**BMC REGIONAL TRAVEL DEMAND MODEL UPDATE:
DEVELOPMENT OF A SYNTHETIC POPULATION GENERATOR**

**TECHNICAL MEMORANDUM 1.2
PROJECT INITIATION PLAN**

Ram M. Pendyala and Karthik C. Konduri, Arizona State University, Tempe, AZ
Chandra R. Bhat, The University of Texas at Austin, Austin, TX
Konstadinos G. Goulias, The University of California at Santa Barbara, Santa Barbara, CA
Maren Outwater, Resource Systems Group, Inc., White River Junction, VT
Xin Ye and Fred Ducca, National Center for Smart Growth, University of Maryland, College Park, MD
Julie Dunbar, Dunbar Transportation Consulting, Bloomington, IL

December 2010

Table of Contents

SECTION 1. PROJECT INITIATION AND APPROACH TO SYNTHETIC POPULATION GENERATION

1. BACKGROUND

The Baltimore Metropolitan Council (BMC) has embarked on the development of a synthetic population generator as the first step in transitioning towards advanced activity- or tour-based travel demand model systems that are founded on the principle of microsimulation wherein the activity-travel patterns of individual agents are simulated in time and space. Although the current project is focused on the development of a synthetic population generator, it is also intended to provide a vision and framework for the development of an activity-based travel model system that would meet the planning needs and computational constraints of the agency.

This technical memorandum is intended to describe the various considerations that go into the design and formulation of a synthetic population generator and an activity-based travel microsimulation model system. It should be noted that this memo only provides some of the design considerations and criteria, but does not actually present the respective frameworks or procedures in detail. The detailed formulation, methodology, framework, and operational procedures for the synthetic population generator and the activity-based travel model system will be described in subsequent technical memoranda submitted at the end of Tasks 3 and 4. However, the information contained in this memo provides some initial considerations and ideas that will guide the formulation of methodologies underlying the respective model systems.

2. KICK-OFF MEETING AND WORKSHOP

A project kick-off meeting for the development of a synthetic population generator was held at the Baltimore Metropolitan Council (BMC) offices during November 9-10, 2010. The kick-off meeting was intended to provide an opportunity for the project team to meet with BMC staff and other travel model stakeholders (e.g., Maryland Highway Administration) and learn about the planning needs and challenges confronting planners in the region. Moreover, the project team was able to make presentations about synthetic population generation, review activity-based travel model systems, and provide initial thoughts on phased approaches to the implementation of activity-based model systems in practice.

A copy of the meeting agenda is provided at the end of this technical memorandum as Appendix A. The meeting started with a presentation on synthetic population generation by Ram Pendyala of Arizona State University. This presentation had several components including:

a) An overview of synthetic population generation and the motivation for developing and implementing synthetic population generators in the context of activity-based travel model systems

b) A review of synthetic population generation approaches with identification of the drawbacks of traditional procedures and recent approaches to remedy the drawbacks

c) A review of the synthetic population generation process that has been implemented within the PopGen open-source software package developed at Arizona State University

d) Identification of data required for synthetic population generation

e) Illustration of the application of PopGen to Maricopa County in Arizona and results obtained from the sample application

f) An intuitive interpretation and explanation of the enhanced algorithm implemented in PopGen for controlling both household and person level variables/dimensions of interest

g) Criteria for identifying appropriate household and person control variables and the categories for each variable

h) Methods and test-statistics employed within PopGen for testing the goodness-of-fit of the synthetic population to known distributions of population characteristics

i) Approaches implemented in PopGen to correct for zero-cell and zero-marginal value problems in the context of implementing iterative proportional fitting (IPF) and enhanced iterative proportional control of household and person-level control variables

The presentation also touched on software development considerations from the perspective of developing a stand-alone synthetic population generator that can be applied in the context of any travel demand model system. The presentation ended with a description of the many components that would comprise a comprehensive population evolution model system that may be employed for forecasting populations in a future year. Within the context of this project, the project team envisions developing an initial functioning prototype of a population evolution model consisting of a small subset of the model components presented in the kick-off meeting. The exact design of the initial prototype to be implemented in this project will be spelled out in detail in the Technical Memorandum submitted at the end of Task 4, following close consultations with BMC staff. The agreed-upon prototype will then be programmed and implemented as part of Task 5B.

Following the detailed presentation on PopGen, Maren Outwater of Resource Systems Group, Inc. (RSG) provided an overview of the options available for integrating the synthetic population generator within the current travel demand modeling environment. The presentation included a detailed explanation of how synthetic population generators can be integrated with both land use microsimulation model systems on the one hand and with travel demand model systems on the other. Land use microsimulation model systems are generally concerned with the determination of longer term location choices for households and individuals. These include such choices as residential and work place location choices, and school location choices for adults and children. A synthetic population generator can generate a population at the level of a traffic analysis zone or census tract or block group in such a way that known population characteristics are replicated as much as possible. A land use microsimulation model system can then take the synthetic population and allocate each household to an individual parcel (or small grid cell) or building unit in terms of its residence. In addition, a land use microsimulation model system will attach each worker with a specific job, thus establishing the regular work place location for each worker in the synthetic population. Depending on school enrollment and capacity constraints, each child or adult student is allocated a school location choice as well. Thus, a synthetic population generator provides the population, and associated socio-economic attributes, for which disaggregate location choices may be simulated. Once the synthetic population is complete, then aggregate tables providing joint distributions of the number of households (or persons) in multiway tables of household and person variables may be generated. These multiway tables, providing household and person frequencies in individual cells, can be used in conjunction with cross-classification trip generation models to estimate trip productions. While this is one potential way in which the synthetic population generator can be used in the short term in the context of the four-step travel demand model, the merits of synthetic population generator can be fully exploited when the generator is integrated with an activity-based travel microsimulation model system. In the activity-based modeling context, the synthetic population serves as the input to the activity-based travel model. The activity-based travel model works on each individual and household in the synthetic population and simulates their individual and collective daily activity-travel patterns.

The integration of the synthetic population generator within the BMC Cube/TP+ travel model environment was a major theme of the presentation offered by Maren Outwater of Resource Systems Group, Inc. (RSG). Additional discussion on this aspect of the project effort is offered in the next subsection of this technical memorandum. Basically, there are two paths that one could consider for this integration effort. One could code the entire PopGen software and algorithm within Cube/TP+ so that the entire program is native to Cube. The Cube Scripting Language may be used to accomplish this. While this offers some advantages in that the synthetic population generator becomes a native piece of the Cube/TP+ model system, it also suffers from some drawbacks. One drawback is that one would have to keep updating the script every time an improvement or enhancement (or bug fix) is made to PopGen, which is written in Python and uses Python libraries (such as Numpy). Another drawback is that the project resources would be spent to simply duplicate an entity that is already available as a robust software package that has been extensively tested and applied in a variety of urban modeling environments. For this reason, the project team believes that the objectives of the project and the modeling needs of BMC are best met by retaining PopGen as a software package, but enhancing the package, and programming appropriate Cube utilities or scripts, so that PopGen could be integrated with the BMC travel demand model using the Application Manager of Cube/TP+. The core PopGen code that implements the algorithm would be isolated from the PopGen graphical user interface (GUI) so that it can be effectively hooked to the BMC Cube/TP+ model through the Application Manager. Then, every time the PopGen code is modified, revised, or enhanced, the code in Cube/TP+ application manager can be replaced thus making updates and upgrades very easy and quick without the need for extensive programming or scripting within Cube.

The second day of the kick-off meeting started with a presentation by Fred Ducca of the University of Maryland National Center for Smart Growth (NCSG). This presentation offered an overview of the shortcomings of current four-step travel demand model practice in the context of emerging planning and policy needs, particularly as articulated and summarized in TRB Special Report 288 on the state of Metropolitan Travel Forecasting in the United States. The four-step travel demand model is not able to effectively respond to emerging policies of interest (such as dynamic pricing schemes, new transit technologies, flexible work hours and telecommuting arrangements, and managed lanes systems) because the models are not designed to capture the interactions and time-space constraints that govern activity-travel engagement patterns of individuals and households. The inability to analyze behavioral response to policy actions at the level of the disaggregate individual decision maker severely limits the ability of aggregate zone-based travel demand models to accurately forecast changes in system conditions under alternative policy scenarios. As a result, it is difficult to conduct social equity or environmental justice analysis, user-benefits estimation, and policy evaluation using current four-step travel demand modeling tools. In making the transition to activity-based models, Ducca noted that it is possible to consider phased implementation plans that help agencies make the move to activity-based models in a gradual, but effective, manner. The presentation included two possible paths. In the first path, one could replace the first three steps of the four-step travel demand model (namely, trip generation, trip distribution, and modal split) with an activity-based demand model. The resulting activity-travel patterns can be aggregated into origin-destination trip tables (by time of day period) and fed into existing static traffic assignment procedures. In a subsequent phase, one could replace the static traffic assignment procedures with dynamic traffic assignment procedures that fully exploit the availability of disaggregate trip and tour information emanating from an activity-based travel demand model system. In the second path, the transition process can be reversed. Similar to some of the ongoing TRANSIMS deployment and test application efforts, one could replace the current static assignment process with a more dynamic routing and microsimulation process. Origin-destination trip tables coming out of the first three steps of existing models could be disaggregated using time-of-day distributions to produce lists of trips along the time axis. These lists of trips could be provided to the dynamic traffic assignment and microsimulation model so that individual trips are routed through the network and system performance can be accurately measured. In a subsequent phase, one could replace the first three steps of the travel demand model with an activity-based model so that the lists of trips to be microsimulated on the network are generated through a disaggregate activity-based microsimulation model system. As it appears that BMC is interested in working through the development of an activity-based travel demand model system first, it is envisioned that the approach taken by BMC will follow the first path.

Following the presentation by Fred Ducca, Ram Pendyala of Arizona State University provided a very detailed presentation on activity-based travel model systems and the theory behind activity-based modeling. This presentation traced the early development of the theory of activity-based travel modeling to the work in the field of Regional Science in the 1970s. Activity-based paradigms found their way into transportation in earnest in the 1980s with much progress made in the research arena in the 1990s and into the 2000s. Research continues to take place at this time and there are a number of activity-based travel modeling frameworks that have been proposed by researchers around the world. The PowerPoint presentation includes a description and graphical depiction of the many activity-based travel model systems reviewed by Pendyala. However, it is important to note that the tour-based model systems that have been implemented in Metropolitan Planning Organizations (MPO) in the United States adopt a more practical tour-based paradigm that does not necessarily treat time as a continuous entity that is all-encompassing when one considers human activity-travel engagement. An exception to this is the model that is being implemented in the Southern California region by members of the project team for the Southern California Association of Governments (SCAG). Following the review of the research-oriented and SCAG activity-based travel model systems, Maren Outwater of Resource Systems Group, Inc. provided an overview of the model systems implemented at other MPOs around the country. These include the DaySim model family implemented in Sacramento, Puget Sound, and Jacksonville, Florida (as part of the ongoing SHRP2 C10 project) and the CT-RAMP family of models implemented by PB in places such as Atlanta, San Francisco, San Diego, Phoenix, New York, and Columbus. The Puget Sound region has taken a more multiphase approach to the development of activity-based model systems with the first phase involving *only* the replacement of the trip generation step with an activity-based travel model. In this application, a trip chaining model is deployed to estimate the number of trips, and then individual trips are passed through the traditional trip distribution, modal split, and traffic assignment steps.

The Baltimore Metropolitan Council (BMC) staff provided the project team an overview of the travel demand model and land use model systems in place at BMC. The project team noted that BMC has databases needed to support the development of an activity-based travel model, and a sophisticated land use modeling process that engages stakeholders and could work very effectively in conjunction with a synthetic population generator. The land use modeling process that is currently in place at BMC would provide the traffic analysis zone (TAZ) level control totals for a variety of socio-economic and demographic characteristics that would, in turn, drive the synthetic population generation process. Once the synthetic population is generated, one can aggregate back up to get joint distributions of population characteristics in each zone that could be applied in the context of an existing cross-classification model of trip generation. Once an activity-based travel model is put in place, then the synthetic population would directly feed into that model system. Through such a design, the land use modeling process currently in place at BMC does not have to be disturbed and the valuable participation of the local stakeholders can be preserved. Even after a population evolution model is implemented for BMC, the land use modeling process should continue to provide population control totals at the TAZ level. The population evolution model system will evolve the population over time, but with the constraint that marginal control totals should match those provided by the land use modeling process of BMC.

Due to time constraints, it was not possible to engage in an extended discussion on the specifics of the synthetic population generator development project. In general, it was agreed that deliverables would follow the schedule as noted in the executed contract, and that a refined work plan and project operational logistics would be documented in the first technical memorandum (Technical Memorandum 1.1). The project team committed to providing the necessary training and technical support (virtual and on-site) that is necessary to ensure that the BMC staff members are fully capable of engaging in the use, enhancement, and application of the synthetic population generator. It is interesting to note that the project manager at BMC has already developed a comprehensive spreadsheet-based implementation of the PopGen algorithms and procedures, thus demonstrating a thorough knowledge of the underlying algorithms incorporated in PopGen. In fact, this spreadsheet will be used as a teaching tool in future PopGen workshops.

The kick-off meeting concluded with a hands-on computer-based training workshop on the use of PopGen, the stand-alone software application developed at ASU for purposes of synthetic population generation. Several BMC staff members participate in the training workshop and learnt how PopGen reads and processes data, synthesizes population, and graphically displays output diagnostics. It was during the workshop that the project team was notified of the key difference between Baltimore County and Baltimore City, which are mutually exclusive of one another. The project team is currently in the process of making the necessary enhancements to clearly delineate the geographic entity Baltimore County from the geographic entity Baltimore City. Following the hands-on computer-based training session, the kick-off meeting concluded. All PowerPoint presentations from the kick-off meeting have been posted at the BMC project website at: http://www.baltometro.org/transportation-planning/activity-based-modeling. The project wiki site will also link to this BMC site so that visitors to the project wiki site can access the presentations as desired.

3. APPROACH TO SYNTHETIC POPULATION GENERATION

The advent of activity-based paradigms to travel demand forecasting has ushered in a new era of microsimulation-based approaches in transportation planning. The microsimulation of activity-travel patterns essentially involves simulating the daily activities and trips (tours) of each and every individual in the population, including infants and children. It is for this reason that one needs a complete population for the region of interest to implement an activity-based microsimulation model of travel demand.

A synthetic population generator was originally developed by researchers at the Los Alamos National Laboratory (LANL) as part of the TRANSIMS development effort. The synthetic population generator developed in that work has, in many ways, become the foundation for many population synthesizers that have been developed and implemented around the world. In a typical population synthesizer (such as the TRANSIMS population synthesizer), the following basic steps are executed:

a) Select a set of household control variables such as household size, number of workers, number of children, and household income, and set categories for each variable

b) Obtain marginal distributions for each variable of interest based on Census Summary Files or TAZ socio-economic files for the geography of interest (TAZ, census tract, census block group)

c) Create a seed joint distribution matrix of the variables of interest using the Public Use Microdata Sample (PUMS) data set available from the Census or using any other disaggregate survey data available for the region.

d) Use Iterative Proportional Fitting (IPF) algorithm to expand the seed matrix such that the population-level cell frequencies are computed, while ensuring that known population-level marginal control totals are matched. This step will inform the analyst the number of households (in the population) that fall into every cell of the multiway distribution table.

e) Use a probabilistic drawing process to sample households from the PUMS data set to create a synthetic population. The number of households of each type (defined as a cell in the multiway distribution table) drawn into the synthetic population should match the expanded cell frequency obtained after running the IPF algorithm.

There are several additional steps in the process that have not been explicitly identified here. For example, there is a step to correct or account for the zero-cell problem wherein one might encounter a zero-cell frequency in the seed matrix (even after running the IPF algorithm, one ends up with a population frequency for that cell as zero, thus resulting in no households of that type being drawn into the synthetic population). There is a step to round off frequencies in the expanded joint distribution matrix so that one is dealing with whole numbers prior to the household drawing process (after all, one cannot draw a fraction of a household). The probabilistic drawing process itself can take different forms, although most are basic variants of Monte Carlo simulation procedures. Finally, there is a step to assess the goodness-of-fit or performance of the synthetic population generation process in replicating known population characteristics and distributions, with possible iterative loops incorporated to obtain the best fit synthetic population.

In general, this process is quite robust and has therefore served as the foundation for population synthesizers that have followed that initial effort. One of the key drawbacks of many population synthesizers (including the original TRANSIMS population synthesizer) is that they control exclusively for household level variables, but do not control for person level variables. As a result, the synthetic population that is generated invariably matches household-level control variables with respect to distributions and frequencies without any problem, but performs poorly in matching the person-level variable distributions. When a synthetic population is generated, all households of a certain type (cell) receive the same weight or probability of being selected into the synthetic population regardless of the person composition of the household. In a simple case, consider two households, each of household size two. Both households could be a married couple, thus suggesting that the households are similar in composition. However, in reality, these households could be very different from one another. One household may be a couple of retired older individuals, whose children have moved out of the house. These individuals may both be over 65 years of age. The second household may be a young couple, individuals who are below 30 years of age and do not have any children. The activity-travel patterns of these two households are likely to be very different simply because the individuals in the households are in very different stages of their lifecycle. However, a synthetic population generation process that only controls for household attributes such as household size and marital status would not distinguish between these two households. Both of these households would qualify as the two-person married couple type and would receive an equal “weight” or probability of selection into the synthetic population. Such a process may not provide a synthetic population that accurately replicates known distributions of person-level characteristics in the population. In a zone that is in a retirement community, one would expect that the age distribution is skewed towards older individuals. In that case, the household with older individuals who are retired should receive a higher weight or probability of being selected than the household with the young couple. Similarly, if one were to consider a zone of young professionals and families, then the household with the young couple should receive a higher weight or probability of selection than the household with retired older individuals. If one is controlling only for household level control variables, then both households receive exactly the same weight, and the resulting synthetic population is not going to match the age distribution (or any person level control variable) as well as would have been possible had the person level variables been also controlled.

In order to overcome this drawback, recent population synthesizers, including PopGen, have developed mechanisms to also control for person-level variables. It is in this respect that recent population synthesizers differ from one another in the algorithm employed. PopGen offers a rigorous, but practical and intuitively appealing algorithm to simultaneously control for both household level and person level control variables. PopGen is the backbone of the synthetic population generator that will be implemented in BMC in the course of this project and hence the approach followed in PopGen forms the basis of discussion in this technical memorandum. The principle that is implemented in PopGen is founded on the notion that households of the same type, but of different composition, should receive different weights or have different probabilities of being selected or drawn into the synthetic population depending on the person characteristics of interest. The weights allocated to households of a certain type (cell) are re-distributed or re-allocated across the households in such a way that person characteristics of interest are matched as closely as possible. This re-allocation process is achieved through the implementation of an additional component in the population synthesis procedure; this step is known as the Iterative Proportational Updating (IPU) algorithm within the PopGen model.

The approach to synthetic population generation may be summarized as one that simultaneously controls for both household and person level variables of interest to the agency. Once an initial set of weights is obtained based on matching household level control distributions, adjustments or updates to the weights are obtained in such a way that person level control distributions are matched as closely as possible without compromising on the match to household-level controls. Consider once again the two households as described previously. After controlling for household level control variables, both households may have received exactly the same weight of say, 25. If one were to control for a host of person-level control variables such as employment status, age distribution, and gender, then the weights on these two households will be re-allocated in such a way that these person variables are matched as closely as possible. In a zone belonging to a retirement community, the retired household may see its weight greatly increased to about 40, while the weight on the younger household may reduce to just 10. In another zone that belongs to a young professional and family community, the younger couple household may see its weight enhanced to 40 while that of the older household may be adjusted down to 10. In all cases, the sum of the weights is 50, but the allocation between the two households is different based on the person characteristics of interest that need to be replicated.

The enhanced process of population synthesis incorporated in PopGen consists of the following steps:

a) Identify a set of household *and person* control variables whose distributions should be matched in the synthetic population generation process. For each variable, identify an appropriate set of categories – that presents a compromise between preserving detail and avoiding the creation of large multiway distribution tables that have many sparse (or zero) cells in the seed matrix generated from the PUMS data set.

b) Obtain marginal distributions for each variable of interest based on Census Summary Files or TAZ socio-economic files for the geography of interest (TAZ, census tract, census block group). These marginal distributions should be obtained for both household and person variables.

c) Create a seed joint distribution matrix of the variables of interest using the Public Use Microdata Sample (PUMS) data set available from the Census or using any other disaggregate survey data available for the region. Generate separate matrices for the household variables and the person variables.

d) Use Iterative Proportional Fitting (IPF) algorithm to expand the seed matrix such that the population-level cell frequencies are computed, while ensuring that known population-level marginal control totals are matched. This step will inform the analyst the number of households or persons (in the population) that fall into every cell of the multiway distribution table. In other words, the IPF procedure is applied separately to household and person multiway tables to obtain the population frequency in each corresponding cell. These household and person level population frequencies will serve as the inputs to the new step introduced in PopGen.

e) Apply the Iterative Proportional Updating (IPU) algorithm to re-computed or re-distribute or re-allocate weights among households of a certain type such that person level control distributions are matched as closely as possible. This algorithm will be described in detail in the technical memoranda submitted as part of Tasks 3 and 4 of the project. The algorithm essentially re-allocates weights based on the composition of the household in the PUMS data set and the marginal control totals for person level variables of interest in the geography under consideration.

f) Use a probabilistic drawing process to sample households from the PUMS data set to create a synthetic population. Households are drawn probabilistically into the synthetic population based on the re-allocated weights assigned to them. Thus, households with larger weights in the same household matrix cell will have a greater probability of being selected into the synthetic population, thus helping better match the person level variables of interest.

Finally, the project team is committed to customizing PopGen to the Baltimore Metropolitan Council (BMC) implementation as per the specifications and needs of the agency. At this time, PopGen is a generic software package that can be applied to any metropolitan region in the country. It is envisioned that PopGen-BMC will be customized so that it is applicable to the Baltimore region only and includes a suite of features and specially prepared data sets that apply specifically to the region. In addition to customizing the software application to work within the Cube/TP+ modeling environment, it will also be fine-tuned to work specifically with data sets for the region and incorporate features desired by the agency. The project team will work closely with the BMC staff over the next few weeks to identify the specific features desired by the agency and then design a customized version of PopGen for BMC and document the design in a Task 3 technical memorandum.

4. DESIGN CONSIDERATIONS IN SYNTHETIC POPULATION GENERATION

The previous section provided a general overview of the methodological approach that will be adopted in synthetic population generation for BMC. There are several additional design considerations that will be accommodated in PopGen-BMC so that the synthetic population generation process is robust and the performance of the process can be adequately evaluated. This section provides an overview of some additional design considerations that will be incorporated in PopGen-BMC.

The evaluation of the performance of a synthetic population generation process can be done by comparing the distributions of a host of population (household and person level) characteristics against those in the synthetic population. Within PopGen, goodness-of-fit is measured and the performance of the system is evaluated in at least the following three ways:

- *Convergence of the IPF Procedure*: In the PopGen methodology, the IPF procedure is executed first to determine the expanded number of households and persons that fall into each of the cells of the multiway tables defined by a host of household and person level control variables. Note that there is one multiway control table for household characteristics and another such table for person level characteristics. The IPF procedure is an algorithm that systematically adjusts cell frequencies to match row and column totals in turn until both row and column totals are matched as closely as possible. PopGen incorporates a default (very small) tolerance level at which the IPF procedure is deemed to have converged and the process can stop. The tolerance level that is achieved in the IPF procedure is a measure of the performance of this step of the process. In general, it is found that the IPF procedure is quite robust with the ability to converge rapidly and meet low tolerance levels. However, there may be situations where the IPF procedure fails to converge rapidly or meet the desired tolerance levels; in such situations, the data need to be examined carefully and the nature of the specific geographical unit where the problem is being encountered should be explored further to identify potential remedial action. For example, one may have TAZs in which there is an unusually small number of households (just one or two); in such cases, the IPF algorithm may struggle to reach convergence as the multiway table for this geography is inevitably going to be filled with zeros.

- *Convergence of the IPU Procedure*: Once the IPF procedure is complete and the number of households or persons of each type that should be simulated is determined, the IPU procedure actually allocates weights to each household in the sample survey data set (such as PUMS data set) such that the household and person type constraints (cell frequencies) are met. This process is an iterative process wherein the weights are adjusted for each household based on its household characteristics and person type composition. The end goal is to have a set of weights such that the weighted frequencies match the household and person control totals as closely as possible. In general, the procedure adopted in PopGen is such that household level constraints are matched perfectly and no compromise is made in the matching of household control totals. This is done in recognition of the notion that activity-travel demand is a manifestation of household level decision processes and that the household tends to be the appropriate behavioral unit of importance in simulating individual activity-travel patterns. With respect to person-level constraints, PopGen does the best possible job in matching the totals. The process is iterative and PopGen computes the difference between the actual IPF-generated cell frequencies and the IPU-generated cell totals. The average percent difference between these figures across all household and person types (cells) is a good measure of fit indicative of the extent to which the IPU procedure is producing weights such that household and person totals match expanded frequencies estimated through the IPF procedure. As the IPU procedure progresses, this measure of fit is computed at the end of each iteration. Convergence is considered to have been achieved when the *improvement* in the goodness of fit (not the actual goodness of fit measure itself) does not exceed a small tolerance level. If the improvement is greater than the tolerance criterion, then the process continues. Otherwise, the process stops. In addition, for those geographies where the IPU procedure continues indefinitely, a careful examination of the data is warranted to check and see why convergence is not being achieved. Even after such a careful data examination, if convergence is not being achieved, then the number of iterations is constrained to not exceed a certain number.

- *Level of Match on Distributions of Person-level Control Variables*: After the IPU procedure is completed, then the households are drawn probabilistically from the sample survey data set (such as the PUMS data set) in accordance with the weights estimated by the IPU procedure. As the drawing process is probabilistic, one would naturally obtain a slightly different synthetic population every time the population drawing process is executed. In recognition of this stochastic random nature of the synthetic population drawing process, PopGen incorporates a feature wherein multiple populations are drawn and the one with the best fit is taken to constitute the synthetic population for that geography. At the end of each draw, the person-level multiway frequency table in the synthetic population is compared against that obtained at the end of the IPF step. The difference between two frequency tables can be checked for statistical significance through the chi-square (χ^2) statistic and the null hypothesis that the tables are not significantly different from one another can be tested. If the chi-square test statistic is low, then it implies that the null hypothesis cannot be rejected with a relatively high level of confidence (as denoted by a p-value). Thus, if the p-value is high, then it means that the synthetic population closely matches the actual population distributions quite closely. A desired p-value is set as a stopping criterion; if a certain draw of a synthetic population yields a p-value greater than the desired threshold, then the drawn population may be considered satisfactory and included in the synthetic population. Otherwise, the synthetic population draw is rejected and the process is repeated until the desired p-value is achieved. Once again, to accommodate situations where the desired p-value may not be achieved even after many trials, the number of possible trials is subject to a maximum limit. Once the maximum limit is reached, the synthetic population that provided the highest p-value is adopted.

- *Compare Distributions of Non-Controlled Variables*: The synthetic population generation process is accomplished by controlling for household and person-level variables that are deemed important for the transportation modeling process. In general, the synthetic population that is generated by PopGen should have distributions of these controlled variables that closely match and resemble those in the actual population. When a synthetic population is generated, all of the other attributes of the drawn households are automatically appended to each household and person record. Some, and possibly many, of these attributes would not have been controlled in the synthetic population generation process. Then, it is of interest to see how the synthetic population compares with the actual population in the distribution of these non-controlled variables. PopGen incorporates a feature wherein the distribution of a non-controlled variable in the synthetic population can be compared against the distribution of the same variable in the actual population. In general, one would expect the fit to be poorer for these variables simply because they were not controlled in the synthetic population generation process. However, if there are variables where the distributions are completely different or mismatched, then it may be prudent to consider whether the variable in question should also be controlled in the synthetic population generation process. Although it is impossible to match distributions across all household and person variables of potential interest, one would like to see distributions match well on at least some variables considered key determinants of travel behavior.

The last check identified above points to the need to carefully select household and person-level control variables for the synthetic population generation process. A basic criterion for household and person-level control variable selection is that those variables included in travel model specifications and considered key determinants of travel behavior should be included as control variables. However, in some instances, the number of variables of interest based on this criterion may be prohibitively large leading to huge multiway tables with many sparse (zero) cells (in the seed matrix). A careful compromise must be made between the variables included as control variables and the computational aspects of the population synthesis procedure. This process may be iterative in nature and will inevitably require some trial-and-error prior to determining the optimal combination of person and household control variables and their categories. When it is impossible to include all variables of interest as control variables, then the combination of variables that is included as control should capture the effects of uncontrolled variables (of importance) to the extent possible. In other words, the variables that are controlled should be highly correlated with variables of interest (or importance) that are not controlled. For example, number of workers may be correlated with income, the presence of children may be correlated with marital status, and housing unit type may be correlated with home ownership (or tenure). In each of these situations, it may be sufficient to include one of the two variables in an effort to keep the specification of the multiway table parsimonious. At the end of the synthetic population generation process, the frequency distributions on non-controlled variables need to be checked and compared to see if indeed the variables included as control were able to adequately account for the effects of key and important variables not included as control.

Additional design considerations are related to the quality of data available for performing population synthesis. The initial joint seed matrix used for the IPF procedure is generally formulated based on PUMS (or sample survey) data for the PUMA region to which the TAZ or geography of interest belongs. When the multiway table of interest is rather extensive, there may be cells for which zero observations are available in the PUMS or sample survey data set. This zero cell issue is a problem because even the expanded matrix (after running the IPF procedure) will have zero values when, in fact, the cells ought to have some positive numbers to reflect the fact that such households or persons do actually exist in the small geography in the actual population. To overcome this problem, the classic remedy has been to add a small value (e.g., 0.5) to the zero cells so that the IPF procedure will result in a positive number of households (albeit small) in these sparse cells. While this remedy is generally robust and works well, caution needs to be exercised in cases where the number of zero cells is large. If 0.5 is added to many cells where zeros occur, then one may distort the underlying distribution of characteristics in the population, and this may lead to erroneous numbers of households and persons being generated in the synthetic population. The project team is has developed more intelligent ways of dealing with the zero cell problem and will implement a few different options in PopGen-BMC to ensure that all types of zero-cell situations can be adequately addressed. One option currently implemented in PopGen is to borrow a prior from the seed matrix constructed using the entire PUMS data set for the whole region. This prior is borrowed (subject to a maximum limit), and the seed matrix is adjusted to minimize any distortion of underlying distributions of population characteristics. Similarly, in the context of the IPU algorithm, the nature of the algorithm is such that it cannot proceed if there is a zero marginal control total. Once again, PopGen adds a small numeric value (e.g., 0.1 divided by the number of zero marginals for the dimension in question) to the marginal control total to allow the IPU procedure to move forward. Extensive testing has shown that this correction works well and does not compromise the synthetic population generation process in any appreciable way. The project team will undertake more specific testing of this correction process in the context of this project to ensure that the best possible correction mechanism is implemented in PopGen-BMC.

An important data quality related issue that the project team has encountered in past work is the mismatch between person totals implied by household level distributions and person level distributions. For example consider a household size distribution where one has the number of households of each size in the marginal control totals. This household size distribution implies a certain population total (or a range within which the population total must fall). If one compares this population total (or range) against the population total implied by the age or gender distribution, then the numbers should match either perfectly or at least very closely. In some situations, the project team has found that the population totals implied by the household size distribution and the person variable distribution differ substantially, calling into question the quality and consistency of the zonal data. In implementing PopGen for BMC, the project team will undertake these checks and bring them to the attention of BMC so that the data can be corrected and consistency can be enhanced. Where such corrections cannot be done or would not be possible in a timely manner within the schedule of this project, the project team will implement an automated correction procedure that adjusts the household size distribution such that population totals are matched nearly perfectly. This adjustment procedure is rather straightforward and will be described in detail in a Task 3 technical memorandum. Difficulties with data also arise in situations where one may have only one or two households in an entire TAZ. The project team is currently exploring ways to deal with such outlier TAZs, perhaps using heuristics given that the actual population total for such TAZs is extremely low. The initial seed matrix and the marginal control totals are going to be replete with zero values, and the application of any correction mechanism may distort sample distributions from the reality of the population distributions. As such, a more heuristic procedure may be best implemented to deal with these TAZs. The project team is currently developing this procedure and will document the mechanism in a Task 3 technical memorandum.

5. INTEGRATION OF SYNTHETIC POPULATION GENERATOR WITH CUBE/TP+ MODEL SYSTEM

The overall goal of this project is to provide a synthetic population generator that is seamlessly integrated with the Cube/TP+ modeling environment currently in place at BMC. The project team is currently working through the design and details of the model integration framework and the complete design will be documented in a Task 3 technical memorandum. The approach being taken by the project team is one in which the core PopGen algorithm and code (written in Python) is seamlessly connected to the Cube/TP+ model system through the application manager of Cube. The application manager of Cube allows one to access and run third party applications from within the Cube modeling environment. PopGen will be treated as a third party application that can be accessed and run using the application manager. Figure 1 shows a broad schematic of how the process of integration would be designed.

PopGen has a set of control files and parameters that need to be defined for any scenario or run. These control files and parameters are currently established in the stand-alone PopGen software package through a graphical user interface and setup wizard. Once these items are stripped away from the

PopGen core code, the setup must be handled internally within the Cube modeling environment. The project team plans to code and script within Cube, a PopGen set up process so that the control files, variables, and parameters can be specified in a user-friendly and intuitive manner. The specification and set up of the input control files and parameters will therefore be a native Cube process and the analyst will be able to access the set up screens similar to any other travel model component within Cube. Once the set up process is complete, the application manager will run PopGen in a seamless manner without the need for the user to specially open or close any programs. Upon the completion of the PopGen run, the user will be able to view the results, synthetic population output, and other diagnostics within the Cube modeling environment. All outputs will be automatically converted to Cube-readable formats and structures, and the output can be visually displayed or saved to files in a variety of formats. All GIS-based visual displays will be available within the Cube-ArcGIS environment, thus providing a convenient mechanism to generate production quality maps of the synthetic population generation process. In essence, the input and output specification and generation steps will be subsumed into Cube native processes, while the PopGen core code remains a third-party program accessed and run through the application manager. The project team is currently working to separate the PopGen code from the graphical user interfaces coded within the stand-alone PopGen software package. In the process, the program is being modified to generate a XML control file when the user specifies the input parameters and preferences. PopGen will run in accordance with the specifications and preferences written to the XML control file. This process is going to be replicated within the Cube modeling environment when PopGen-BMC is implemented for BMC.

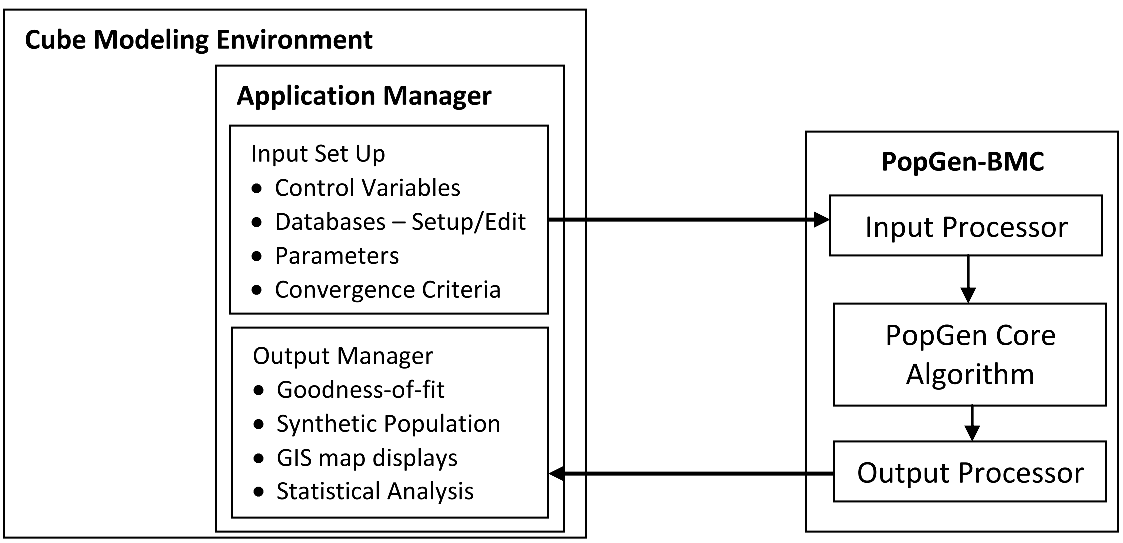

Figure 1. Schematic of PopGen-BMC in Cube Modeling Environment

SECTION 2. ACTIVITY-BASED MODELS: CONCEPTS, STATE-OF-THE-ART, AND DATA NEEDS

1. INTRODUCTION

Over the past few decades, because of escalating capital costs of new infrastructure, and increasing concerns regarding traffic congestion, energy dependence, greenhouse gas emissions, and air-quality, the originally supply-oriented focus of transportation planning has expanded to include the objective of addressing accessibility needs and problems by managing travel demand within the available transportation supply. Consequently, there has been an increasing interest in travel demand management strategies, such as mixed land-use development, parking pricing, and congestion pricing, all of which attempt to change land-use and transport service characteristics to influence individual travel behavior and control aggregate travel demand. The evaluation of such demand management strategies using travel demand models places more emphasis on the realistic representation of behavior to accurately reflect traveler responses to management policies. Enhanced behavioral realism brought to the forefront the need to assess these policies at the level of individual decision makers (individuals and households) and groups of particular interest. This realization has led to the consideration of an activity-based travel demand modeling system that operates at the level of decision-makers (individuals and households) within the context of land-use, urban form, and built environment characteristics.

2. THE ACTIVITY-BASED APPROACH

The fundamental difference between the trip-based and activity-based approaches is that the former approach directly focuses on “travel participation behavior” as the decision entity of interest, while the activity-based approach views travel as a demand derived from the need to pursue activities and focuses on “activity participation behavior”. The underlying philosophy of the activity-based approach is to better understand the behavioral basis for individual decisions regarding participation in activities in certain places at given times, and hence the resulting travel needs. This behavioral basis includes all the factors that influence the why, how, when, with whom, and where of performed activities and resulting travel.

At a fundamental level, therefore, the activity-based approach emphasizes the point that the needs of the households are likely to be translated into a certain number of total activity stops by purpose followed by (or jointly with) decisions regarding how the stops are organized. For example, consider a congestion pricing policy during the evening commute period along a corridor. Also, consider an individual who currently makes a shopping stop during the evening commute at a location that entails travel along the “to-be-priced” corridor (but assume that the person would not be traveling the ‘to-be-priced” corridor if she went directly home from work). In response to the pricing policy, the individual may now stop making the shopping stop during the evening commute, but may generate another stop in the evening after returning home from work. If some of these post-home arrival stops are undertaken in the peak period, congestion may be simply transferred to other locations in the network. The activity-based approach explicitly acknowledges the possibility of such temporal redistributions in activity participation (and hence travel) by focusing on sequences or patterns of activity participation (using the whole day or longer periods of time as the unit of analysis), and thus is able to provide a holistic picture of policy effects.

A second defining aspect of the activity-based approach is its use of "tours" as the basic element to represent and model travel patterns. Tours are chains of trips beginning and ending at a same location, say, home or work. The tour-based representation helps maintain consistency across, and capture the interdependency (and consistency) of the modeled choice attributes among, the activity episodes (and related travel characteristics) undertaken in the same tour. This is in contrast to the trip-based approach that considers travel as a collection of "trips", each trip treated like if it was independent of other trips. The explicit consideration in the activity-based approach of the inter-relationship in the choice attributes (such as time of participation, location of participation, and mode of travel) of different activity episodes within a tour, and therefore the recognition of the temporal, spatial and modal linkages among activity episodes within a tour, can lead to improved evaluations of the impact of policy actions. Take, for example, an individual who drives alone to work and makes a shopping stop on the way back home from work. The home-work and work-home trips in this scenario are not independent. Now consider an improvement in transit between the home and the work place. The activity-based approach would recognize that the individual needs to make a stop on the return home from work, and so may not predict a shift to transit for the work tour (including the home-work, work-to-shop, and shop-work trips), while a trip-based model would break the tour into three separate and independent trips – a home-based work trip, a non-home based non-work trip, and a home based non-work trip, and would be more likely (and inaccurately so) to shift the morning home-based work trip contribution of the individual to transit.

A third defining feature of the activity-based approach relates to the way the time dimension of activities and travel is considered. In the trip-based approach, time is included as a "cost" of making a trip and a day is viewed as a combination of broadly defined peak and off-peak time periods. On the other hand, the activity-based approach views individuals' activity-travel patterns as a result of their time-use decisions within a continuous time domain. Individuals have 24 hours in a day (or multiples of 24 hours for longer periods of time) and decide how to use that time among (or allocate that time to) activities and travel (and with whom). These decisions determine the generation and scheduling of trips. Hence, determining the impact of travel demand management policies on time-use behavior is an important precursor step to assessing the impact of such polices on individual travel behavior. Take the example of a worker who typically leaves work at 5:00 pm (say, the start of the evening peak period), drives to a grocery store 15 minutes away, spends about 25 minutes shopping, and then gets back home by 6:00 pm. In response to an early release from work policy designed by the employer that lets the employee off from work at 4:00 pm instead of 5:00 pm, a naïve model system may predict that the person would be off the road and back home by 5:00 pm (*i.e.*, before the peak period begins). But the individual, now released from work earlier and having more time on his hands after work, may decide to drive a longer distance to a preferred grocery store where he spends more time shopping (say 70 minutes as against 25 minutes) and may eventually return home only at 6:00 pm. So, in the case of this individual, the policy would not only be ineffective in keeping the person off the road during the peak period, but also the longer time spent at the grocery (soak duration between successive trips) has adverse air quality implications. The activity-based model is able to consider such interactions in space and time due its emphasis on time-use, and thus can produce more informed evaluations of policy actions.

Another feature of the activity-based approach is the recognition of interactions among household members, which leads to the accommodation of linkages among trips of household members. As a result, policy actions could have complex responses. Consider that Person 1 (the worker in a household with children) was originally dropping off a child at school in the mornings and picking up the child from school in the evenings, as part of her commute. Assume a pricing strategy on a corridor that connects the school location and the worker's work location. Because of this pricing policy, the worker may not pursue the drop-off/pick-up tasks himself and has a simple home-work-home pattern. But now person 2 (the non-worker) generates drop-off and pick-up trips, perhaps supplemented with shopping stops during his drop-off/pick-up trips. Such an explicit modeling of inter-individual interactions and the resulting joint travel is particularly important to examine the effects of occupancy-specific tolling strategies such as high occupancy vehicle (HOV) lanes and high occupancy toll (HOT) lanes. Another way that household linkages in activities can have an effect on responses to policies is through a reluctance to change the spatial and temporal attributes of joint activity episode participations. For instance, serve passenger trips (such as dropping/picking children from daycare/school or other extracurricular activities) and joint social/recreational out-of-home activities of household members may not be moved around much because of schedule constraints. Acknowledging such joint interactions can therefore potentially lead to a more accurate evaluation of policy actions.

A final important feature of activity-based approaches relates to the level of aggregation of decision-makers used in the estimation and application of the models. In the trip-based approach, several aspects of travel (number of trips produced and attracted from each zone, trip interchanges, and mode split) are usually (though not always) estimated and/or applied at a relatively aggregate level of decision-makers (such as at the spatial level of travel analysis zones). The activity-based models, on the other hand, have the ability to relatively easily accommodate virtually any number of decision factors related to the socio-demographic characteristics of the individuals who actually make the activity-travel choices. Using microsimulation techniques, activity-based models predict the entire activity-travel patterns at the level of individuals (while recognizing temporal/spatial constraints across individuals of a household due to joint activity participations and serve passenger activities). Such a methodology ensures a realistic, consistent, and integral prediction of activity-travel patterns, which should lead to the better aggregate prediction of travel flows on the network in response to demographic changes or policy scenarios. Thus the activity-based models are well equipped to forecast the longer-term changes in travel demand in response to the changes in the socio-demographic composition and the activity-travel environment of urban areas, as well as in response to land-use and transportation policies.

In addition to the fundamental elements of an activity-based system as discussed above, other desirable attributes of a state-of-the-art modeling system include the following:

(a) Models the activity-travel patterns of all individuals in a household, including children and adults.

(b) Accommodates intra-household decisions of activity-travel choices among all individuals (children and adults) in a household in a compact and computationally efficient manner,

(c) Incorporates spatial-temporal dependencies and constraints in activity-travel patterns between and within individuals of a household by using a temporal resolution that is in the order of minutes as opposed to hours. Such a fine resolution of time is also important for time-varying policy analysis, such as time-varying pricing strategies.

(d) Allows enhanced sensitivity to land-use, built environment and development patterns, and multi-modal (and inter-modal) transportation policies and demographic changes in the population, by using an agent-based micro-simulation platform that is designed conceptually to accommodate any level of spatial resolution and incorporate time-varying levels of accessibility measures within a day (the spatial resolution scale can be anywhere from the disaggregate scale of parcels to the traditional use of TAZs, including combinations of resolution scales across different modeling modules.

(e) Enables a holistic assessment of the effects of land-use, built environment, and transportation policies on entire activity-travel patterns rather than assuming, for example, that congestion pricing or land-use changes will only impact mode choice or route choice

(f) Facilitates environmental justice (EJ) analyses by having the ability to examine the effects of policies on any defined segment of the population by type of activity, by spatial unit of interest, by travel mode, and for any time-of-day.

(g) Allows seamless interfacing and integration with land-use and demographic model outputs, GIS/Geo-data base input layers, GIS output visualization abilities/needs, querying and reporting capabilities, population synthesizer outputs, and freight forecasting and external-external trip model outputs. The output should be easy to interface with existing multi-period static and dynamic traffic assignment as well as the newly designed EPA software MOVES.

(h) Enables multimodal planning and allow the accurate analysis of the introduction of new transit modes, including for FTA New Starts modeling applications that require the application of robust travel demand model systems founded on sound data and that respond appropriately to changes in system conditions and the introduction of new modes.

(i) Involves a portable and flexible software architecture design that is based on ensuring data and data processing integrity, parallel processing and multi-threading capability (so that computational speed can be almost linearly increased by dedicating more processors), extensibility and modifiability in structure, and usability.

(j) Allows tracking intended and unintended consequences of a single policy action or combinations of policy actions at different levels of aggregation including individuals, households, segments of the population, neighborhoods, and entire regions or parts thereof.

3. DATA NEEDS

3.1 Data Needs for Estimation of Activity-Based Systems

The primary sources of data for the estimation of tour-based and activity-based models are household activity and/or travel surveys. As the term “household activity and/or travel surveys” suggests, the surveys can be either travel surveys (that collect information on out-of-home travel undertaken by the household members) or activity-travel surveys (that collect information on out-of-home activities and associated travel). Both the surveys implicitly or explicitly collect information on: (1) household-level characteristics (2) individual-level characteristics, and (3) information on the activity/travel episodes undertaken by the individuals. Activity surveys, however, may also collect additional information on the activities participated by individuals, specifically the participation, timing, and duration of in-home and joint activities.

It should be noted here that the development of several activity-based models to date has involved the use of household travel survey data that is not any different from those collected and used by regional

MPOs for their trip-based model development and calibration. Thus, the notion that activity-based models are data hungry is not necessarily accurate, at least at the estimation stage (though it should also be mentioned here that activity-based models would perhaps benefit more from larger sample sizes than would trip-based models, especially from the standpoint of estimating models of joint activity participation).

Data on regional land-use and transportation system networks are also typically used in model estimation. Land-use data includes information on the spatial residential characteristics of households, employment locations, and school and other locations at the level of spatial resolution (for example, zones or parcels) used in the models. The typical land-use information includes size and density measures, such as number of households, population, area (or size), employment by each category of employment, household density, population density, and employment density for each category of employment. In addition, one or more of the following land-use data are also used by some activity modeling systems: (1) land-use structure information, such as the percentage of commercial, residential, other, and open areas, percentage of water coverage, and the land-use mix, (2) socio-demographic characteristics, such as average household size, median household income, ethnic composition, housing characteristics such as median housing value, and housing type measures (single family and multiple family dwelling units), and (3) activity opportunity measures such as activity center intensity (*i.e.,* the number of business establishments per square mile) and density (*i.e.,* the number of business establishments per square mile) for each of several activity purposes.

Transportation network data needed in activity models are similar to data used in trip-based models and typically include highway network data, transit network data, and non-motorized mode data. The transportation system performance data should be of high quality, with time-varying level-of-service (LOS) characteristics (in-vehicle travel time, out-of-vehicle travel time, access, egress, and waiting times) across different time periods, as well as across different location pairs. However, as one moves to finer resolution of time and space in behavioral models, one also needs to move to finer levels of time and space to represent the environment in which behavior happens. To provide this fine detail of resolution, agencies are moving towards smaller traffic analysis zones, and assignment techniques with increased detail of the multimodal travel environment.

3.2 Data Needs for Application of Activity-Based Systems

Once the activity-based modeling system has been estimated using the data sources discussed above, the application of these advanced models for a study area for a base year requires as inputs the information on all individuals and households of the study area for the base year. Synthetic population generation techniques are used for this purpose, sometimes supplemented with a series of other demographic models (see next section). For a future year forecasting exercise, the inputs should consist of the future year synthetic population and land-use and level-of-service data. Thus, advanced travel demand model development should be supported with the development of detailed input data (*i.e.,* the synthetic population, level-of-service and land-use data) for future years. This can be done by either using aggregate demographic and land-use projections for future years and applying a synthetic population generator (just as in the base year), or by “evolving” the base synthetic population.

3.3 Data Needs for Calibration and Validation of Activity-Based Systems

In this section, the data sources that can be used to calibrate and validate activity-based model systems are discussed.

Validation of input data

- The base year synthetic population inputs can be validated against the census data and data available from state and regional agencies.
- To validate the input work locations, the home-work trip lengths and patterns can be matched against that in the census data.
- To validate the vehicle ownership inputs, census data and perhaps other sources such as motor vehicle department estimates of auto registrations can be used.

Calibration and validation of activity-travel outputs

- Each component of the activity-travel model system can be validated by comparing its predictions to the observed activity-travel patterns in the household activity-travel survey.
- The commute mode choice model can be validated using the journey to work data.
- The entire model system can be validated by comparing the traffic assignment outputs with the observed traffic volumes in the study area or even turning movements at major intersections.
- Highway traffic assignment validation can be undertaken by using observed traffic volumes by time of day, while transit traffic assignment validation can be pursued by using transit boarding/alighting data by route and stop by time of day from an on-board transit survey/count.

Along with the above identified base year calibrations and validations, it is essential to understand the forecasting ability and the policy sensitivity of advanced models for non-base year conditions.

To test the forecasting ability, the model performance for past years (for example, year 1990) and for existing “future” years (relative to the base year for the travel modeling effort; for example, year 2010) can be compared with the observed patterns in those years. For this purpose, complete input data (including the aggregate socio-demographic variable distributions for synthetic population generation, and the land-use and level-of-service data), observed traffic volumes, household activity and/travel survey data, and the Census data (if available) are required for past years and existing “future” years. In this regard, it is important that the regional planning agencies store and document the land-use data and transportation network data of past and existing “future” years.

An examination of the policy sensitivity of advanced models for non-base year conditions can be undertaken by assessing the impact on activity-travel patterns of changes in transportation system and land-use patterns. Such an examination of the response to several policy scenarios can be a useful assessment of the abilities of the activity-based model system (especially when compared with the outputs from a trip-based model system). The scenario approach discussed above to assessing the policy-sensitivity of activity-based models, however, may not completely represent the complexity of real life projects and policies. Further, sensitivity testing using test scenarios serves only as a broad qualitative reasonableness assessment of performance, rather than a quantitative performance measurement against observed data. A more robust way to quantify and assess the predicted policy sensitivity from activity-based models is to compare the model predictions with real-world data before- and after- real-life transportation infrastructure investments or policy actions. Hence it is important to collect traffic counts and other travel pattern data before- and after- any major transportation infrastructure investments or policy actions.

4. INTEGRATION WITH OTHER MODEL SYSTEMS

The recognition of the linkages among socio-demographics, land use, and transportation is important for realistic forecasts of travel demand, which has led practitioners and researchers to develop approaches that capture socio-demographic, land-use, and travel behavior processes in an integrated manner. Such behavioral approaches emphasize the interactions among population socioeconomic processes, the households' long-term choice behaviors, and the employment, housing, and transportation markets within which individuals and households act. The resulting integrated model system should be able to capture the supply-demand interactions in the housing, employment, and transportation markets. A conceptual simplified framework of such a system is provided in the diagram below.

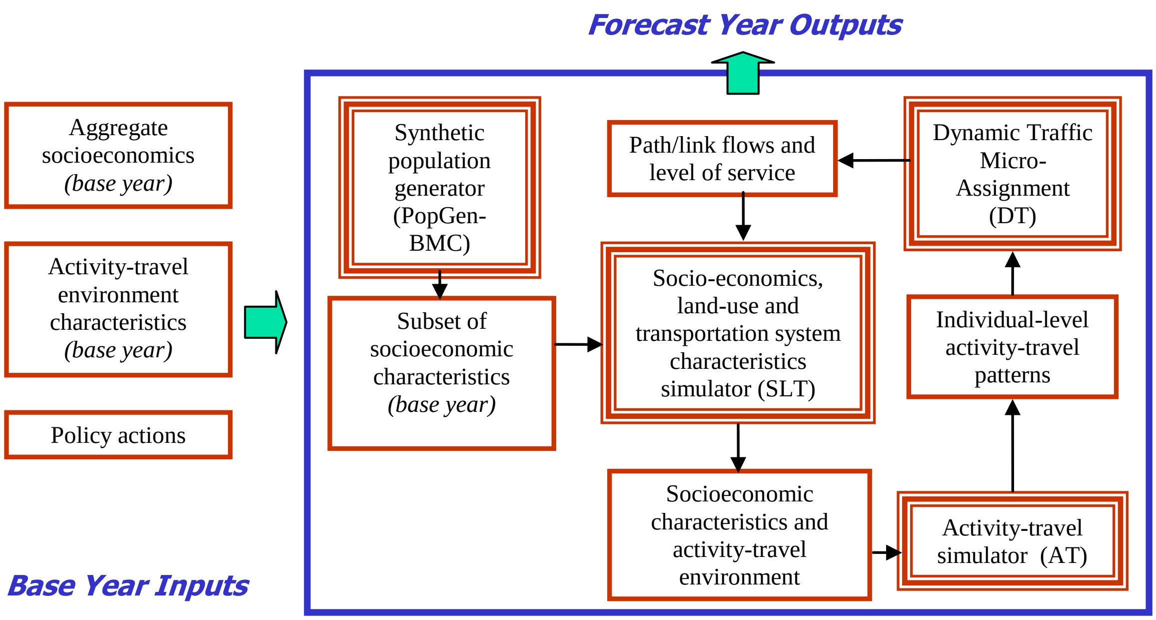

Figure 2. A Simplified Integrated Microsimulation Model System

The integrated system places the focus on households and individuals, and businesses and developers that are the primary decision makers in an urban system. The system takes as inputs the aggregate socio-economics and the land-use and transportation system characteristics for the base year, as well as policy actions being considered for future years. The aggregate-level base year socioeconomic data are first fed into a synthetic population generator (PopGen-BMC) module to produce a disaggregate-level synthetic dataset describing a subset of the socioeconomic characteristics of all the households and individuals residing in the study area. Additional base-year socioeconomic attributes related to mobility, schooling, and employment at the individual level, and residential/vehicle ownership choices at the household level, that are difficult to synthesize (or cannot be synthesized) directly from the aggregate socioeconomic data for the base year are simulated by the socio-economics, land-use, and transportation (SLT) system simulator. The base year socioeconomic data, along with the land-use and transportation system attributes, are then run through the daily activity-travel pattern (AT) simulator to obtain individual-level activity-travel patterns. The activity-travel patterns are subsequently passed through a traffic micro-assignment (DT) scheme to determine path flows, link flows, and transportation system level-of-service by time of day. The resulting transportation system level-of-service characteristics are fed back to the SLT simulator to generate a revised set of activity-travel environment attributes, which is passed through the AT simulator along with the socioeconomic data to generate revised individual activity-travel patterns. This "within-year" iteration is continued until base-year consistency in activity and travel patterns is achieved. This completes the simulation for the base year.

Environmental impacts in terms of pollutants produced can be computed in a system like this at different levels depending on available software. The path link flows and level of service (output of traffic assignment) can be used to estimate GHG emissions, CO, NOx, PM, and HC in a similar way as for trip based models. The DT micro-assignment offers the opportunity to also use the more recent advances in the MOVES software and possibly the assessment of ecodriving-based policies.

The next phase, which takes the population one step forward in time (*i.e*. one year), starts with the SLT simulator updating the population, urban-form, and the land-use markets (note that PopGen-BMC is used only to generate the disaggregate-level synthetic population for the base-year and is not used beyond the base year). An initial set of transportation system attributes is generated by SLT for this next time step based on (a) the population, urban form, and land-use markets for the next time step, (b) the transportation system attributes from the previous year in the simulation, and (c) the future year policy scenarios provided as input to the integrated system. The SLT outputs are then input into the AT system, which interfaces with the DT scheme in a series of equilibrium iterations for the next time step (just as for the base year) to obtain the "one time step" outputs. The loop continues for several time steps forward until the socioeconomics, land-use, and transportation system path/link flows and transportation system level of service are obtained for the forecast year specified by the analyst. During this iterative process, the effects of the prescribed policy actions can be evaluated based on the simulated network flows and speeds for any intermediate year between the base year and the forecast year.

APPENDIX A

BMC Regional Travel Demand Model Update
Development of Synthetic Population Generator

Project Kickoff Meeting and Workshop
Baltimore Metropolitan Council
Offices @ McHenry Row
1500 Whetstone Way, Suite 300
Baltimore, MD 21230
http://www.baltometro.org/contact-us/directions-to-bmc

Agenda

Tuesday, November 9, 2010

10:30 AM	Welcome and Introductions	BMC
10:45 AM	Synthetic Population Generation – Concept and Methodology	Ram Pendyala
12 Noon	Lunch/Break	
1:15 PM	Synthetic Population Generation – Data Requirements and Algorithm	Ram Pendyala
2:15 PM	Break	
2:30 PM	Integration of Synthetic Population Generator with Travel Models in CUBE/Voyager and with Land Use Model (PECAS)	Maren Outwater
4:00 PM	Adjourn	

Wednesday, November 10, 2010

8:15 AM	Activity-Based Microsimulation Models of Travel – An Overview	Ram Pendyala
9:30 AM	A Phased Approach to Activity-Based Model Development	Fred Ducca
10:15 AM	Break	
10:30 AM	Overview of BMC Travel Demand and Land Use Model and Databases	BMC
11:30 AM	General Discussion - BMC vision for activity-based model development - Project schedule, management, and coordination processes - Training and technical support - User documentation and deliverables	
12:15 PM	Lunch/Break	
1:15 PM	Hands-on Installation and Training on PopGen (BMC Staff Only)	Karthik K & Bhargava S
4:00 PM	Adjourn	