. It considers household characteristics as forecast by the population synthesizer along with transit and walk accessibilities to different activities.
Workplace Location Choice Model
This model has two parts. First, a binary choice model predicts, for every employed person in the synthetic population, whether he or she has a usual work location. Second, a destination choice model predicts the usual workplace zone, if there is one. This model takes into account accessibility measures (i.e., distance) from home to work; location measures (i.e., home and work in same district or home in suburb and work in downtown), and interactions of the labor force and employment characteristics (i.e., worker’s occupation and employment by sector specified in size functions).

School Location Choice Model
The school location choice model assigns a school location zone to every student in the synthetic population. As expected, the survey data showed that all students have school purpose destinations as their regular school locations, and so the model excludes “no regular location” from the choice set. These models take into account accessibility measures from home to school (i.e., distance); location measures (i.e., school districts); and include interaction of siblings (i.e., siblings close in age tend to go to the same school). The school location choice model has fewer variables than other destination choice models since distance to school or school district boundaries determine the school location for most students.

Tour Generation
Tour generation models include the daily activity pattern (DAP), school escorting, fully joint, and individual non-mandatory models. The DAP model is applied next after the long-term models, but mandatory tour-level models are then run on the generated DAPs before school escorting tours are generated. Similarly, the fully joint tour-level models are run before the individual non-mandatory tours are generated.

Daily Activity Pattern
These models predict every person’s daily activity pattern, including the number of mandatory tours and stops, for each person type. These models capture the tradeoffs among activities, for example, full-time workers tend to make fewer (and shorter) tours for non-work related travel purposes while part-time workers and nonworking adults make more maintenance tours. The following combinations of tours were most common in the survey data and are the DAP alternatives:

- One Work Tour No Stops
- One Work Tour With Stops
- Two Work Tours No Stops
- Two Work Tours Stops On One
- Two Work Tours Stops On Both
- One University One Work Tour No Stops
- One University One Work Tour Stops On Work Tour
- One School One Work Tour No Stops
- One School One Work Tour Stops On Work Tour
- One University Tour
- Two University Tours
- One School Tour
- Two School Tours
- Non Mandatory Travel
- Stay At Home
- Out Of Area
- External Travel Only

School Escorting Model
The school escorting model simulates whether a child is escorted by another household member on either half of their school tour and, if so, by whom. Potential chauffeurs are split into two groups. The first are adults that could perform the escorting task as a stop on their own mandatory tour (e.g., drop child off at school on the way to work). The second are adults that could perform the escorting task as a stand-alone tour.

There are a number of availability conditions that must be processed in order to generate the choice set for each school tour. For adults with mandatory tours, the timing of the mandatory tour (which is simulated prior to school escorting) must meet two conditions for the adult to be considered available for escorting as part of the mandatory tour. First, the adult’s mandatory activity must begin after the school activity (on outbound journey to school) or end before the school activity ends (on inbound journey from school). If this condition is not met, the adult is not considered a viable escort candidate. Second, the adult’s mandatory activity must begin or end within 3 half-hour periods of the start or end of the school tour. If this condition is not met, the adult can still be considered a viable escort candidate, but only in the stand-alone school escorting category, not as part of the adult’s mandatory tour. Only if both conditions are met is the adult considered a candidate for school escorting as part of the mandatory tour.

**Fully Joint Travel**

A fully joint tour is one where all segments of the tour is shared across two or more individuals from the same household, including tour purpose, number and purpose of stops, activity location for each stop, timing of each stop, and mode choice for each trip made on the tour. The H-GAC ABM models explicitly only joint tours with non-mandatory purposes.

The fully joint tour model consists of two submodels: tour generation at the household level (which simulates the number and purpose of joint tours for a household) and tour participation at the individual level (which simulates which household members will participate in each tour). The joint tour generation model is estimated on and applied for all households that meet the condition of two or more household members with an active DAP (i.e., a mandatory or non-mandatory travel pattern). The participation model simulates participation in each generated fully joint tour for all active DAP household members where more than three household members have an active DAP (if there are only two household members have an active DAP then, by definition, they must both participate in all fully joint tours for that household).

**Individual Non-Mandatory Travel**

Individual non-mandatory (INM) travel represents demand for activities such as meal, shopping, social-recreation, personal business, or non-school escorting. This travel activity is considered to be a lower priority than the mandatory work and school tours therefore it is modeled after those tours have been generated and their times-of-day simulated. The INM travel model predicts the number and purpose of non-mandatory tours for each person that does not have a Stay at Home, Out of Area, or External Travel Only DAP based on person and household characteristics and the time windows available.

**Tour Level Choices**

Tour level models include destination choice, time of day choice, intermediate stop generation, and mode choice for all tour types (mandatory, fully joint, and INM).

**Tour Destination**

The first tour-level model run is destination choice, which predicts the activity location for each tour type, except school tours, which are constrained to be located at the school location simulated in the long term model. Note that, although the person may have a usual work location, work tours are not constrained to the usual work location, although it is a likely outcome.

The destination choice model alternatives are every zone in the region and the models are applied independently across persons without constraint, i.e. there are no controls over how many persons choose a particular destination zone.
The destination choice model takes as inputs person and household attributes, attributes of the destination zone, and the impedance between the home and destination zones. The impedance is represented through the logsum of the mode choice model.

Tour Time-of-Day
The tour level time of day choice models predict joint arrival and departure times at the primary activity location in 30-minute intervals (48 periods across the day) beginning and ending at 3:00 AM. Thus, for any given arrival period there are multiple departure periods available, except the last period of the day when there is the same period available for departure. Each arrival/return pair represents an alternative, for a total of 1,176 alternatives.

The model involved using alternative specific constants for various groupings of arrival periods, departure periods, and durations, plus shift effects that push arrivals earlier or later and durations of stay longer or shorter (as a result, shifting the departure period as well). The models also make extensive use of the concept of available time windows for scheduling tours. Tour times of day are simulated sequentially and, as each tour is simulated, periods within a tour are not available for other home-based tours.

Stop Generation
The stop generation model predicts the number and purpose of stops on each half-tour (outbound and inbound). For mandatory tours, the presence of stops on one or both legs of the tour is simulated by the DAP. Fully joint and individual non-mandatory tours may also have stops on either or both legs of each tour. Stops are generated depending on person and household attributes and on the primary tour attributes, including tour purpose, time of day, and duration as well as the DAP.

Tour Mode Choice
The mode choice model is not run until after the tour destination, time of day, and stop models so that the information from these models can be used in the mode choice utilities. The H-GAC ABM model includes 8 mode alternatives: three auto; two transit; two non-motorized; and school bus, which is only available for school tours.

The predicted tour mode is not necessarily the mode used on every trip of the tour, instead the tour mode determines which modes are available for each trip of the tour. This ensures that the trip chains have a reasonable sequence of modes (e.g. a Shared Ride 2 tour mode precludes use of transit on any trip).

Work-Based Sub-Tour Models
All of the tour models discussed to this point are home-based tours, meaning that the person departs from and returns to the home location. The H-GAC ABM allows for one type of sub-tour, work-based sub-tours, where the person departs and returns to the work location. Work-based sub-tours must occur within a work tour, thus they cannot be modeled until after the work tours are simulated. The work-based sub-tour models follow the same paradigm as other tours: generation, destination, time of day, stop generation, and mode choice.

Stop / Trip Level Choices
The stop and trip level choice models simulate the destination and time of day for intermediate stops and the mode choice for each specific trip. Each of these models are tightly-coupled with the results of the tour-level models.

At the conclusion of the stop / trip level choices, a trip roster with mode, origin/destination, time of day, and person attributes is available. The H-GAC ABM then aggregates the trips by zone and time period for static assignment.

VALIDATION PROCESS
Based on the number, complexity and interdependency of models in an ABM, it is evident that the validation process of an ABM must be much more elaborate, comprehensive and data intensive than that of
a four-step trip-based model system. For the H-GAC ABM validation, CS has executed the comprehensive model validation plan (10) outlined below.

Validation Data Compilation
The first step of this process is data assembly and assessment. This includes reviewing and summarizing all kinds of travel survey data (household travel, transit, etc.), count data (by facility type, vehicle class, transit boardings, etc.), speed data, and other publicly available data (census, ACS, etc.). This review assured us that there was sufficient data to validate the ABM system for the eight-county Houston-Galveston region. It also provided us with a logical way of compiling data for the various ABM components, and helped us set up validation targets for various model components.

Operational Checks
Many of the models are complex with a large number of parameters and alternatives, so it is important to verify correct operation of each model before calibration.

Initially, basic tests were conducted for all ABM components to provide enough insights into the performance of each component. Then, specific “debugging” tests were run to provide more information at disaggregate levels and reveal any anomalies found in the basic tests.

Population Synthesis Validation
The population synthesis results for the base year (2010) were validated prior to starting single pass calibration of the model components. While the population synthesis procedures were outside of the ABM, inaccuracies in the population synthesis results could have cascaded through and, possibly, compounded within the ABM.

Since the ABM uses information on the structures of individual households such as household income, number of household members, and age and employment characteristics of household members, it is important that distributions for these characteristics be reasonable throughout the region. It is insufficient for marginal distributions, such as number of households by household size for the number of households by income group, to be reasonable for the region and small areas within the region. Rather, joint distributions, such as a number of households by household size and number of workers, must be reasonable.

Single-Pass Calibration of Model Components
First, the estimated models are applied in a single pass for a base year using the validated congested speed network data from the model estimation process. Subsequent models are dependent on prior model outputs so the single-pass calibration is conducted in the sequence that the models are implemented. For each model, the following actions are taken:

- Execute the model using synthesized population and loaded network skim data and compare model result summaries against observed data.
- Calibrate the model by adjusting model parameters and also re-estimating certain models to include new parameters, wherever necessary. Model results are compared to observed data with multiple stratifications of socioeconomic, demographic, and geographic variables.
- Examine subsequent model sensitivity to calibration. As all the ABM components are linked to one another and applied in sequence, every subsequent model component is affected by models upstream. So doing a single-pass calibration gives insight into the magnitude and direction of error propagation through the model system.

Full Feedback Validation
After completing a single-pass calibration, the loaded networks from the ABM are skimmed and fed back into the model. Feedback loops ensure that the network state that is input to the demand models is comparable to the network state produced by the output of the demand models. The previous model network state is used in the single-pass calibration to “prime the pump”, but the network state produced by
the ABM may have systematic differences, e.g. by time of day, that need to be understood. Nearly every model in the ABM is sensitive to the network state, including the generation, destination, mode, timing and routing of tours and trips. After a number of feedback loops, the key results from each ABM component is compared against the observed (survey) data and additional calibration actions taken.

**Sensitivity Testing**

The next step is to perform a series of sensitivity tests consisting of a number of different scenarios designed to vary input parameters across multiple components. Sensitivity tests are meant to examine the variation in the model forecasts across plausible scenarios and how sensitive the forecasting model is to variations in level-of-service parameters and other variables. As such, these tests are focused on the most important factors that will have a direct impact on the projection of future travel behavior in the region, including the following:

- Socioeconomic and demographic factors:
  - Alternate growth rates of population, employment;
  - Alternate growth rates of different market segments such as aging of population, presence of more females in the workforce, increase in low income households, etc.

- Auto Mode Parameters:
  - Adjustments to fuel costs.

- Impact of new highway projects:
  - New managed lanes, or pricing scenarios;
  - Widening of highways.

- Impact of new transit projects:
  - Extension of rail lines;
  - Addition of new transit modes, e.g. bus rapid transit, streetcar, light rail.

**Elasticity Tests**

Similar to sensitivity testing, both direct and cross-elasticities are estimated for certain key explanatory variables.

**Backcast Test**

Finally, a backcast of the year 2000 is performed as a means to evaluate model performance against a known data set that was not used in model estimation or validation is performed.

**CALIBRATION AND VALIDATION RESULTS**

ABM validation is a major undertaking and each proposed test is assigned a priority describing its importance. Priorities are based on the following considerations:

- Level 1 priority tests include those tests that can be produced with relatively little effort and provide good measures of the validity of the models. This priority level includes aggregate validation measures like VMT and VHT summaries, tours per person and per household, and average tour durations and distances.

- Level 2 priority tests include more detailed and non-standard validation tests. These tests can be considered to be more directly focused on the tour-based models being developed for the ABM. Since the tests are more detailed and/or non-traditional, they may be more difficult to produce or interpret, and it may be more difficult to acquire data or comparable tests from other regions for comparison. In addition, these tests can lead to “information overload” (e.g. summarizing results for four income groups by five household sizes by eight districts produces two sets of 160 numbers for comparison). Innovative methods might be required to display and interpret the results.

- Level 3 priority tests include those that are likely to be costly or difficult to produce or interpret, including those tests that do not have readily available validation data.
Priorities were assigned to each validation test in order to maximize the return on the investment in the validation testing. Tests were selected that were expected to reveal the greatest insight and provide a comprehensive view into how the ABM functions. Each validation test may be compared against different stratification levels including geographic data (such as county, area type), household characteristics (such as income, size, and number of workers), and person attributes (such as age and gender). Stratifications may also be made based on simulated variables from prior models. Table 1 and Table 2 present the validation tests selected for the H-GAC ABM single pass. A sample selection of test results are provided later in this section. The table or figure number of the selected test results are indicated by the green shaded cells of Tables 1 and 2.
# Table 1 Single Pass Validation Tests: Population Synthesis and Long-Term Choices

<table>
<thead>
<tr>
<th>Model</th>
<th>Modeled Unit</th>
<th>Model Segmentation</th>
<th>Comparisons</th>
<th>regional</th>
<th>county</th>
<th>school district</th>
<th>area type</th>
<th>transit / walk accessibility</th>
<th>income</th>
<th>hh size</th>
<th>hh workers</th>
<th>hh vehicles</th>
<th>income x hh size</th>
<th>hh worker x hh size</th>
<th>hh worker x children</th>
<th>hh worker x vehicles</th>
<th>hh size x vehicles</th>
<th>person type</th>
<th>age</th>
<th>gender</th>
<th>age x gender</th>
<th>distance and transit attributes</th>
<th>stops per tour</th>
<th>tour duration</th>
<th>trip time of day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Synthesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Characteristics of individual households and persons</td>
<td>Households</td>
<td>Households by Household Income &amp; Household Size</td>
<td>x x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Households</td>
<td>Households by Household Income &amp; Number of Employed Household Members</td>
<td>x x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Persons</td>
<td>Persons by Age &amp; Sex</td>
<td>x x</td>
<td>x x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Persons</td>
<td>Persons by Employment Status &amp; Sex</td>
<td>x x</td>
<td>x x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Persons</td>
<td>Persons by Age by Person Type</td>
<td>x x</td>
<td>x x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long-Term Choices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vehicle Availability</td>
<td>Household</td>
<td>Frequency of vehicles per household</td>
<td>x x</td>
<td>x x x</td>
<td>x x x</td>
<td>x x x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Average vehicle per household</td>
<td>x x</td>
<td>x x x</td>
<td>T3</td>
<td>x x x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Autos per driver</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usual Workplace existence</td>
<td>Person</td>
<td>Frequency of usual workplaces</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usual Workplace location</td>
<td>Person</td>
<td>Trip length distributions (congested time and straight-line distance)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average trip lengths (congested time and straight-line distance)</td>
<td>x x x</td>
<td>x x x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intra-zonal percentages</td>
<td>x x x</td>
<td>x x x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Destination by socio-economic data</td>
<td>P2</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School Location</td>
<td>Person</td>
<td>Trip length distributions (congested time and straight-line distance)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average trip lengths (congested time and straight-line distance)</td>
<td>x x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intra-zonal percentages</td>
<td>x x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intra-district percentages</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2 Single Pass Validation Tests: Tour Generation, Tour-Level Choices, Stop / Trip-Level Choices

<table>
<thead>
<tr>
<th>Model</th>
<th>Modeled Unit</th>
<th>Modeled Segmentation</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Daily Activity Pattern</strong></td>
<td>Person</td>
<td>Person Type</td>
<td>Frequency of DAP</td>
</tr>
<tr>
<td><strong>School Escort</strong></td>
<td>Intra-Household</td>
<td></td>
<td>Escort activity by child age group</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Chauffer activity by person type</td>
</tr>
<tr>
<td><strong>Fully Joint Tour Generation / Participation</strong></td>
<td>Household / Person</td>
<td></td>
<td>Tours per household</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tour frequency by purpose and group size</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Persons by tour purpose</td>
</tr>
<tr>
<td><strong>Individual Non-Mandatory Tour Generation</strong></td>
<td>Person</td>
<td></td>
<td>Tour frequency by purpose</td>
</tr>
<tr>
<td><strong>Work Based Tour Generation</strong></td>
<td>Tour</td>
<td></td>
<td>Tour frequency by purpose</td>
</tr>
<tr>
<td><strong>Tour-Level Choices</strong></td>
<td>Tour</td>
<td>Tour Purpose</td>
<td>Trip length distributions</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Average trip lengths</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Intrastral percentages</td>
</tr>
<tr>
<td><strong>Tour Time of Day</strong></td>
<td>Tour</td>
<td>Tour Purpose</td>
<td>Arrival period frequency and coincidence ratio</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Departure period frequency and coincidence ratio</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Average duration</td>
</tr>
<tr>
<td><strong>Stop Destination</strong></td>
<td>Trip</td>
<td>Tour Purpose</td>
<td>Trip length distributions</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Average trip lengths</td>
</tr>
<tr>
<td><strong>Tour Stop Generation</strong></td>
<td>Half-Tour</td>
<td>Tour Purpose</td>
<td>Stop frequency by half-tour and time of day</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Stop frequency</td>
</tr>
<tr>
<td><strong>Tour Mode</strong></td>
<td>Tour</td>
<td>Tour Purpose</td>
<td>Mode frequency</td>
</tr>
<tr>
<td><strong>Stop/ Trip-Level Choices</strong></td>
<td>Trip</td>
<td>Time period frequency</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trip</td>
<td>Mode frequency</td>
<td></td>
</tr>
</tbody>
</table>
Table 3 compares the expanded household survey and modeled average vehicles per household by size and income group. Note that the model was not calibrated to match every segment exactly, although that would have been possible. For example, in the survey data, 2-person households have more autos than 3-person households (2.57 vs. 2.54). This is a counter-intuitive result and, instead of calibrating the model to match it, heat maps were used to reveal the general trends across segments and guide calibration.

**TABLE 3 Average Vehicles Per Household**

<table>
<thead>
<tr>
<th>HHIncome</th>
<th>HHSize</th>
<th>Total</th>
<th>&lt;$20000</th>
<th>20,000-39,999</th>
<th>40,000-69,999</th>
<th>70,000-99,999</th>
<th>&gt;$100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td></td>
<td>2.05</td>
<td>1.35</td>
<td>1.88</td>
<td>2.37</td>
<td>2.52</td>
<td>2.64</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1.13</td>
<td>0.88</td>
<td>1.19</td>
<td>1.29</td>
<td>1.43</td>
<td>1.48</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2.05</td>
<td>1.59</td>
<td>1.93</td>
<td>2.16</td>
<td>2.33</td>
<td>2.31</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>2.57</td>
<td>1.86</td>
<td>2.29</td>
<td>2.82</td>
<td>3.00</td>
<td>3.01</td>
</tr>
<tr>
<td>4+</td>
<td></td>
<td>2.54</td>
<td>1.79</td>
<td>2.37</td>
<td>2.80</td>
<td>2.80</td>
<td>2.92</td>
</tr>
</tbody>
</table>

**Model**

<table>
<thead>
<tr>
<th>HHIncome</th>
<th>HHSize</th>
<th>Total</th>
<th>&lt;$20000</th>
<th>20,000-39,999</th>
<th>40,000-69,999</th>
<th>70,000-99,999</th>
<th>&gt;$100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td></td>
<td>2.05</td>
<td>1.20</td>
<td>1.73</td>
<td>2.15</td>
<td>2.40</td>
<td>2.59</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1.13</td>
<td>0.85</td>
<td>1.13</td>
<td>1.33</td>
<td>1.39</td>
<td>1.47</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2.12</td>
<td>1.51</td>
<td>1.88</td>
<td>2.15</td>
<td>2.27</td>
<td>2.39</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>2.23</td>
<td>1.43</td>
<td>1.88</td>
<td>2.30</td>
<td>2.47</td>
<td>2.65</td>
</tr>
<tr>
<td>4+</td>
<td></td>
<td>2.59</td>
<td>1.61</td>
<td>2.15</td>
<td>2.66</td>
<td>2.85</td>
<td>3.00</td>
</tr>
</tbody>
</table>

**Percentage Difference (model - survey)/survey**

<table>
<thead>
<tr>
<th>HHSize</th>
<th>Total</th>
<th>&lt;$20000</th>
<th>20,000-39,999</th>
<th>40,000-69,999</th>
<th>70,000-99,999</th>
<th>&gt;$100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td></td>
<td>0%</td>
<td>-15%</td>
<td>-15%</td>
<td>-22%</td>
<td>-12%</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>3%</td>
<td>-3%</td>
<td>-6%</td>
<td>4%</td>
<td>-3%</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>7%</td>
<td>-8%</td>
<td>-5%</td>
<td>-1%</td>
<td>-6%</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>-34%</td>
<td>-43%</td>
<td>-41%</td>
<td>-52%</td>
<td>-53%</td>
</tr>
<tr>
<td>4+</td>
<td></td>
<td>5%</td>
<td>-18%</td>
<td>-22%</td>
<td>-15%</td>
<td>5%</td>
</tr>
</tbody>
</table>

Figure 2 shows the locations of usual workplaces divided by the total employment in each zone to identify cases where the simulated usual workplace locations differ from employment totals. While there are a few zones with more usual workplace locations than jobs, these are scattered in the more rural areas of the region; so the actual over prediction of usual workplace locations is small.
FIGURE 2  Comparison of Chosen Workplaces and Employment in the Workplace Location Choice Model

Tour time of day results are modeled and validated by tour purpose because there are systematic differences in the time of day profiles by tour purpose (Figure 3). Validation of these models includes a visual comparison of the model and survey time profiles and examination of the coincidence ratio for the arrival and departure distributions.
CALIBRATION AND VALIDATION CHALLENGES
The primary challenge of an ABM single-pass validation is to synthesize a lot of data and determine what warrants adjustments to the estimated parameters or a re-estimation with additional variables. In a single pass calibration it is expected that some adjustments to the estimated parameters are necessary because the
household survey data is not a true random sample and the survey expansion factors do not account for the high level of segmentation in an ABM. In this case, early tests revealed substantial discrepancies between the expanded household survey and the synthesized population. There were more controls in the synthesized population generation process than the household survey expansion process so the survey data was normalized to match the results of the synthesized population. This normalization process made it possible to compare model and survey results at aggregate levels.

The H-GAC model update included changes in the zone structure and network that occurred between model estimation and validation, which made calibration more imperative. However, parameter adjustments must be made with caution to avoid over fitting the model, particularly when the comparison segment is based on a small number of household records.

To address the challenge of few observations, heat maps (as shown in Table 2) were used to help reveal trends in the survey data and confirm that the model is reproducing reasonable trends. In some cases, a conditional formatting mask was used to de-emphasize comparisons based on a small number of survey records.

Another challenge to ABM calibration is the dependency that later models have on prior ones. School escorting is an example of a model that is dependent on not only the outputs, but the interactions of prior models. Specifically, the interactions between the work tour time of day and school tour time of day models determine if any adults would be available to escort children to and from school. In calibrating these models, it was not sufficient to fit the arrival and departure frequency distributions of work tours. Rather, it was critical that the arrival and departure times for work tours were sensitive to any school tours made by children in the same household. Calibrating an ABM requires that the impacts of any parameter adjustments on subsequent models be continually monitored.

CONCLUSIONS
ABMs, such as the ABM developed for the H-GAC region, have been developed to more realistically model traveler behavior and provide improved sensitivity to changes in model inputs such as system characteristics, household and person characteristics, and policies in a region. The improved sensitivity, however, increases the burden on model estimation, model calibration, and model validation.

This paper has demonstrated the approach used for the H-GAC ABM to directly address the increase in efforts to calibrate and validate the model. Specifically, the approach:

- used explicitly developed calibration/validation tests to be applied for each model component,
- applied those tests in a sequential, single step process, with requirement that each model component be reasonably calibrated/validated prior to starting calibration/validation of the subsequent model component, and,
- performed a full model system validation with feedback only after successfully completing the single step calibration/validation process.

The benefits of the approach demonstrated our improved understanding of model sensitivities and interactions, and improved confidence in the calibrated/validated ABM.

While the benefits of the approach far outweigh the additional burden placed on the calibration/validation process, it is important to note the following:

- information overload is a concern; innovative methods, such as graphical displays, may be required to gain a better understanding of the calibration/validation results for specific model components,
- the availability and statistical accuracy of observed data for the calibration/validation process decreases as the amount of stratification in the tests increases,
- substantial time and cost are required to fully implement quality model calibration/validation process like the stepwise process used for the H-GAC ABM, and,
- we are still building the experience necessary with ABMs to establish guidelines for the calibration/validation of specific ABM model components.
ACKNOWLEDGEMENTS
The authors would like to acknowledge the staff from H-GAC, Chris van Slyke, Chi-Ping Lam, Sharon Ju, Heng Wang, David Gao and Michael Onuogu for their extensive support with the development of the model networks, TAZ shapefiles and SED data, as well as providing their feedback throughout the modeling process. The authors would also like to acknowledge Vincent Sanders of Houston Metro for providing invaluable input to the model design process from a transit perspective. Andy Mullins of Texas Transportation Institute was also instrumental in assisting with the model design plan, reviewing model estimation results and providing feedback on the validation plan.

REFERENCES
8. PB Americas, Activity-Based Model for Highway Pricing Studies at CMAP., Prepared for Chicago Metropolitan Agency for Planning, October 2011.