Revised TRANSIMS Implementation Plan for Portland

Program Support for Environment and Planning

DTFH61-04-D-00015

Draft Technical Report

Revisiting the Portland GEN2 Modeling Process with TRANSIMS Version 4.0 Software Methods

Prepared for

U.S. Department of Transportation
Federal Highway Administration

Prepared by

AECOM

August 1, 2009

Table of Contents

List of Figures

List of Tables

Preface

The Transportation Analysis and Simulation System, or TRANSIMS, is an integrated system of travel forecasting tools designed to give transportation planners accurate and complete information on traffic impacts, congestion, and pollution. It was developed by the Los Alamos National Laboratory (LANL) as part of the Travel Model Improvement Program (TMIP) sponsored by the U.S. Department of Transportation (USDOT), the Environmental Protection Agency (EPA), and the Department of Energy (DOE).

The goal of TRANSIMS is to develop technologies that offer transportation planning agencies increased policy sensitivity, more detailed vehicle-emission estimates, and improved analysis and visualization capabilities. The philosophy underlying TRANSIMS is that effective analysis of transportation systems requires detailed temporal and spatial simulation of travel conditions. The TRANSIMS Microsimulator tracks the activity patterns of individuals, households, and vehicles on a second-by-second basis in order to create a realistic representation of traffic dynamics and travel behavior. Additional information about TRANSIMS, technical documentation, and software downloads can be found on the TRANSIMS Open Source website http://transims-opensource.org.

In 2002, the Federal Highway Administration (FHWA) sponsored a project entitled "Support for the Implementation of TRANSIMS in Portland, Oregon". The project was designed as a full implementation of the TRANSIMS software for regional modeling within the Portland metropolitan area. It followed efforts by LANL to use the Portland region as a proof of concept test case for a large scale implementation of activity-based microsimulation techniques. As part of the proof of concept, LANL developed, with the assistance of Portland METRO, a regional network that contained all of the local streets and transit routes within the greater Portland metropolitan area. This "All-Streets" network was used for the initial model develop and software evaluation efforts.

In addition to the "All-Streets" implementation, FHWA pursued a parallel track designed to evaluate the feasibility of using the networks and trip tables from existing regional planning models as input to a TRANSIMS microsimulation. By comparing the results of the "Track 1" effort to the more detailed "All-Streets" implementation, FHWA could evaluate the relative costs and benefits of pursuing one approach over the other. The "Track 1" approach is a quick and cost-effective way for regional planning agencies to investigate the potential of regional simulation as an extension or enhancement to current planning practices.

The Portland study also implemented an activity-based regional model using the TRANSIMS Population Synthesizer, Activity Generator, and Router. One of the objectives of this effort was to utilize existing trip-based models to the full extent possible. If trip-based models could be re-calibrated within a TRANSIMS framework, an effective transition strategy from current practice to advanced activity-based analysis could be demonstrated. The so called "GEN2" model incorporates several components of the existing Portland regional model into a TRANSIMS model that tracks individual and vehicle movements on a second by second basis.

The original Portland study generated two reports that provide important background information for this report:

1. Using Traditional Model Data as Input to TRANSIMS Microsimulation, January 2006.
2. Generating TRANSIMS Activities from Trip-based Location and Mode Choice Models, May 2006.

A number of problems and complications using the TRANSIMS Version 3.x software were identified and worked around as much as possible as part of the original GEN2 implementation. This project addressed many of these issues through the development of the TRANSIMS Version 4.0 software. This includes new software implementations of all major components: Population Synthesis, Activity Generation, Mode Choice, Router and Microsimulator as well as numerous support tools for data processing, analysis, and presentation. This report highlights the implementation of the Portland Track 1 and GEN2 models using the TRANSIMS Version 4.0 software and compares these results to the original findings and to observed data. The primary objective of this work was to link the activity generation and location choice components of the activity model with the Microsimulator to validate the traffic assignments and feedback simulated travel times by time of day to the activity model.

Chapter 1 Overview

1.1 Original Study

The original Portland study was divided into two primary research tracks. The first track was to apply the Router and Microsimulator as a dynamic traffic assignment procedure using regional networks and trip tables. The second track was designed to implement a regional modeling process using the Population Synthesizer, Activity Generator, and Router-Microsimulator assignments.

1.1.1 Track 1 Microsimulation

Track 1 was completed successfully using incremental assignment techniques and random feedback procedures. In order to generate the necessary network throughput using the TRANSIMS 3.x Microsimulator, the network size and times of day were doubled. This doubled the fidelity of the Microsimulator for lane changing and significantly reduced the number of lost vehicle problems. The results were validated against traffic counts, but little was done to identify or confirm a user equilibrium solution.

1.1.2 Population Synthesizer and Activity Generator

Track 2 successfully implemented the Population Synthesizer and Activity Generator using the Version 3.x software. The tour-based location choice components of the Activity Generator were implemented by adjusting a trip-based location choice model previously calibrated for the Portland region. The inputs for this model included zone-to-zone logsums from the METRO EMME/2 mode choice model and zone attraction weights by trip purpose. The model was applied to independent tours in order to minimize the routing overhead and ignore the inter-tour relationships involved in modeling all tours within a household as a coordinated unit. For most trip purposes, the calibrated model was able to replicate the average trip lengths by tour type, complexity and trip purpose. A simple adjustment factor was included to balance the work attractions to employment data for downtown Portland. Other trip interchanges were not validated in any significant way.

1.1.3 Two Stage Mode Choice

Several additional software tools were developed to implement a two stage mode choice process. The first stage identified the overall tour mode and the second stage selected compatible trip modes. The stage 1 mode choice model was based on the METRO EMME/2 multi-nomial logit mode choice model. The zone-to-zone skims by mode were extracted from the EMME/2 software. These were production-attraction skims for peak and off-peak periods. The stage 1 model applied these data to the anchor activity purpose of each household tour. The model parameters were adjusted to replicate the overall mode shares to the primary trip purpose. Based on the tour mode, various rules and the relative distance to intermediate activities, the mode for each trip within a tour was selected. The model parameters were then adjusted to replicate the overall mode shares for auto, transit, and walk trips as observed in the household travel survey.

1.1.4 Transit Routing Problems

Additional work on the Track 2 model was frustrated when attempting to route the scheduled tours by the selected mode. The zone-to-zone skims used to implement and calibrate the mode choice procedures did not adequately capture the complexity and feasibility of transit trip making by time of day. In many cases the return transit trip from the primary tour purpose could not be constructed due to incompatibilities between the return trip time and the transit schedules at that time of day. If the tour included intermediate stops, it was either impossible to reach the activity location using transit or the running times and schedules for the transit option could not be completed in the specified trip duration. Efforts to relocate the intermediate stops to transit accessible activity locations helped make more of the trips feasible, but did not address the trip duration problems caused by the fixed activity schedules.

1.1.5 Feedback

A swapping algorithm was designed to exchange an infeasible transit tour with an auto tour in order to keep the regional mode shares consistent with the calibrated totals. This proved to be overly complex and theoretically suspect given the number of tours that needed to be swapped.

Another suggestion was to generate and locate each tour using an assumed mode and then base the mode choice on the relative utility of the mode-specific tours. The fact that the walk, transit, and auto tours for a given household activity pattern and schedule would not include the same activity locations or travel characteristics made the calculation and comparison of mode-specific utilities theoretically difficult to calibrate. The prior experience using zone-to-zone travel attributes to identify feasible transit options also raised concerns. It became increasingly clear that realistic transit options could only be generated from point-to-point paths for a specific time of day. The computational challenges for generating this level of detail for all potential locations on all tours was beyond the resources of the original project. Time budgets, sampling, and spatial constraints were among the concepts considered for addressing the computational challenge.

It was eventually decided that the remaining project resources would be better spent investigating the interaction of the activity generation with the Router and Microsimulator for auto-based tours. Developing methods and procedures for feeding back simulated travel times to location choice and tour scheduling was perceived as the most practical research direction.

1.2 Current Study

1.2.1 New Microsimulator Software

The original Track 1 work for Portland was performed using the TRANSIMS Version 3.x software. In order to support the needs of this and other FHWA sponsored studies, the Router and Microsimulator were re-written as part of the TRANSIMS 4.0 software suite. The new Microsimulator included improved lane changing algorithms which ultimately made it possible to simulate congested networks without using the double size and time concepts originally developed for Portland. Other research efforts also generated better procedures for implementing user equilibrium convergence techniques. As a way of quantifying the impacts of these software and procedure changes on model results, the

Portland Track 1 application was re-run using the TRANSIMS 4.0 software. The resulting travel time skims were used as input to the new GEN2 model.

1.2.1 2000 Census Data Integration

The Population Synthesizer and Activity Generator software were also re-written for TRANSIMS Version 4.0. This includes a flexible specification of population models using both 1990 and 2000 Census data. A flexible modeling structure for defining household types (e.g., CART trees) and location choice models was provided. Location choice by activity purpose can be specified using zone-based or location-based resolution and distance, travel time, or user-defined utility functions. Travel time budgets can also be used with user-defined scripts to better synchronize locations with specified trip durations. Activity time scheduling can now use mode-specific travel times rather than distance and mode-specific speeds.

The Population Synthesizer and Activity Generator models previously developed for Portland were converted to the TRANSIMS Version 4.0 software and re-calibrated. The activity generation models were then recalibrated again by replacing the zone-to-zone logsums from the EMME/2 software with zone-to-zone travel time skims by time of day from the Track 1 application. This model also generated activities for a whole household rather than split the household data into independent tours. A number of problems related to matching shared activity patterns, coordinating vehicles and replicating trip lengths by purpose were addressed as part of this effort. Unfortunately, the resulting simulation generated many problems and very low volumes.

To address these concerns, the household type model was reviewed and redefined to better match the characteristics of the synthetic households to the survey households. Logic was also added to better distribute the household persons in the survey to persons in the synthetic households by reducing the likelihood that a household member will be re-used in the person matching process. This helped to reduce vehicle conflicts and provide better coordination between drivers and passengers. This effort also uncovered a number of problems with the activities coded in the household survey.

1.2.3 Location Choice with Time Budgets

Despite these improvements, the activity generator, Router and Microsimulator still had considerable difficulty fitting the travel requirements into the activity schedule. It was at this point that travel time budgets were introduced into the activity generator script. Time budgets provide the user script with the total travel time available for the outbound and inbound legs of a tour after all intermediate activity durations are removed. This information is compared to the outbound and inbound travel times estimated for the specified travel mode from the zone-to-zone skims by time of day. If the travel times to a potential activity location are reasonably close to the travel time budget, the utility to the location is included in the location choice set. If not, the location is ignored.

After re-calibrating the activity generation model using time budget constraints and implementing 20 Router-Microsimulator feedback iterations, the resulting were much improved. The overall number of trips with time schedule or routing problems was significantly reduced. Comparisons of the simulated volumes to traffic counts showed higher volumes. The travel times from this application were summarized as zone-to-zone skims for inclusion in the next round of feedback to the activity generator.

1.2.4 Feedback from the Microsimulator

To facilitate feedback from the Microsimulator to activity generation, the ActGen program was modified to permit re-generation of selected households. This enabled the feedback strategy to incorporate the concepts that have proved effective in feeding back travel times from the Microsimulator to the Router. This strategy involves re-generating the activities for a small percentage of the households rather than all of the households. The households selected for the first few rounds of feedback are selected because they were identified as problematic tours by the activity generator, Router or Microsimulator. A list of problem households is passed to the ActGen program for re-generation. The synthetic household is re-matched to a survey household and the activities generated. Minor adjustments to the location choice model are considered at this point to help improve the overall average trip lengths. The new activities are added to the regional activity set and re-assigned through the Router-Microsimulator process.

Several rounds of problem feedback were implemented to generate feasible activity patterns for nearly all of the households. The travel times that result from simulating these activities were used to recalculate the travel time skims. The overall activity generation model was fine-tuned using these travel times, incrementally merged with the original travel plans and re-simulated. After each simulation, the travel times were updated and used to generate the next set of activities. Once all of the households were updated, a second round of incremental updates was implemented.

1.2.5 Additional Model Refinements

Unfortunately, several iterations of feedback between the activity generator and the Microsimulator did not produce a stable solution. This was in large part due to the instability of the zone-to-zone travel times fed back to the location choice model from the Microsimulator. Averaging methods were applied to dampen these congestion affect on location choice with minimal success. The Microsimulator congestion levels also had detrimental effects on the ability to load and simulate multi-stop tours. A large percentage of travelers experienced time schedule errors caused by simulated travel times that far exceeded the expected travel time in the plan file. This resulted in incomplete tours and vehicles not arriving at the planned location at the expected time. Once the tours started experiencing problems the traffic during the later hours of the day was less than expected and therefore the travel times and zone-to-zone skims were less than they should be.

Validation checks against screenline traffic counts and comparisons of the district-to-district travel demand generated by ActGen to the trip tables used in Track 1 showed significant differences. These differences generated significant congestion crossing the Columbia River and in Northeast Portland. Attraction weight adjustments were included to balance the generated demand to the employment by zone. This improved the distribution of trips considerably, but it still did not address the Microsimulator problems. It was not until the trips in the tours were unlinked and made independent so the paths could be equilibrated without schedule complications that the travel times stabilized.

Chapter 2 Network Simulation

To make the activity-based demand models useful within the TRANSIMS process, the demand must ultimately be loaded to the TRANSIMS network and the simulated travel times must be fed back to the activity models to adjust the demand and equilibrate the results. The network simulation components of the GEN2 model build on the Track 1 procedures developed as part of the original Portland study. This task also includes generating travel time skims from the Microsimulator output to feed back to the location choice model.

2.1 Update the Track 1 Model

As a part of the original Portland study, AECOM and METRO successfully converted the existing regional networks and trip tables to TRANSIMS and utilized the Router and Microsimulator to replicate observed traffic counts. This approach is commonly referred to as a Track 1 model to imply that it is the first phase of a multi-phase effort to implement a complete TRANSIMS application. The GEN2 model replaces the trip tables in the Track 1 process with activity-based demand models and then loads this demand to the network using the Track 1 Router-Microsimulator procedures.

The original Track 1 work was performed using the TRANSIMS Version 3.x software. This software contained a number of constraints that made the application more complicated than it needed to be. The most significant work-around was an artificial scaling of link lengths, activity times, vehicle sizes and simulation duration to increase the fidelity and throughput of the Microsimulator under congested conditions. The TRANSIMS Version 4.0 improvements addressed these constraints and streamlined the Track 1 procedures. Since this study uses the Version 4.0 software to implement the activity-based models, it was useful to re-run the Track 1 process with the new TRANSIMS software as well.

The key changes to the original Portland Track 1 model are outlined and discussed below:

1. Remove network and time scaling factors
2. Reduce the cell size and maximum link speeds
3. Integrate the new capabilities of the Version 4.0 software
4. Apply improved equilibrium convergence methods

2.1.1 Rescaling

In the original Portland application, the network and activity times were scaled by a factor of two in order to increase the fidelity of the Microsimulator. In a rescaled simulation run, link lengths, activity times, vehicle sizes, and simulation duration are all twice as long as normal. This enabled the Microsimulator to make lane changing decisions every one-half second rather than every second and thereby complete turning movements under congested conditions more effectively. It also made creating and interpreting model inputs and outputs considerably more complicated and substantially increased the computer processing time.

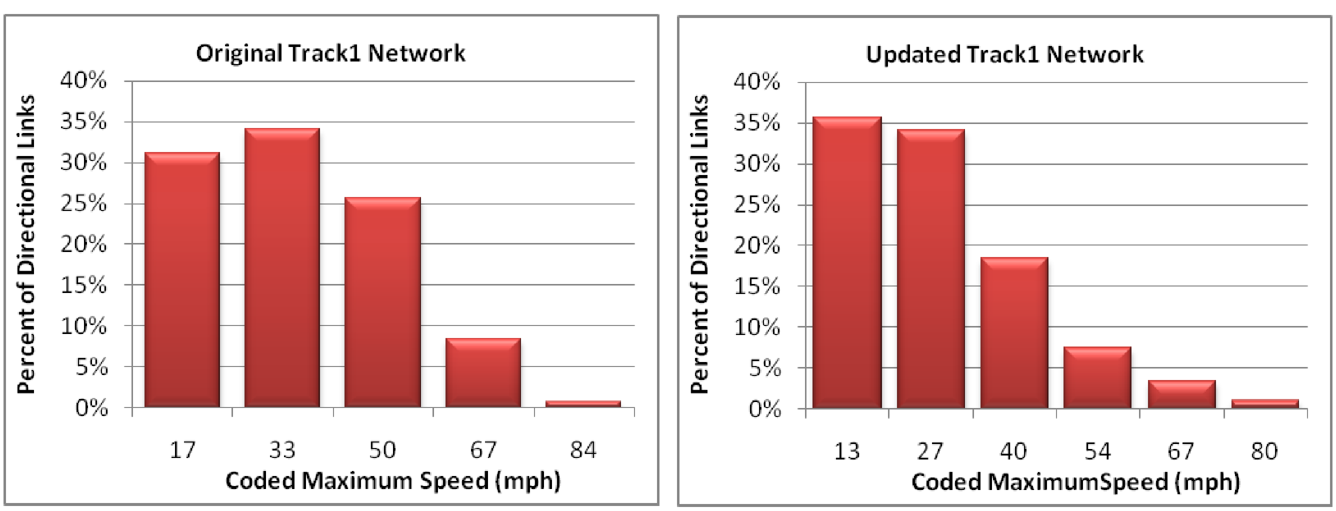
The Version 4.0 software enables the modeler to request sub-second simulations, but it also includes several changes to the simulation algorithms to make this unnecessary. The lane changing algorithm now enables vehicles to move right or left at every time step rather than alternate between right and left on odd or even time steps. It also includes new concepts such as permissive probability, lane swapping and priority movements that enable vehicles to coordinate and resolve conflicting or competing movements. As a result of these features, the Portland TRANSIMS network and trip tables were converted back to their normal, unfactored state, and simulated using one second time steps. This makes the whole process much cleaner and easier to work with and take less time to run.

2.1.2 Network Changes

The Version 3.x Microsimulator also had a fixed the simulation cell size of 7.5 meters. A fixed cell size with a fixed time step of one second result in six speed increments (cells / second) within the simulation. In other words, the vehicles can travel at 0.0, 7.5, 15.0, 22.5, 30.0 or 37.5 meters per second (or 0, 17, 33, 50, 67, or 84 mph). In order to compensate for some of the limitations of the Microsimulator and replicate observed counts, freeways and major arterials were coded with higher maximum speeds than we would have liked. For example, the freeway links in the TRANSIMS network were coded with maximum speeds of either 30.0 meters per second (67 mph) or 37.5 meters per second (84 mph) even though the speed limits were 55 mph or 65 mph.

The Version 4.0 Microsimulator still uses the cells per second concept, but adds the flexibility to adjust the cell size and calibrate random slow down probabilities and driver reaction time parameters to control the simulated speeds more effectively. Research done as part of other studies found that a 6.0 meter cell size simulated congested networks reasonable well. This provides the Microsimulator with seven speed increments of 0.0, 6.0, 12.0, 18.0, 24.0, 30.0, and 36.0 meters per second (or 0, 13, 27, 40, 54, 67, and 80 mph) that more closely reflect to the range of free flow speeds observed in the real world. Figure 1 shows the number of links coded with each of the speed bins available in the original and the updated Portland network. The overall goal was to make the speeds generated by the Microsimulator more reasonable (i.e., lower).

Figure 1: Distribution of Coded Maximum Speeds



2.1.3 New Software Capabilities

Several of the new parameters available in the TRANSIMS Version 4.0 software were adjusted and calibrated to improve the Track 1 results. One key difference was the addition of the value of distance parameter in the Router. The distance parameter adds incentives to select a path that minimizes trip length in addition to minimizing travel time. The result is a minimum impedance path that considers both time and distance in the path choice. A distance value of 0.4 impedance units per meter was added to the in-vehicle time value of 10 impedance units per second to build the minimum impedance path. This means that the traveler would be willing to add 1500 meters (~1 mile) to their trip length to save one minute of travel time.

In addition to the distance parameter, the Router was configured to restrict the use of local facility types for the line-haul portion of the trip and penalize turning movements. In this case, local links cannot be used by the path builder if the link is more than 1,500 meters (~1 mile) away from the trip origin or destination. This parameter implies that travelers have imperfect knowledge and avoid path options that wind through local neighborhoods when they are far away from home or their trip destination.

Turning penalties were also added for right, left and U-turns. A U-turn was valued at 500 impedance units, a left turn at 100 impedance units, and a right turn at 50 impedance units. This is equivalent to 50, 10, and 5 seconds of in-vehicle travel time. These penalties help the Router construct paths that are more direct and less susceptible to lane changing problems in the Microsimulator.

The Version 4.0 Microsimulator also includes several new parameters to control the gap acceptance and lane changing behaviors. The driver reaction time, slow down probability, look ahead weights, and permissive probability parameters were calibrated to reflect the driving conditions in Portland and produce reasonable flows.

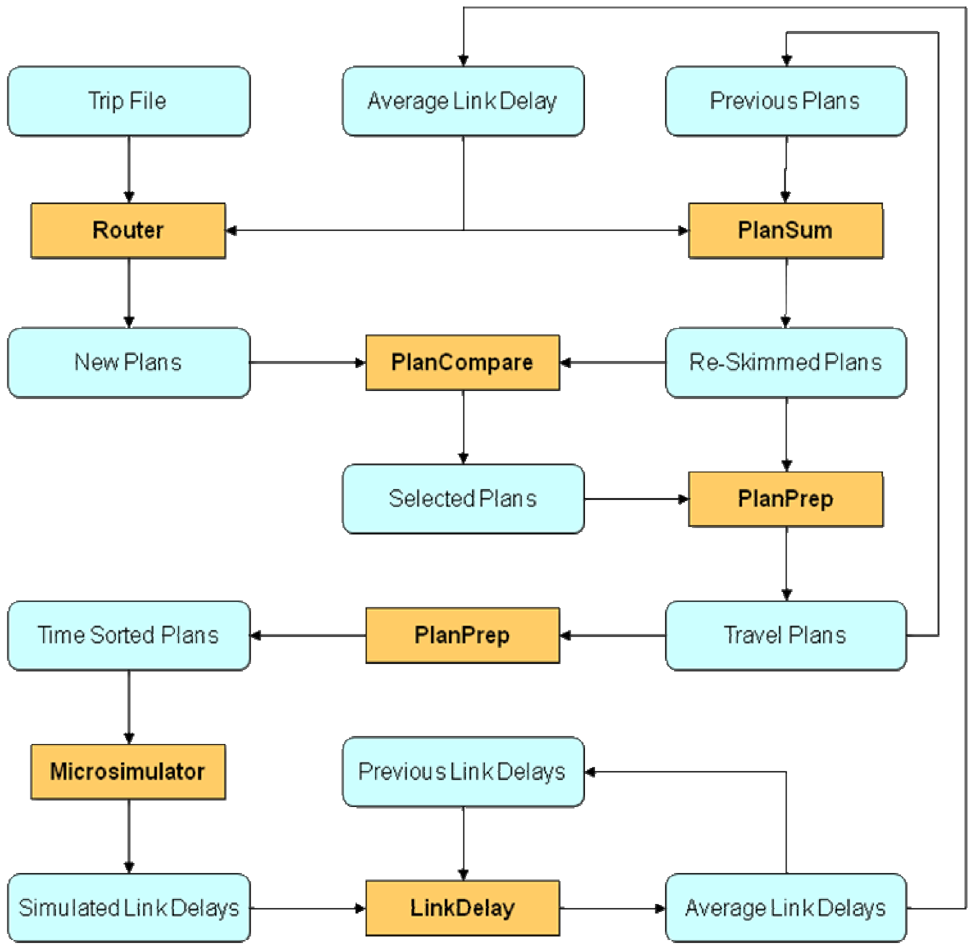
2.1.4 Feedback Mechanism

TRANSIMS uses iterative feedback mechanisms to adjust and refine travel paths based on simulated travel times and turning movement delays. The original Track 1 model used an incremental loading technique where a specific percentage of the trips were loaded onto the network and simulated. After all of the trips are loaded, a series of Router and Microsimulator feedback loops are executed until the results stabilize.

More recent implementations have developed feedback and convergence algorithms based on user equilibrium concepts. These iterations start from the link delays in 15 minute increments generated by the original Track 1 effort and re-routes all of the travelers using these travel times. The new travel plans are then compared to the previous travel plans to identify travelers whose path or total travel time are significantly different. A percentage of the travelers with the highest differences are selected for inclusion in the full plan file for the next simulation. This process continues until the number of travelers with significant differences is small. This would be defined as user equilibrium within the TRANSIMS context.

The overall process is outlined in Figure 2. This flow chart shows the primary programs and data sets used by the equilibrium convergence process. It also shows that the link delays are averaged after each iteration before they are fed back to the Router. The LinkDelay program uses an MSA (method of successive averaging) technique to generate weighted average link travel times and turning movement delays for each 15 minute time increment. Link delay averaging has been found to be an effective way of dampening oscillation affects between parallel roadways and speed up the convergence process.

Figure 2: Equilibrium Convergence Process

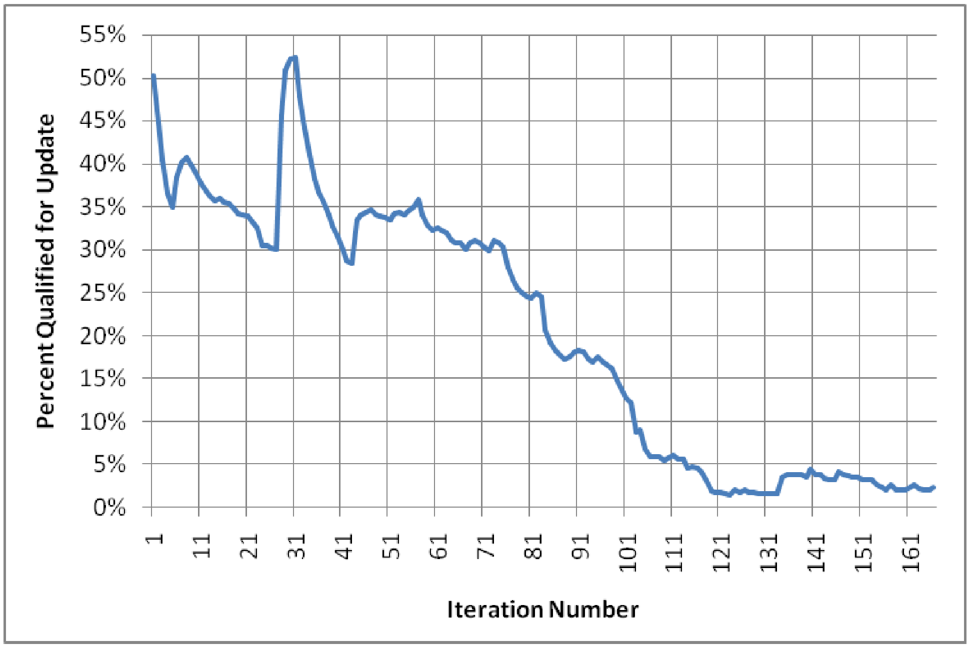


Convergence Statistics

One of the statistics used to evaluate the convergence process is the number of travelers with significantly different travel times between the current simulation and the previous simulation. This value is known as the percent qualified for re-routing. In order to qualify for re-routing the travel times on the two paths must be different by five percent or a minimum of one minute. At any point in the feedback process only a subset of the qualified travelers are integrated into the full plan file for the next simulation. For this model, no more than 50 percent of the qualified travelers or five percent of all travelers are selected at any one time.

The results of applying the user equilibrium convergence process to the Portland regional network are shown in Figure 3. This chart shows that the percent qualified for re-routing starts around 50 percent and after about 120 feedback iterations is down below three percent. Selective network refinements to address problem locations were made along the way to help improve the simulation. These changes often disrupt the convergence for several iterations while the travelers adjust to the changing conditions. In this case, relatively few network changes were made in order to simplify model comparisons.

Figure 3: Router-Microsimulator Convergence



Lost Vehicles

In addition to user equilibrium, a converged TRANSIMS model should also have network performance stability. One of the key measures of network stability is the number of lost vehicles. Lost vehicles are vehicles that are removed from the simulation because they violate one of the Microsimulator problem constraints. Vehicles may be removed from the simulation when they fail to load onto the network or complete their trip in a specified amount of time; there is an inconsistency between their travel plan and the current network; or they are unable to move from their current positions for a period of time. The Microsimulator removes these vehicles from the simulation to avoid cascading queues that impact the network performance and other vehicles unreasonably.

In many cases, lost vehicles can be addressed through additional feedback iterations that enable travelers to select paths that avoid heavily congested areas. In other cases, lost vehicles point to a network coding problem that needs to be fixed for the Microsimulator to replicate the real world effectively. For this study, very few network corrections were made to address lost vehicle problems. The objective was to apply the Version 4.0 Microsimulator to the same network used with the Version 3.x Microsimulator to provide a fair assessment of the software and methodological improvements on the model results. The original Track 1 model produced in excess of 250,000 lost vehicles per day (or 5 percent of total trips). Rescaling the network improved the results dramatically and the number of problems was ultimately reduced to about 70,000 (or 1.4 percent or total trips). The updated Track 1 model using the Version 4.0 Microsimulator and improved convergence methods reduced the lost vehicles to 50,000 per day (or 1.0 percent or total trips). One percent lost vehicles is considered good for a regional simulation of this type.

2.2 Validation

Figure 4 and Figure 5 below show the daily validation of the original Portland Track 1 simulation and the METRO regional assignment compared to 1996 daily traffic counts. The model replicated the observed counts reasonably well in total, but the volume assigned to freeways and major arterials was higher than the counts and the METRO assignment. In addition, the root mean squared error (RMSE) statistics by functional class and for the total region were reasonable, but not as good as the METRO assignment (23 percent versus 35 percent).

Figure 4: Original Track 1 Daily Validation by Functional Class

Summary Statistics by Functional Class

Functional Class	Num. Links	Total Volume	Total Count	Difference	% Diff	Avg. Error	Avg % Error	Std. Dev.	% RMSE
Freeway	52	3025656	2874830	150826	5.2	6427	11.6	5313	15.0
Principal Arterial	24	505135	409706	95429	23.3	5091	29.8	3978	37.5
Major Arterial	104	1312825	1143236	169589	14.8	3525	32.1	2739	40.5
Minor Arterial	82	401095	507048	-105953	-20.9	2900	46.9	2268	59.4
Collector	130	139455	462788	-323333	-69.9	2736	76.9	2064	96.1
Other	18	284068	286236	-2168	-0.8	3852	24.2	2727	29.4
TOTAL	410	5668234	5683844	-15610	-0.3	3624	26.1	3234	35.0

Figure 5: METRO Daily Assignment Validation by Functional Class

Summary Statistics by Functional Class

Functional Class	Num. Links	Total Volume	Total Count	Difference	% Diff	Avg. Error	Avg % Error	Std. Dev.	% RMSE
Freeway	52	2845091	2874830	-29739	-1.0	3913	7.1	3149	9.1
Principal Arterial	24	459029	409706	49323	12.0	2651	15.5	2893	22.7
Major Arterial	104	1111055	1143236	-32181	-2.8	1825	16.6	1392	20.8
Minor Arterial	82	451512	507048	-55536	-11.0	2442	39.5	1619	47.3
Collector	130	431111	462788	-31677	-6.8	1920	53.9	1563	69.4
Other	18	331709	286236	45473	15.9	3785	23.8	2818	29.4
TOTAL	410	5629507	5683844	-54337	-1.0	2378	17.2	2086	22.8

Figure 6 shows the same validation statistics for the updated Track 1 model. The overall percent error is greater than both the original Track 1 model and the METRO assignment, but the RMSE errors are less. The freeway assignment have also improved to be comparable to the METRO results. In general, the METRO assignment is still better than the updated Track 1 model. It is clear from examining the network performance, that network refinements at a number of key locations would noticeably improve these results. Once again, these types of improvements would bias the assessment of the software and methodological changes on the model results.

Figure 6: Updated Track 1 Daily Validation by Facility Type

Summary Statistics by Facility Type

Facility Type	Num. Obs.	------Volume------ Estimate	Observed	---Difference--- Volume	%	--Abs.Error-- Avg.	%	Std. Dev.	% RMSE
Freeway	51	2843610	2874828	-31218	-1.1	4489	8.0	3855	10.5
Expressway	24	511466	409706	101760	24.8	4918	28.8	4198	37.5
Major Arterial	104	1333357	1143236	190121	16.6	2971	27.0	2426	34.8
Minor Arterial	82	542767	507048	35719	7.0	2468	39.9	1955	50.8
Collector	130	392231	462788	-70557	-15.2	1959	55.0	1570	70.4
Bridge/Other	18	300580	286236	14344	5.0	3105	19.5	2681	25.5
TOTAL	409	5924011	5683842	240169	4.2	2858	20.6	2648	28.0

Tier 1 Screenlines

Model validation at Portland METRO focuses on a subset of screenlines identified as Tier 1 screenlines. These screenlines represent the boundaries between the eight districts defined in the Portland modeling area. METRO compared the simulation results to the daily and two hour PM peak period Tier1 screenline counts. Figure 7 shows the detailed validation statistics for each of these screenlines. Table 1 shows a comparison of the METRO assignment, the original Track 1 model, and the updated Track 1 model for the Tier 1 screenlines.

The original Track 1 model has the least overall difference at 3.3 percent, but the absolute difference is higher than the METRO assignment due to several largest positive and negative differences. The updated Track 1 model reduces the large negative error on the I-205 Columbia River Bridge from 21.6 percent to 15.3 percent, but significantly increase the positive error on the two East side screenlines. The largest positive error is 19.4 percent in the Western suburbs. The Willamette River bridges in downtown were 8.1 percent high. This improved to 2.4 percent in the new TRANSIMS model. The majority of screenlines in the new Track-1 model has errors in low single digits. In general, the new Track-1 screenline validation results are reasonable when compared to the counts.

Figure 7: Updated Track 1 Daily Validation by Screenline

Link Group	Num. Obs.	------Volume------ Estimate	Observed	---Difference--- Volume	%	--Abs.Error-- Avg.	%	Std. Dev.	% RMSE	R Sq.
R-1 St. John's Br	2	23167	22610	557	2.5	924	8.2	394	8.5	0.000
R-2, William. River Bridg	16	489129	464966	24163	5.2	4018	13.8	2618	16.3	0.975
R-3, Sellwood Bridge	2	33248	31382	1866	5.9	1447	9.2	1319	11.0	0.000
R-4 OrCty 205 Br	4	120052	109571	10481	9.6	2784	10.2	1906	11.8	0.994
R-5, I-5 Bridge	2	121148	113838	7310	6.4	3655	6.4	895	6.5	0.000
R-6 I-5 Hayden	2	126306	128300	-1994	-1.6	997	1.6	1372	2.2	0.000
R-7 I205	2	85679	104074	-18395	-17.7	9198	17.7	1459	17.8	0.000
E-9 W of 122nd	34	398186	348370	49816	14.3	2970	29.0	3020	41.0	0.912
E-21 W of Tacoma	20	285016	272032	12984	4.8	3907	28.7	3559	38.4	0.941
W-7 Westhills	10	165133	172226	-7093	-4.1	1684	9.8	1635	13.3	0.994
W-09 Tigard Int	6	151630	154962	-3332	-2.2	3035	11.8	2315	14.3	0.975
W-16 NW of Tigard	10	163593	140922	22671	16.1	3004	21.3	2747	28.2	0.984
TOTAL	110	2162287	2063253	99034	4.8	3201	17.1	2915	23.0	0.962

Table 1: Tier1 Screenline Daily Volume Comparisons

Tier 1 Screenline	Num. Obs.	Daily Counts	METRO			Original Track-1			New Track-1		
			Estimate	Diff	%	Estimate	Diff	%	Estimate	Diff	%
R-1 St. John's Br	2	22,610	22,829	219	1.0%	20,337	-2,273	-10.1%	22,491	-119	-0.5%
R-2 Williamette River	16	464,966	490,575	25,609	5.5%	502,649	37,683	8.1%	467,142	2,176	0.5%
R-3 Sellwood Bridge	2	31,382	39,715	8,333	26.6%	28,704	-2,678	-8.5%	32,605	1,223	3.9%
R-4 OrCty 205 Br	4	109,571	128,572	19,001	17.3%	123,308	13,737	12.5%	116,833	7,262	6.6%
R-5 I-5 Bridge	2	113,838	127,737	13,899	12.2%	117,385	3,547	3.1%	116,636	2,798	2.5%
R-6 I-5 Hayden	2	128,300	134,904	6,604	5.1%	124,273	-4,027	-3.1%	120,584	-7,716	-6.0%
R-7 I205	2	104,074	99,780	-4,294	-4.1%	81,551	-22,523	-21.6%	84,609	-19,465	-18.7%
E-9 W of 122nd	34	348,370	339,296	-9,074	-2.6%	363,304	14,934	4.3%	392,340	43,970	12.6%
E-21 W of Tacoma	20	272,032	275,566	3,534	1.3%	280,889	8,857	3.3%	278,688	6,656	2.4%
W-7 Westhills	10	172,226	178,463	6,237	3.6%	164,781	-7,445	-4.3%	160,794	-11,432	-6.6%
W-09 Tigard Int	6	154,962	154,966	4	0.0%	155,573	611	0.4%	148,138	-6,824	-4.4%
W-16 NW of Tigard	12	144,000	164,597	20,597	14.3%	171,911	27,911	19.4%	156,952	12,952	9.0%
TOTAL	**110**	**2,066,331**	**2,157,000**	**90,669**	**4.4%**	**2,134,665**	**68,334**	**3.3%**	**2,097,812**	**31,481**	**1.5%**
Absolute Difference				**117,405**	**5.7%**		**146,226**	**7.1%**		**122,593**	**5.9%**

2.3 Generate Travel Time Skims for Location Choice

Track 1 assignment results (plans) can be aggregated into travel time skims by time period using the PlanSum program. However, METRO trip tables do not include trips from all origins to all destinations at all times of day. Over 60 percent of origin-destination-time period combinations do not have trips or travel plans. This makes it difficult to use these travel time skims in the location choice model to identify activity locations. The trip interchanges will be constrained to the Track 1 travel patterns.

The problem was addressed by synthesizing a complete skim file using the Router control keys. The first task is to identify activity locations to represent the centroid of each zone. The Router is then executed with the centroid activity locations as the list of trip origins and destinations. Eight time periods were selected to represent the time of day differences in network travel times. The Router builds a mode specific path between each zone centroid using eight start times and Track 1 link delays. The resulting plan file was then read by PlanSum to generate the zone-to-zone travel time and distance skims.

Figure 8 summarizes the overall process. The Router control keys are shown in Figure 9 and the PlanSum control keys are shown in Figure 10.

Figure 8: Travel Time Skims for Location Choice

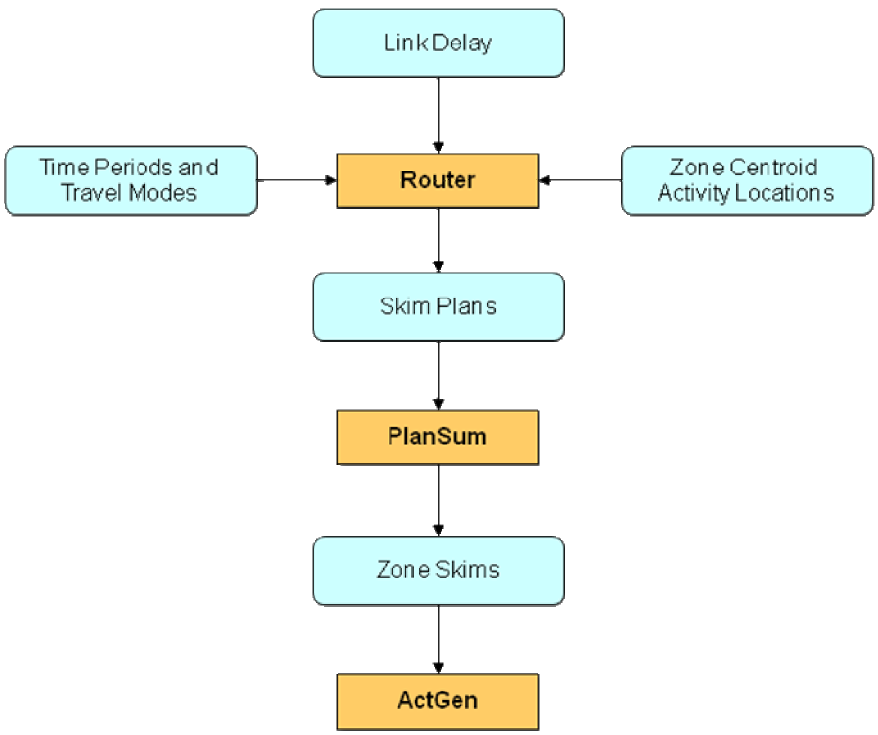


Figure 9: Router Control Keys for Zone-to-Zone Travel Times

```
TITLE                                   Plans for Building Skims by Drive Modes

#-- input --#

LINK_DELAY_FILE                         ../msim/Track1_LinkDelay

#-- output --#

NEW_PLAN_FILE                           ../plans/Drive_Paths.t*
NEW_PLAN_FORMAT                         BINARY
NEW_PROBLEM_FILE                        ../output/Drive_Problems.t*

#-- network --#

NET_DIRECTORY                           ../network
NET_NODE_TABLE                          Node
NET_LINK_TABLE                          Link
NET_LANE_CONNECTIVITY_TABLE             Lane_Connectivity
NET_PARKING_TABLE                       Parking
NET_ACTIVITY_LOCATION_TABLE             Activity_Locations_One_per_Zone_For_Skims
NET_PROCESS_LINK_TABLE                  Process_Link_For_Skims
NET_LANE_USE_TABLE                      Lane_Use
NET_TURN_PROHIBITION_TABLE              Turn_Prohibition

#-- parameters --#

ROUTE_FROM_SPECIFIED_LOCATIONS          ALL
ROUTE_TO_SPECIFIED_LOCATIONS            ALL
ROUTE_AT_SPECIFIED_TIMES                0:00..24:00
ROUTE_BY_TIME_INCREMENT                 180
ROUTE_WITH_SPECIFIED_MODE               2
TIME_OF_DAY_FORMAT                      SECONDS

WALK_PATH_DETAILS                       NO
IGNORE_TIME_CONSTRAINTS                 YES
END_TIME_CONSTRAINT                     60
IGNORE_ROUTING_PROBLEMS                 NO

WALK_SPEED                              1.0
WALK_TIME_VALUE                         20
VEHICLE_TIME_VALUE                      10
DISTANCE_VALUE                          0.6
COST_VALUE                              28.9855     #--- $12.42/hr (1994 $s) ---

MAX_WALK_DISTANCE                       2000
MAX_CIRCUITY_RATIO                      0

U_TURN_PENALTY                          500
LEFT_TURN_PENALTY                       100
RIGHT_TURN_PENALTY                      0

LOCAL_ACCESS_DISTANCE                   1500
LOCAL_FACILITY_TYPE                     LOCAL
```

Figure 10: PlanSum Control Keys for Zone-to-Zone Skims

```
TITLE                              Generate Zone-to-Zone Travel Time Skims

#-- input --#

PLAN_FILE                          ../plans/Drive_Paths.t*
PLAN_FORMAT                        BINARY

LINK_DELAY_FILE                    ../msim/Track1_LinkDelay

#-- output --#

NEW_ZONE_SKIM_FILE                 ../skims/Track1_Drive_Skims_8_Periods
NEW_ZONE_SKIM_FORMAT               TAB_DELIMITED

#-- network --#

NET_DIRECTORY                      ../network
NET_NODE_TABLE                     Node
NET_LINK_TABLE                     Link
NET_LANE_CONNECTIVITY_TABLE        Lane_Connectivity
NET_PARKING_TABLE                  Parking
NET_LANE_USE_TABLE                 Lane_Use
NET_TURN_PROHIBITION_TABLE         Turn_Prohibition
NET_ACTIVITY_LOCATION_TABLE        Activity_Location_One_per_Zone_For_Skims
NET_PROCESS_LINK_TABLE             Process_Link_For_Skims

#-- parameters --#

SUMMARY_TIME_PERIODS               0:00..24:00
SUMMARY_TIME_INCREMENT             180
TIME_OF_DAY_FORMAT                 SECONDS

SKIM_TOTAL_TIME                    YES
SKIM_TRIP_LENGTH                   YES
NEAREST_NEIGHBOR_FACTOR            0.5
```

Chapter 3 Transit Path Building

One of the findings of the original Portland study was the sensitivity of the tour-based mode choice process to the feasibility of point-to-point transit paths. This study investigated these issues further by calibrating the TRANSIMS Router parameters using observed transit trips and then validating the parameters by comparing the ridership estimates generated from METRO transit trip tables against Tri-Met ridership counts.

3.1 Calibration of Impedance Parameters

The transit path building parameters used by the Router were calibrated using the 1994 household activity survey. The survey contained approximately 510 tours that reported using transit. These tours accounted for a total of 970 transit trips. Out of these transit trips, 95 trips had the same starting and ending location and were therefore ignored by the Router. The 875 remaining transit trips were iteratively routed using various combinations of path building parameters until a reasonable match between the reported paths and the modeled paths was achieved. One of the primary objectives was to match the observed transfer rate of 1.34 transit boardings per trip.

The key impedance parameters adjusted as part of the calibration included:

1. Coefficient for Walk Time
2. Coefficient for In-Vehicle Time
3. Coefficient for Transfer Wait
4. Coefficient for First Wait
5. Transfer Penalty
6. Maximum Walking Distance

The top portion of Table 2 shows the software defaults and the adjusted impedance values used in the various iterations during the calibration process. The bottom section shows the effect of these changes on the transfer ratio and distribution.

In the first run (Run 1), the number of transfers was high because each transit trip required two or more transfers. This is reflected in the high transfer ratio for that run. Travelers took transit for very short legs, which in reality could be completed by walking. Also, there were many long walks (~2 miles) during off-peak periods. The walk weight was increased to 25 for the next run and a transfer penalty of 1 minute was introduced. The results for Run 2 were better than Run 1 with about a 50 percent reduction in overall number of transfers. Still, there were many trips with more than two transfers. So, only the transfer penalty was increased for the next few runs to bring the transfer ratio closer to 1.34. Even though Run 4 had the desired transfer ratio, Run 5 was made to test the effect of further increase in transfer penalty on the transfer ratio. Since there were still many long walks in Run 5, the maximum walking distance was reduced to a more reasonable value of 2,000 meters (i.e., the total walk distance in a transit path cannot be greater than 1.25 miles). Since the transfer penalty was found to be high, Run 7 attempted to adopt FTA New Starts guidelines in the path building process with out-of-vehicle to in-vehicle time ratio of 2.5. Because of the high first wait value, there was evidence of travelers taking the first available transit route and transferring to the more logical transit route later instead of just waiting for the transit route. So, the transfer and first wait values were reset to the default values for the last run.

The number of trips with all-walk (i.e., no transit boarding in their plan) appeared to be sensitive to the first weight and transfer weight parameters based on the increase in the all-walk trips in Runs 7 and 8. It is worth noting that the average walk distance for trips in Run 8 was 0.26 miles and more than 85 percent of the all-walk trips had a walk less than 0.52 miles, which seems reasonable.

As can be seen from Table 2, the default values of most of the parameters with a transfer penalty of three minutes (1,800 impedance units) were sufficient to produce the desired results. Since the survey data is only a sample, the impedance parameters may require additional adjustments when applied to a larger data set.

Table 2: Router Impedance Parameter Calibration

Router Calibration Run	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Impedance Parameters								
Walk Time Value (imp/second)	20	20	25	25	25	25	25	25
Vehicle Time Value (imp/second)	10	10	10	10	10	10	10	10
First Wait Value (imp/second)	20	15	15	15	15	15	15	25
Transfer Wait Value (imp/second)	20	15	15	15	15	15	15	25
Transfer Penalty (impedance)	0	0	600	1,200	1,800	2,400	2,400	1,800
Max Walk Distance (meters)	2,000	4,000	4,000	4,000	4,000	4,000	2,000	2,000
Transit Trips								
Unlinked Trips (tranist boardings)	1,734	1,288	1,163	1,094	1,044	1,023	941	964
Linked Trips (observed = 875)	807	809	809	809	809	783	701	722
Transfer Ratio (observed = 1.34)	2.15	1.59	1.44	1.35	1.29	1.31	1.34	1.34
Trips Routed as All Walk	20	20	18	18	18	18	93	93
% Trips with a valid path	95%	95%	95%	95%	95%	92%	91%	93%
Transfer Distribution								
% 0 Transfer Trips	27.6%	50.8%	60.7%	67.4%	72.4%	71.0%	67.8%	68.0%
% 1 Transfer Trips	39.8%	40.2%	35.1%	30.2%	26.2%	27.3%	30.4%	30.5%
% 2 Transfer Trips	22.7%	8.0%	4.0%	2.3%	1.2%	1.7%	1.7%	1.5%
% 3+ Transfer Trips	9.9%	1.0%	0.2%	0.1%	0.1%	0.0%	0.1%	0.0%

3.1.1 Survey Routing Problems

Initially, all of the routing problems were due to either walk access problem or the maximum walk distance criterion. The origin and destination activity locations of about 40 survey trips were modified before Run 1 to correct these access problems. In most cases, the origin activity location was attached to a link that did not permit walk access (e.g., freeway/expressway/ramp). In these cases, the origin or destination activity location was modified (wherever possible) to a nearby location where walk access to a transit stop was possible. After the adjustments, almost all of the routing problems were due to the maximum walk distance limitation.

3.2 Transit Model Validation

The initial TRANSIMS calibration exercise was performed using observed transit trips from the 1994 household survey. The calibrated parameters were then applied to the regional transit trips from the METRO model. The Router parameters were adjusted to replicate the ridership targets METRO used to calibrate the 1994 regional model. TRANSIMS path and travel time attributes were then compared to the regional model data. The final results were validated against the line ridership estimates from the 1994 on-board survey.

3.2.1 Converting Trip Tables

The first step in validating the regional transit trips is converting the METRO transit trip tables to TRANSIMS format so that they can be routed. The ConvertTrips program uses diurnal distribution curves by trip purpose to convert zone-to-zone transit trip tables from the regional model to TRANSIMS trips from specific activity locations by second of the day.

For this exercise, the walk-to-transit trip tables from the regional model were aggregated into two groups, 1) Home Based Work and 2) all other purposes. These tables were converted from production-attraction format to origin-destination format using METRO's P-A and A-P split factors. The Home Based Work and Home Based Other park-&-ride (PNR) trips were combined into a single table, but split factors were not applied. ConvertTrips requires park-&-ride trips to be input in production-attraction format in order to link the inbound and outbound trips for each traveler. This ensures that the return trip will retrieve the car from the park-&-ride lot selected by the outbound trip.

METRO's production-attraction factors are listed in Table 3. The diurnal distribution curves by trip type are presented in Figure 11, Figure 12 and Figure 13.

Table 3: Production-Attraction Factors

Trip Type	*P-A Factor*	*A-P Factor*	*P-A Trips*	*A-P Trips*	*Total*
Home Based Work	0.54	0.46	37,125	31,610	68,735
Other Walk-to-Transit	0.46	0.54	48,700	57,160	105,860
Park-and-Ride	0.50	0.50	8,019	8,019	16,038
Total	**-**	**-**	**93,844**	**96,789**	**190,633**

Figure 11: Home-Based-Work Diurnal Distribution Curves

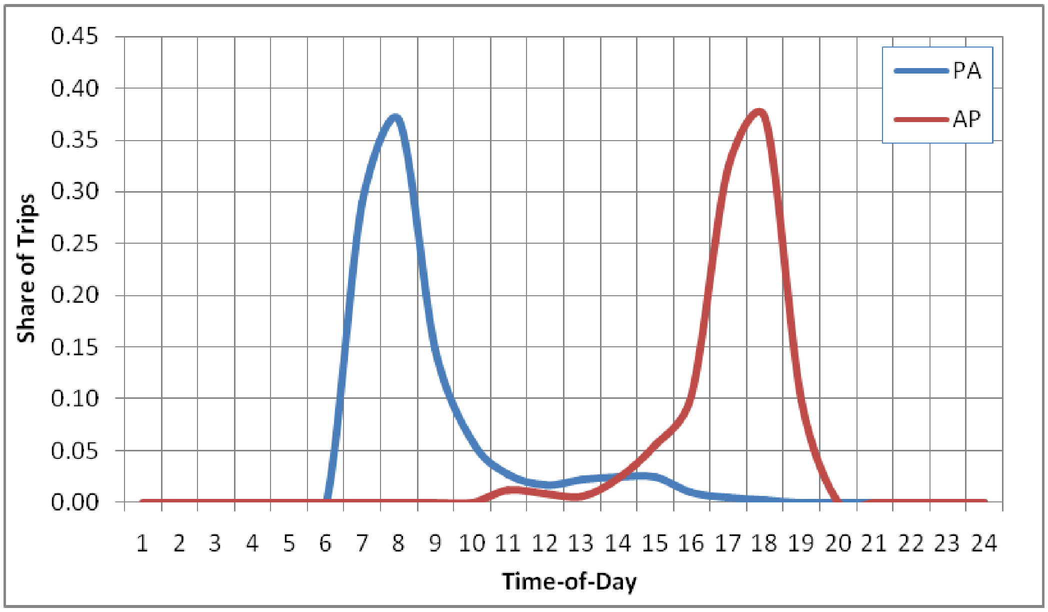


Figure 12: Other Transit Diurnal Distribution Curves

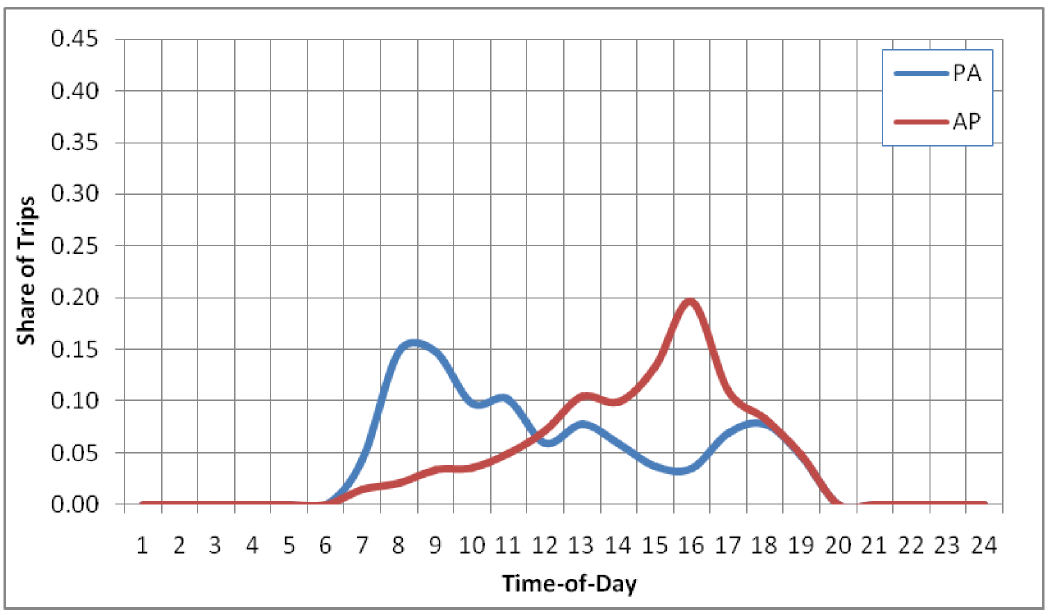
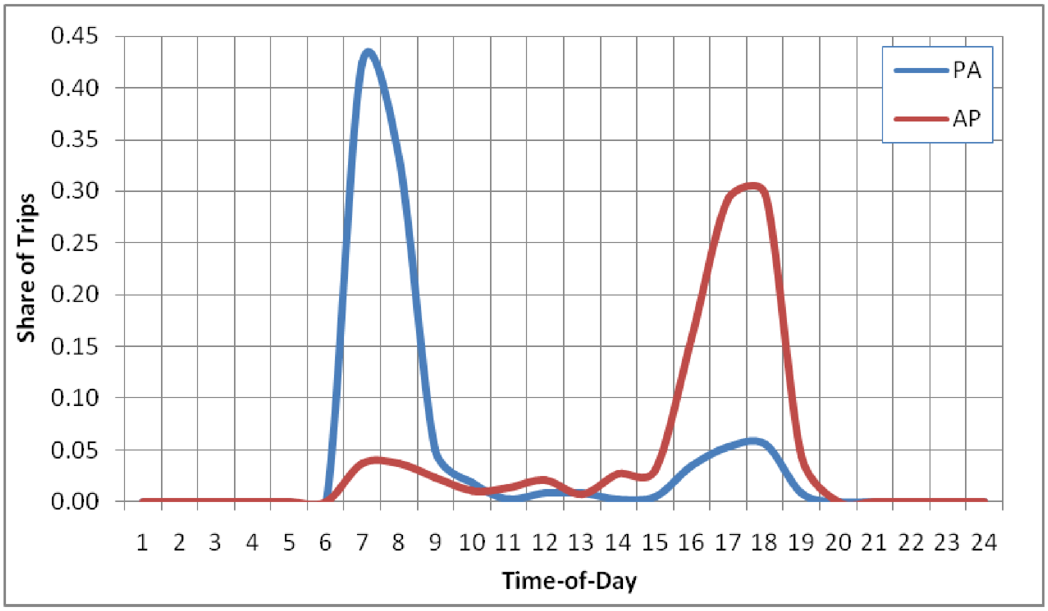


Figure 13: Park-&-Ride Diurnal Distribution Curves



The distribution of trips to activity locations within the origin and destination zone was controlled by the origin and destination weight fields. A non-zero weight value was assigned to all activity locations within 750 meters (0.5 mile) of a transit stop. For walk-to-transit trips, both the origin and destination locations must have walk accessibility. Park-&-Ride trips select from auto accessible activity locations in the origin zone and walk accessible locations in the destination zone.

3.2.2 Ridership Targets

The impedance parameters calibrated using the travel survey data were used as the starting point for routing the regional transit trips. The trips were routed separately by trip group (HBW, other walk-to-transit, and park-&-ride) and ridership statistics were extracted from the output plan files using PlanSum. These data included ridership by transit route and boarding and alighting data by transit stop by time of day and trip group.

These results were compared to the two observed ridership statistics METRO used to validate the 1994 regional model shown in Table 4:

1. total average weekday originating riders (linked trips), and
2. light rail (LRT) boardings (unlinked trips).

Table 4: Transit Validation Criteria

Boardings	*Linked*	*Unlinked*	*Transfer Ratio*
Tri-Met + Clark County	162,900	214,900	1.32
LRT Ridership	20,600	25,800	1.25

In the initial run, the LRT boardings were about 3,000, which is very low compared to the observed number. Several light rail stations were found to have no boardings due to incorrect access restrictions. The end-to-end running times for the LRT were found to be up to 16 percent higher than the published schedule, especially during the off-peak periods. After fixing these coding problems the light rail boardings increased to 5,260, still far from the observed target. The overall transfer ratio for this run (1.31), however, was very close to the observed value of 1.32.

To improve the "attractiveness" of LRT in the routing algorithm, a bias factor was applied to the rail travel time to encourage travelers to choose rail over competing bus services. The first adjustment was based on the rail bias factor suggested by FTA New Starts guidelines. The 0.80 rail bias factor increased the LRT boardings to 7,811 or 30 percent of the target ridership. The bias factor was then further reduced until the LRT boardings reached the observed boardings. As expected, the overall transfer ratio increased as more travelers transferred to LRT than before. To compensate, the transfer penalty was increased to 4 minutes. The variation of LRT boardings, total transit boardings and the system wide transfer ratio with rail bias factor and transfer penalty changes is shown in Table 5.

Table 5: Rail Bias Factor Calibration

Bias Factor	*Transfer Penalty (min)*	*LRT Boardings*	*Total Transit Boardings*	*Overall Transfer Ratio*
1.00	3	5,257	202,141	1.31
0.80 (FTA)	3	7,811	203,855	1.33
0.50	3	12,272	203,693	1.32
0.25	3	26,108	214,212	1.39
0.40	3	25,489	208,075	1.35
0.40	3.5	22,347	203,646	1.32
0.35	4	22,974	201,709	1.31
0.32	4	25,610	202,989	1.32

The number of all-walk trips in the final run was about 25,750 with an average walk distance of 0.55 miles. There were about 17,000 intra-zonal trips in the METRO transit trip tables that were used as input in the trips conversion process. Most of these trips were routed as all-walk trips. About 5.9 percent of the input trips could not be routed due to various problems. More than 85 percent of those problems were due to the maximum walking distance limitation. These problems could be reduced through smarter allocation of trips to origin and destination activity locations or including additional network detail to make walk paths more direct.

3.2.3 Path Comparisons

In addition to matching ridership targets, the TRANSIMS paths and travel times were compared to the paths and travel times generated by the METRO model to gauge the overall compatibility of the two methods. Fourteen activity locations were picked at representative locations around the region and paths were built to a downtown location during the evening peak period. The traffic analysis zones corresponding to the selected activity locations were identified and paths between these zones were built using the METRO model. Figure 14 shows the origin and destination locations along with the paths built by the Router for each of the origins. A summary of the comparison results is provided in Table 6.

Figure 14: Router Transit Path Tests

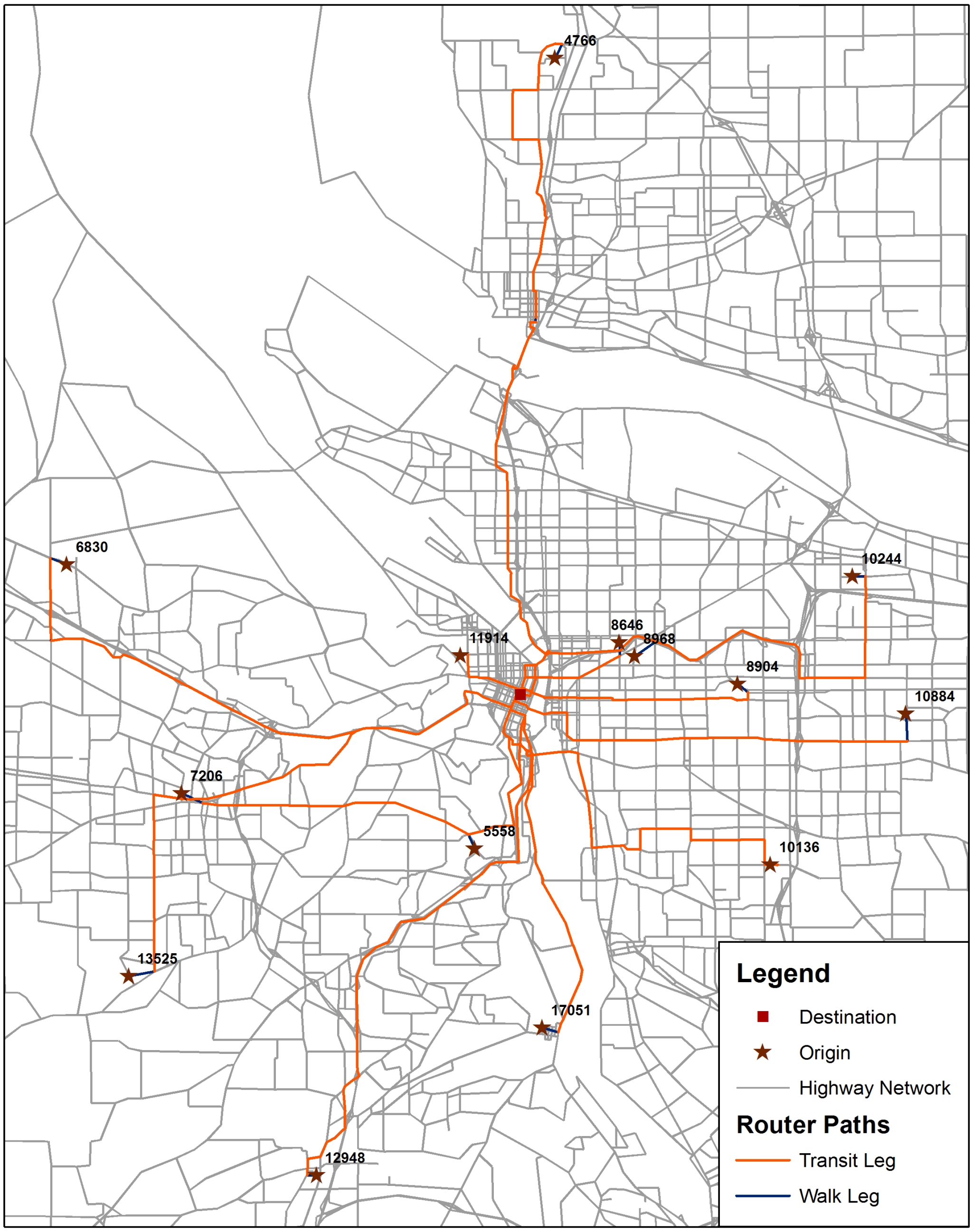


Table 6: METRO vs. TRANSIMS Path Comparisons

Origin – Activity Location (TAZ)	METRO		Router		Comments
	Route(s)	Total Time (min)	Route(s)	Total Time (min)	
11,914 (29)	15THUR*	11.3	15THUR	23.5	Same route
7,206 (125)	57HTC*	36.3	54B	56.9	Different route between same O-D
13,525 (150)	62MURR, 56S*, 57HTC*	60.0	62MURR, 57HTC	84.0	Match. Better Router path.
6,830 (205)	52O, 88K*	48.2	52O, 88K	54.1	Match
17,051 (327)	35CAN	34.3	36TCBD	36.3	Different route along same path
12,948 (363)	96WILS	36.0	96WILS	40.6	Match
10,244 (607)	71T122*, 01BANF	44.4	71T122, 01BANF	52.5	Match
8,646 (715)	09B	23.3	12S	20.1	Different route between same O-D
10,884 (754)	27M, 01BANF	54.4	04D	52.3	Different route; better Router path
8,968 (766)	12S	20.1	01BANF	37.6	High rail bias factor in the Router
8,904 (774)	15MTAB*	26.7	15MTAB	34.8	Match
10,136 (844)	19W	42.3	19W	40.1	Match
5,558 (945)	56S*	23.1	02VCBD	25.9	Different route between same O-D
4,766 (1101)	C006, C134X	55.6	C006, 05IVAN	86.5	Half Match; second route between same O-D

Note: '' indicates that it was combined with other routes. ',' indicates a transfer.*

Table 6 shows that most of the paths matched for the selected O-D pairs. For the O-D pairs where the paths did not match, the Router either took a different transit route along the same corridor or a better path. In general, the Router travel times are higher. Since only the light rail (MAX) run times were validated in the TRANSIMS transit network, differences in total time between METRO and TRANSIMS were expected. Another source of difference is the way wait times are calculated by METRO and TRANSIMS. Wait time in METRO is half of the combined headway of all the transit routes that serve a given boarding and alighting pair, while the wait time in TRANSIMS is the actual time the traveler waits based on the transit schedule.

3.3 Validation Results

The Tri-Met observed boardings from the 1994 on-board survey were not used in METRO's 1994 model calibration due to a number of data processing and compatibility problems. These problems were subsequently addressed and the data were made available for use in this study. Since the METRO validation focused only on the LRT boardings and overall originating transit trips, the total daily transit boardings generated by the METRO model were much higher than the observed boardings reported by Tri-Met. TRANSIMS, on the other hand, replicated the Tri-Met line ridership relatively well.

In the mean time, transit calibration work performed in other regions discovered that the TRANSIMS Router generates fewer illogical paths if the out-of-vehicle impedance parameters are nearly equal to the in-vehicle impedance parameters. Conventional wisdom and FTA guidelines suggest out-of-vehicle time should be valued between 2.0 to 2.5 times greater than in-vehicle time. In TRANSIMS this relationship can cause a traveler to board a bus traveling in the opposite direction, get off, and board the bus headed in the desired direction in order to avoid accumulating wait time impedance. Replacing the out-of-vehicle impedance values with a bus stop boarding penalty and a minimum wait time constraint has proved effective in controlling transfers without dampening rail ridership.

The Portland Router parameters were recalibrated using this approach. The final parameters are shown in Table 7. Note that the value of walking is reduced to 1.5 times in-vehicle time and the wait time values are both 1.0 times in-vehicle time. The general transfer penalty is four minutes and the rail bias factor is a much more acceptable 0.75. Two new parameters were introduced since the original calibration. These are the six minute penalty for each boarding at a bus stop and the minimum wait time of 30 seconds. These values help to account for the environmental impacts of waiting by the side of the road and the general variability of transit schedules.

Table 7: Calibrated Parameters for Transit Path Building

Router Control Key	*Value*	*Parameter Description*
WALK_TIME_VALUE	15	Walk time impedance - 1.5 * IVT (seconds)
VEHICLE_TIME_VALUE	10	In-vehicle time (IVT) impedance (seconds)
FIRST_WAIT_VALUE	10	First wait time impedance - 1.0 * IVT (seconds)
TRANSFER_WAIT_VALUE	10	Transfer wait time impedance - 1.0*IVT (seconds)
COST_VALUE	48	Transit fare and parking cost impedance (cents) - $7.50/hour
TRANSFER_PENALTY	2,400	Transfer impedance penalty - 4 minutes of IVT
STOP_WAITING_PENALTY	3,600	Impedance penalty for boarding at a bus stop - 6 minutes
RAIL_BIAS_FACTOR	0.75	Rail travel time bias factor - 0.75
MAX_WALK_DISTANCE	4,000	Maximum cumulative walk distance - 2.5 miles
MAX_WAIT_TIME	60	Maximum waiting time at a given transit stop - 60 minutes
MIN_WAIT_TIME	30	Minimum waiting time at each transit stop - 30 seconds
MAX_NUMBER_OF_TRANSFERS	3	Maximum number of transfers on a transit trip – 3
MAX_PARK_RIDE_PERCENTAGE	50	Maximum percentage of total trip distance by auto access

Table 8 shows the TRANSIMS transit ridership compare with the observed ridership counts and METRO model loadings on a route-by-route basis. The total TRANSIMS ridership is within 1.5 percent of the observed counts and 64 percent of the routes have ridership estimates that are better than the METRO estimates. The percent root mean squared error for the TRANSIMS routes is a very respectable 36 percent compared to the 158 percent for the METRO estimates.

Table 8: Transit Ridership by Route

Route Name	Observed Weekday Riders	METRO – Observed			TRANSIMS - Observed		
		METRO	Difference	% Diff	TRANSIMS	Difference	% Diff
01BANF (MAX)	25,800	25,883	83	0.3%	24,854	(946)	-3.7%
02G/02VCD	3,228	6,554	3,326	103.0%	4,762	1,534	47.5%
04D/04F	12,043	43,720	31,677	263.0%	8,839	(3,204)	-26.6%
05IVAN & 05JTZB	6,183	14,447	8,264	133.7%	6,234	51	0.8%
06MLKB	3,882	3,471	(411)	-10.6%	2,812	(1,070)	-27.6%
08D15/08H15	7,996	10,192	2,196	27.5%	6,161	(1,835)	-22.9%
09B/09PGL	8,901	5,965	(2,936)	-33.0%	9,699	798	9.0%
10H/10T	3,956	3,975	19	0.5%	5,033	1,077	27.2%
12BKC & 12BKCL	9,653	11,700	2,047	21.2%	10,509	856	8.9%
14H & 14HX	8,336	162	(8,174)	-98.1%	5,296	(3,040)	-36.5%
15S/15MTAB	6,509	13,558	7,049	108.3%	7,022	513	7.9%
17SLIN & 17SMPK	7,807	11,457	3,650	46.8%	7,048	(759)	-9.7%
18HILL	172	682	510	296.5%	72	(100)	-58.1%
19G/19W	6,805	5,499	(1,306)	-19.2%	8,605	1,800	26.5%
20BBTC/201ARP	5,109	7,002	1,893	37.1%	7,512	2,403	47.0%
22ROSE	715	2,532	1,817	254.1%	508	(207)	-29.0%
23S182 & 23S223	588	19,171	18,583	3160.4%	862	274	46.6%
24H	1,351	1,929	578	42.8%	1,330	(21)	-1.6%
25G	248	456	208	83.9%	214	(34)	-13.7%
26S	2,146	2,164	18	0.8%	1,551	(595)	-27.7%
27M	296	1,727	1,431	483.4%	444	148	50.0%
28LMTC	379	2,582	2,203	581.3%	350	(29)	-7.7%
29LCBD	504	2,341	1,837	364.5%	437	(67)	-13.3%
31CTC	2,304	2,275	(29)	-1.3%	3,348	1,044	45.3%
32OCCC	1,418	2,393	975	68.8%	1,095	(323)	-22.8%
33CCC	3,272	2,533	(739)	-22.6%	2,809	(463)	-14.2%
34RCBD	449	1,697	1,248	278.0%	377	(72)	-16.0%
35CAN	2,152	2,338	186	8.6%	2,586	434	20.2%
37NSHR	162	3,848	3,686	2275.3%	260	98	60.5%
39L	439	2,382	1,943	442.6%	360	(79)	-18.0%

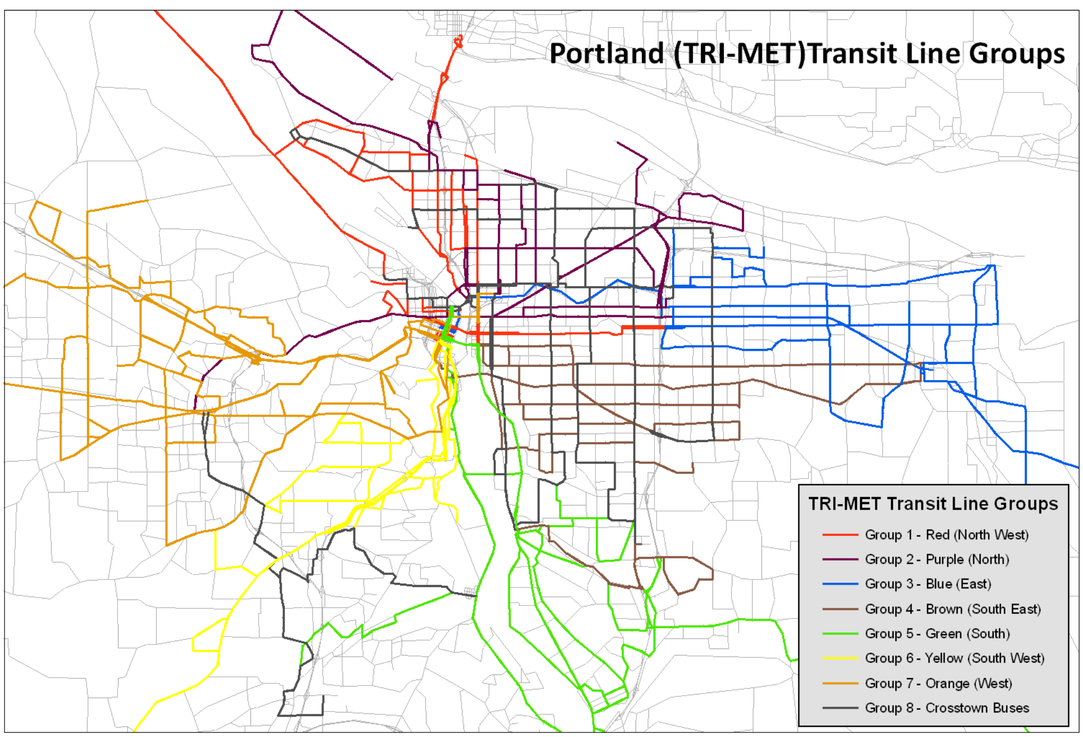
Route Name	Observed Weekday Riders	METRO – Observed			TRANSIMS - Observed		
		METRO	Difference	% Diff	TRANSIMS	Difference	% Diff
40M/TAC	3,772	3,246	(526)	-13.9%	4,747	975	25.8%
41CHWY/41F92	6,585	4,226	(2,359)	-35.8%	7,510	925	14.0%
43TFTV	982	1,289	307	31.3%	1,151	169	17.2%
45G & 45GX	1,395	0	(1,395)	-100.0%	2,002	607	43.5%
51CCPL, 51CDSH	796	0	(796)	-100.0%	619	(177)	-22.2%
52O	1,518	3,379	1,861	122.6%	1,287	(231)	-15.2%
54B	2,472	2,537	65	2.6%	2,253	(219)	-8.9%
55HAML	474	1,488	1,014	213.9%	694	220	46.4%
56S	2,070	1,346	(724)	-35.0%	2,676	606	29.3%
57FFGV/57HTC	7,386	3,049	(4,337)	-58.7%	7,232	(154)	-2.1%
59CBTC	1,638	2,498	860	52.5%	2,468	830	50.7%
60CM	121	393	272	224.8%	288	167	138.0%
62MURR	680	727	47	6.9%	1,105	425	62.5%
67B	1,212	731	(481)	-39.7%	703	(509)	-42.0%
68CORN & 68 HLTN	74	0	(74)	-100.0%	29	(45)	-60.8%
70T13, 70T17	3,020	0	(3,020)	-100.0%	2,088	(932)	-30.9%
71T122	6,368	3,371	(2,997)	-47.1%	6,873	505	7.9%
72K82	12,803	4,675	(8,128)	-63.5%	6,919	(5,884)	-46.0%
75TMTC	9,965	3,806	(6,159)	-61.8%	7,544	(2,421)	-24.3%
76TGTU	593	573	(20)	-3.4%	913	320	54.0%
77NGTC	4,261	2,083	(2,178)	-51.1%	4,295	34	0.8%
78LOBV	2,435	1,421	(1,014)	-41.6%	2,710	275	11.3%
79CAN	586	1,069	483	82.4%	287	(299)	-51.0%
80TTRT	351	1,007	656	186.9%	455	104	29.6%
81G257	121	420	299	247.1%	120	(1)	-0.8%
84SAN & 84 SANX	128	655	527	411.7%	161	33	25.8%
88K & 88CORN	1,394	0	(1,394)	-100.0%	1,904	510	36.6%
89R	982	1,099	117	11.9%	2,591	1,609	163.8%
91X	892	1,192	300	33.6%	1,629	737	82.6%
95X	183	219	36	19.7%	442	259	141.5%
99X	378	310	(68)	-18.0%	702	324	85.7%
Total	**208,438**	**261,444**	**53,006**	**25.4%**	**205,386**	**(3,052)**	**-1.5%**
% Root Mean Squared Error			**158.1%**			**35.8%**	

Table 9 shows how the results compare on a line group basis. The ridership estimates for the eight Tri-Met line groups shown in Figure 15 are very respectable. Almost all of the TRANSIMS line group results match the ridership counts better than the METRO model. The greater level of aggregation also results in improvements in the root mean squared error for both the METRO and TRANSIMS estimates. The 12.5 percent RMSE for TRANSIMS is outstanding.

Table 9: Transit Validation Summary by Line Group

Line Group Name	*Observed Weekday Riders*	*METRO - Observed*			*TRANSIMS - Observed*		
		METRO	*Difference*	*% Diff*	*TRANSIMS*	*Difference*	*% Diff*
Group 1 - Red (NorthWest)	28,764	43,225	14,461	50.3%	28,271	(493)	-1.7%
Group 2 - Purple (North)	24,965	28,145	3,180	12.7%	26,506	1,541	6.2%
Group 3 - Blue (East)	31,623	55,524	23,901	75.6%	30,499	(1,124)	-3.6%
Group 4 - Brown (SouthEast)	30,786	56,418	25,632	83.3%	31,033	247	0.8%
Group 5 - Green (South)	13,842	28,302	14,460	104.5%	14,550	708	5.1%
Group 6 - Yellow (SouthWest)	17,304	20,086	2,782	16.1%	17,707	403	2.3%
Group 7 - Orange (West)	21,709	22,108	399	1.8%	25,478	3,769	17.4%
Group 8 - Crosstown	39,445	15,929	(23,516)	-59.6%	31,342	(8,103)	-20.5%
Total	**208,438**	**269,737**	**61,299**	**29.4%**	**205,386**	**(3,052)**	**-1.5%**
% Root Mean Squared Error			**63.9%**			**12.5%**	

Figure 15: Transit Line Groups



Chapter 4 Population Synthesizer

This chapter documents the steps taken to synthesize a population for Portland for the year 1996 using the 2000 U.S. Census data and Portland METRO's 1996 household data at the block group level – the smallest geographic unit for which relevant Census data are available. The activity-based model generates daily activities for each individual living in the region based on the household and person attributes of the synthetic population.

TRANSIMS Version 4.0 added a suite of tools for manipulating and extracting 2000 U.S. Census data for input into PopSyn (the population synthesizer program). These tools, namely PUMSPrep, SF3Prep, ZoneData, and LocationData, streamline the preparation of input data to PopSyn. The PopSyn program was also re-written to remove the constraints and limitation from the original software and run on both Windows and Linux operating systems.

The population synthesizer imposes the correlation structure from known joint-distributions of household attributes available at a regional (aggregate) level onto smaller (disaggregate) geographies or zones. The marginal distributions (or totals) of various household attributes are matched for each zone. In other words, the households assigned to each zone are distributed across demographic attributes in such a way that the correlation structure of the resulting joint-distribution is consistent with the correlation structure of the region. The regional joint-distribution acts as a seed matrix that is modified to fit the marginal distributions using a statistical method known as iterative proportional fitting (IPF).

4.1 Basic Population Synthesis Algorithm

The population synthesizer uses the 2-step IPF algorithm proposed by Beckman et al. in 1996[1]. The algorithm requires two sets of household data: 1) regional household sample data which contains full detailed information regarding the demographic attributes of a sample of representative households in a given region and 2) zonal household summary data which contains household totals, summarized by select demographic categories – i.e. the total number of households that belong to each demographic category.

These two sets of data are then assembled to produce two types of household tables (or distributions). The zone-level summary data represents marginal distributions (i.e., zone control totals) for the cross-classification tables extracted from the regional joint distribution. The 2-step IPF algorithm then uses the regional cross-classification table as a seed to estimate the zonal cross-classification. This is done by imposing the correlation structure found in the regional cross-classification table while maintaining the household totals found in the corresponding zone-level summary tables. The zonal cross-classification table is used to randomly assign complete household records from the regional household sample data to each synthetic household.

[1] Beckman, R.J., K.A. Baggerly, and M.D. McKay. Creating Synthetic Baseline Populations. Transportation Research Part A: Policy and Practice, 30(6), pp. 415-429, 1996.

In the U.S. Census data, regional sample data can be found in the Public Use Microdata Sample (PUMS) – a 5 percent sample of households with complete records contained in a collection of Census Tracts or smaller geographies called Public Use Microdata Areas (PUMAs). PUMAs are geographic areas containing approximately 100,000 households, so PUMS will contain complete information for approximately 5,000 households in each PUMA.

The household summary data for the full Census population are provided in Summary Files 3A (SF3). These data files contain demographic summaries for various geography levels such as counties, Census Tracts, Census block groups and blocks. Since privacy rules suppress much of the data at the block level, block groups represent the finest level of geographic detail for which reliable data are available.

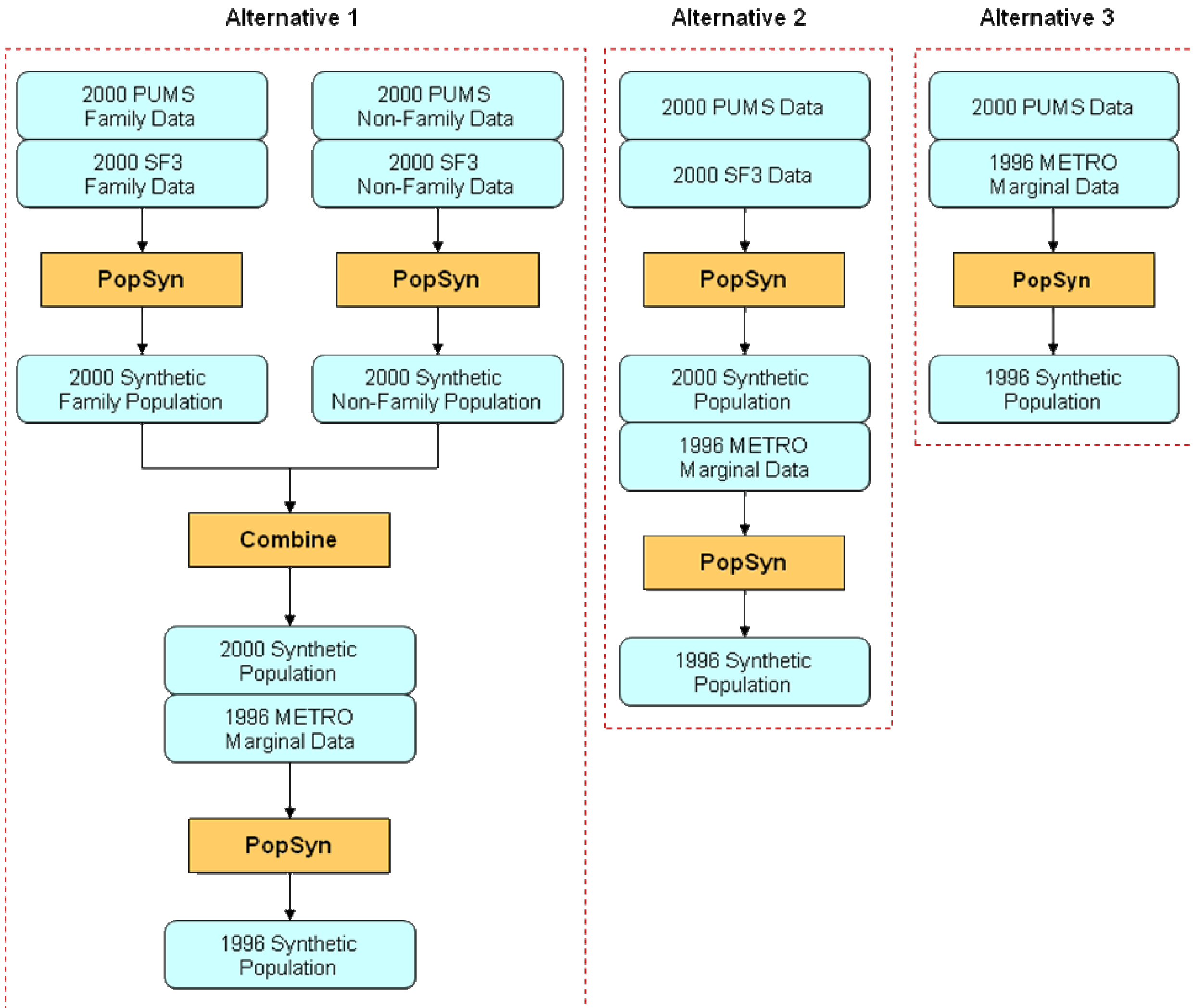
4.2 Synthesizing a 1996 Population for Portland

The calibration year for the Portland GEN2 model is 1996. Given that U.S. Census data are available every 10 years, with the last census being conducted in year 2000, household estimate for 1996 at the block group level are required to synthesize the 1996 population. METRO developed block group household data for the first phase of the Portland Study to forecast 1996 population from 1990 Census data. This application will use the same household estimates to backcast from the 2000 Census to 1996. Since both phases use the same 1996 marginals, differences in the results are attributed to changes in the household correlation structure between the 1990 and 2000 Census. Comparing the two estimates demonstrates the likely stability of the population synthesis process for future forecasting applications.

There are three alternative ways to generate the 1996 synthetic population using the 2000 Census data. A high level depiction of these alternatives is shown in Figure 16. Alternatives 1 and 2 start by estimating a baseline population for the year 2000, and then use that population as a seed along with the 1996 METRO marginal data to back-cast to 1996. These alternatives differ in how the baseline population is obtained. Unlike the U.S. Census data, the 1996 marginals are not differentiated by family and non-family household types. Therefore, to be able to use the 1996 marginals, the baseline population must reflect the true composition of households. This can be done by either using separate models for family and non-family households (Alternative 1) or by using the PUMS household weights (Alternative 2). Alternative 3, on the other hand, skips the baseline population and uses the PUMS household weights to directly estimate the 1996 population using the 2000 PUMS and 1996 marginals.

All alternatives assume that the correlation structure in the 2000 PUMS is representative of field conditions in 1996. In other words, they all assume that correlation structure changes from 1996 to 2000 are negligible. Alternatives 1 and 2, however, will be of a superior spatial fidelity than Alternative 3 because the population seed for the former two (2000 baseline population) is available at the block group level whereas the 2000 PUMS is only available at the PUMA level. Tests performed using all three methods showed only minimal differences between the three methods. Alternative 2 was selected for the GEN2 model under the assumption that expanding the population first provided slightly higher fidelity for future expansion factors than the method that simply depended on PUMS household weights.

Figure 16: Alternative Methods of Synthesizing a 1996 Population

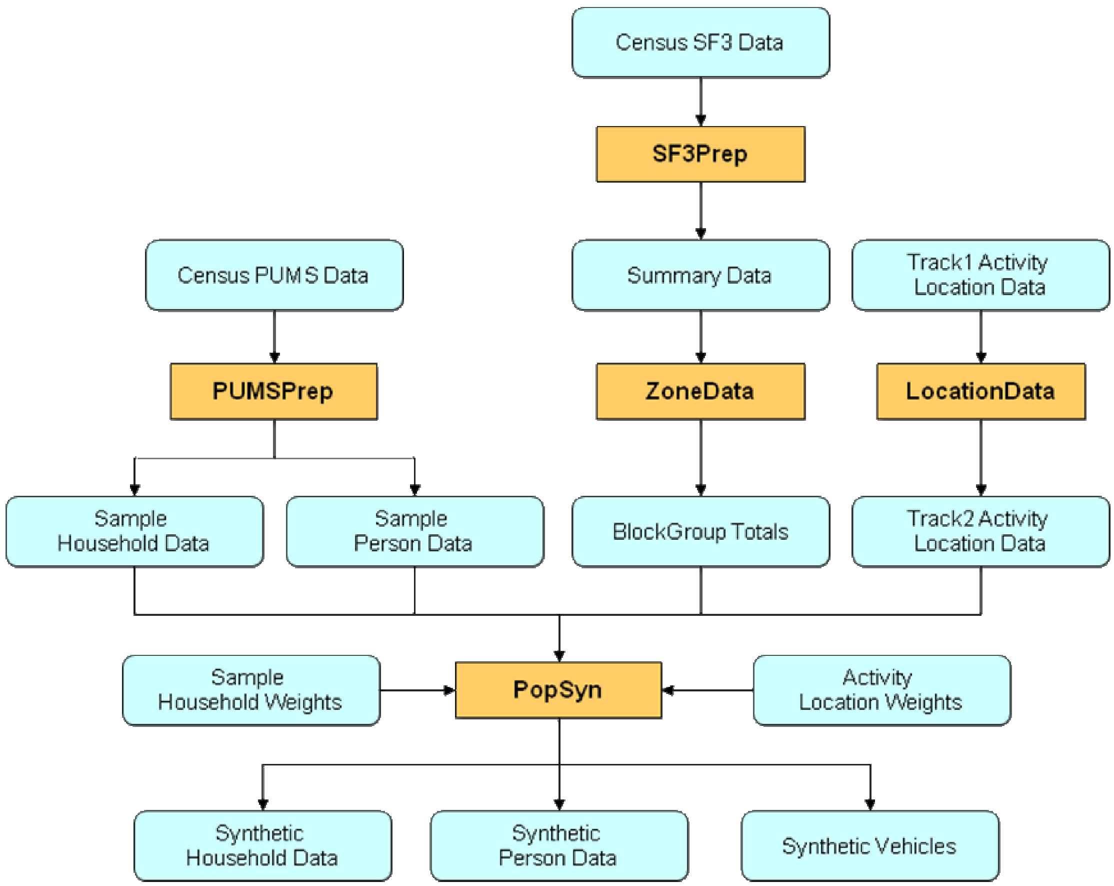


4.3 Population Synthesis Process

The data processing steps used to generate synthetic households for the Portland GEN2 model are outlined in Figure 17. The process starts by downloading PUMS and Summary File 3A (SF3) data from the U.S. Census for the area of interest. The household and person attributes needed for population synthesis, activity generation, and mode choice are identified and extracted from these datasets and stored in TRANSIMS file formats. Activity locations are mapped to block groups and PUMAs in order to link the datasets. The household-related sample records, block group summary or marginal distributions, and activity location weights are used by the PopSyn program to generate the synthetic population.

Given a list of attributes and their corresponding classification system, summary tables are assembled from SF3 files for each attribute at the block group level. Summary tables represent the marginal distribution of households with respect to demographic attributes, hence the terms marginals, marginal totals, summary totals, and summaries all mean the same and will be used interchangeably in this report.

Figure 17: Population Synthesis Process



A representative cross-classification table of the selected control attributes is then extracted from PUMS data at the PUMA level. A two-step IPF process is used to estimate the cross-classification tables first at the PUMA level, and then at the block group level. The two-step IPF process ensures that household totals are preserved for each block group while the correlation structures are equal across all estimated block group cross-classification tables. Moreover, the aggregation of the estimated block group cross-classification tables within a PUMA will match the correlation structure within the PUMA sample cross-classification table.

After obtaining the estimated cross-classification table at the block group level, a PUMS household with the corresponding control attributes is randomly selected from the PUMA data and assigned to the synthetic household. This copies all of the controlled and uncontrolled attributes from the sample household to the synthetic household. Finally each synthetic household is randomly assigned to an activity location within the block group based on location weights and the household vehicles are assigned to a parking lot attached to the household location.

The following steps describe the process in more detail.

Step 1: Download the PUMS Data

PUMS data and technical documentation were downloaded free of charge from the Census website:

http://www2.census.gov/census_2000/datasets/PUMS/FivePercent

Since the Portland modeling region includes parts of Oregon and Washington, PUMS data for these two states were downloaded. The Oregon subdirectory includes a file called PUMS5_41.TXT and the Washington subdirectory includes a file called PUMS5_53.TXT. The PUMS user's manual was downloaded from:

http://www.census.gov/prod/cen2000/doc/pums.pdf

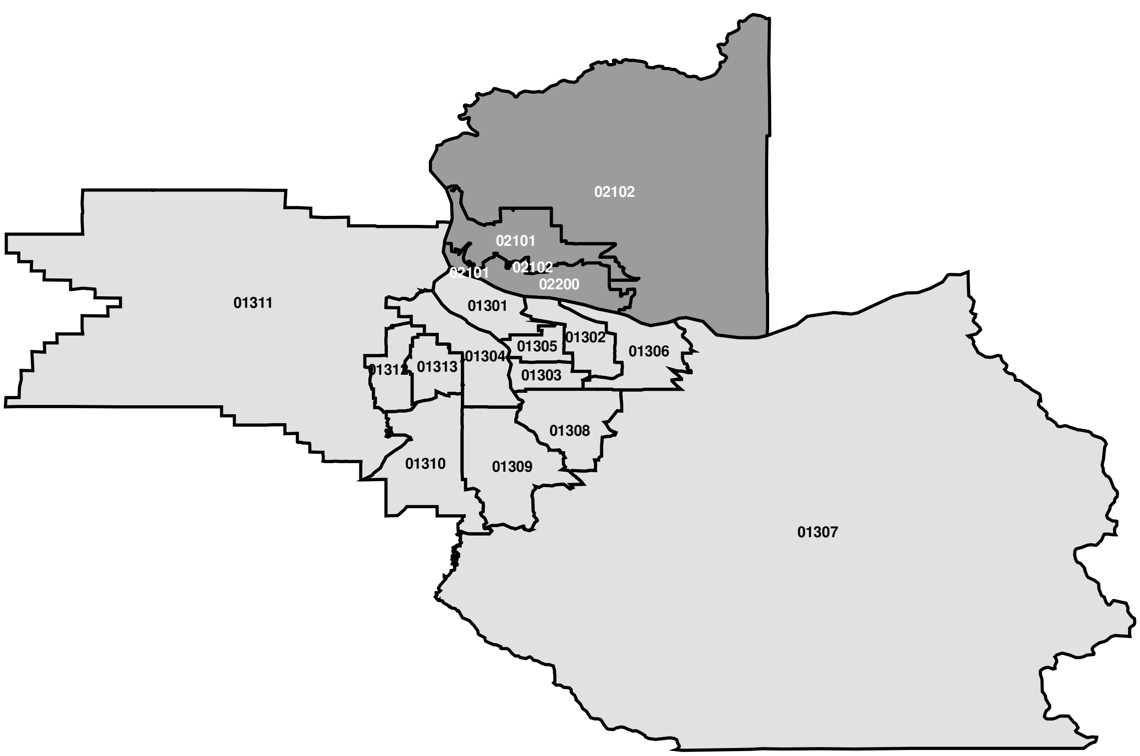
This document was used to determine the appropriate data items to extract from the dataset. The PUMAs were mapped to block groups and activity locations using the ArcView shapefiles provided by the Census at:

http://www.census.gov/geo/www/cob/pu5_2000.html

The compressed files called p541_d00_shp.zip and p553_d00_shp.zip contain the PUMA shapefiles for Oregon and Washington. The coordinate system used by the Census files is GCS_North_American_1983.

The PUMA boundary files were then read into ArcGIS to identify the PUMA numbers for the Portland region. The 13 PUMAs in Oregon (1301 to 1313) and the three PUMAs in Washington (2101, 2102, and 2200) that constitute the Portland region are shown in Figure 18.

Figure 18: Year 2000 PUMAs for the Portland Region



Step 2: Download the Summary File Data

Census Summary File 3 (SF3) provides over 800 summary tables aggregated to 10 levels of geography and stored in 76 data files. Each summary table is identified with a code number and each value in the table is assigned an element number. The table index also has a segment number to identify in which of the 76 data files the table is located. The code and segment numbers for summary tables are found in:

http://www2.census.gov/census_2000/datasets/Summary_File_3/0SF3_table_matrix.doc

The summary tables were downloaded free of charge from the Census website:

http://www2.census.gov/census_2000/datasets/Summary_File_3

Within a given state's subdirectory are the 76 data files corresponding to the 76 segment numbers. The segment 1 and 7 summary tables were downloaded for Oregon (or00001_uf3.zip and or00007_uf3.zip) and Washington (wa00001_uf3.zip and wa00007_uf3.zip). In addition to data tables, the geography files (orgeo_uf3.zip and wageo_uf3.zip) were downloaded in order to relate the record code numbers to the equivalent geography codes and locate the centroid coordinates of each geographic area.

ArcView shapefiles for geographic boundaries are available from the Census website:

http://www.census.gov/geo/www/cob/bdy_files.html

The year 2000 block group boundaries for Oregon (bg41_d00_shp.zip) and Washington (bg53_d00_shp.zip) were downloaded and unzipped.

Step 3: Identify Demographic Attributes

The next step is to identify the demographic attributes needed to synthesize the population, generate activities, and choose travel modes. These attributes will be extracted from the Census datasets for use by TRANSIMS. Household control demographics refer to the attributes that constrain the IPF process within the population synthesizer. For the GEN2 model, the control demographics defined for METRO's 1996 household estimates were used. These include:

- Household size
- Household income
- Age of head of household

Uncontrolled demographics are attributes that are kept when a household is synthetically drawn from PUMS. The uncontrolled demographics for this model include:

- Relationship status (head of household, spouse, child, etc…)
- Gender
- Age
- Employment status
- Number of workers
- Number of vehicles

- Number of children less than 5 years
- Number of children between 5 and 15 years
- Number of adults between 26 and 45 years

The number of workers and the number of vehicles are required for validating the synthetic population against METRO's independent estimates. The gender, age, employment status, and the last three demographics are included because they are needed for the household matching step in the activity generator. These attributes were selected due to their significance in describing trip making behavior in Portland.

Step 4: Extract PUMS Records for Project Area PUMAs

The PUMS file contains two types of records: household records and person records. The household records start with an "H" identifier whereas the person records start with a "P" identifier. Both records are linked based on SERIALNO. Table 10 lists the household and person attributes extracted from PUMS to populate the controlled and uncontrolled demographics identified in Step 3.

Table 10: PUMS Household and Person Attributes

Attribute Name	*PUMS Name*	*Begin Column*	*End Column*
Household Attributes			
HHOLD	SERIALNO	2	8
STATE	STATE	10	11
PUMA	PUMA5	14	18
WEIGHT	HWEIGHT	102	105
PERSONS	PERSONS	106	107
VEHICLE	VEHICL	134	134
TYPE	HHT	213	213
WORKERS		247	247
INCOME	HINC	251	258
Person Attributes			
HHOLD	SERIALNO	2	8
PERSON	PNUM	9	10
RELATE	RELATE	17	18
GENDER	SEX	23	23
AGE	AGE	25	26
WORKER	WRKLYR	236	236

The PUMSPrep program extracts the relevant household and population data from PUMS data files. The process also creates five additional household fields needed by the activity generator, namely:

1. WORKERS (number of workers in the household),
2. HHAGE (age of head of household),
3. NUM_LT5 (number of children less than 5 years old),
4. NUM_5TO15 (number of children between the ages of 5 and 15 years old), and

5. NUM_26TO45 (number of adults between the ages of 26 and 45 years old).

The WORKERS field is derived by totaling the Worker field of household members in the newly generated PUMS population file. The remaining fields are derived through logical tests using the AGE field of household members. This is accomplished via a script that is read and compiled within PUMSPrep. The HHAGE field is derived in a similar way for use as one of the control attributes for synthesizing the population. The remaining fields will be needed for the household matching step in the activity generation process.

Since the Portland region extends into Vancouver, Washington, the PUMS and SF3 data for Oregon and Washington must be used. While the population from each state may be synthesized separately and then merged to form the final synthetic population, PopSyn can also process multiple states in the same application. Unfortunately, PUMS household numbers are not unique across states so a unique household ID must be created. PUMS household ID and the state ID were concatenated to produce the unique ID.

Figure 19 shows the PUMSPrep control file used to extract the attributes shown in Table 10 and Figure 20 shows the data processing script.

Figure 19: PUMSPrep Control Keys

```
#---- Input ----

PUMS_DATA_FILE_1              ../../PUMS/PUMS5_41.TXT          #---- Oregon
PUMS_DATA_FILE_2              ../../PUMS/PUMS5_53.TXT          #---- Washington
CONVERSION_SCRIPT             ../../Scripts/PUMS_Script.txt

#---- Output ----

NEW_PUMS_HOUSEHOLD_FILE       ../../PUMS/2000.PUMS.Household.txt
NEW_PUMS_POPULATION_FILE      ../../PUMS/2000.PUMS.Population.txt

#---- Parameters ----

HOUSEHOLD_DATA_FIELD_1        HHOLD, INTEGER, 0, 10
HOUSEHOLD_DATA_FIELD_2        STATE, INTEGER, 10, 2
HOUSEHOLD_DATA_FIELD_3        PUMA, INTEGER, 14, 5
HOUSEHOLD_DATA_FIELD_4        WEIGHT, INTEGER, 102, 4
HOUSEHOLD_DATA_FIELD_5        PERSONS, INTEGER, 218, 2
HOUSEHOLD_DATA_FIELD_6        VEHICLES, INTEGER, 134, 1
HOUSEHOLD_DATA_FIELD_7        INCOME, INTEGER, 251, 8
HOUSEHOLD_DATA_FIELD_8        TYPE, INTEGER, 213, 1
HOUSEHOLD_DATA_FIELD_9        WORKERS, INTEGER, 0, 1
HOUSEHOLD_DATA_FIELD_10       HHAGE, INTEGER, 0, 2
HOUSEHOLD_DATA_FIELD_11       NUM_LT5, INTEGER, 0, 2
HOUSEHOLD_DATA_FIELD_12       NUM_5TO15, INTEGER, 0, 2
HOUSEHOLD_DATA_FIELD_13       NUM_26TO45, INTEGER, 0, 2
HOUSEHOLD_DATA_FIELD_14       PUMS_HHOLD, INTEGER, 2, 7

POPULATION_DATA_FIELD_1       HHOLD, INTEGER, 0, 10
POPULATION_DATA_FIELD_2       PERSON, INTEGER, 9, 2
POPULATION_DATA_FIELD_3       RELATE, INTEGER, 17, 2
POPULATION_DATA_FIELD_4       GENDER, INTEGER, 23, 1
POPULATION_DATA_FIELD_5       AGE, INTEGER, 25, 2
```

```
POPULATION_DATA_FIELD_6         WORKER, INTEGER, 236, 1
POPULATION_DATA_FIELD_7         HHOLD, INTEGER, 2, 7

STATE_PUMA_LIST                 4101301, 4101302, 4101303, 4101304, 4101305, 4101306,
                                4101307, 4101308, 4101309, 4101310, 4101311, 4101312,
                                4101313, 5302101, 5302102, 5302200
```

Figure 20: PUMS Conversion Script

```
New_HH.HHOLD = New_HH.PUMS_HHOLD * 100 + New_HH.STATE
New_Pop.HHOLD = New_HH.HHOLD
IF (New_Pop.WORKER == 1) THEN
    New_HH.WORKERS = New_HH.WORKERS + 1
ENDIF
IF (New_Pop.RELATE == 1) THEN
    New_HH.HHAGE = New_Pop.AGE
ENDIF
IF (New_Pop.AGE < 5) THEN
    New_HH.NUM_LT5 = New_HH.NUM_LT5 + 1
ELSE
  IF (New_Pop.AGE <= 15) THEN
     New_HH.NUM_5TO15 = New_HH.NUM_5TO15 + 1
  ENDIF
ENDIF
IF (New_Pop.AGE <= 45  AND New_Pop.AGE >= 26) THEN
    New_HH.NUM_26TO45 = New_HH.NUM_26TO45 + 1
ENDIF
```

Step 5: Extract Base Year Control Totals for Project Area Block Groups

Base year control totals were extracted from the 2000 Census Summary File 3 (SF3) database. SF3 contains 329 housing tables and 484 population tables compiled from a sample of approximately 19 million housing units (about 1 in 6 households) that received the Census 2000 long-form questionnaire. The data are aggregated to the ten geographic levels shown in Table 11. For privacy reasons, most tables do not provide data below the block group level.

Table 11: Geographic Levels in the 2000 Census

Geography Level	*2000 Census Summary Level*
United States	10
Regions	20
Divisions	30
State	40
County	50
County Subdivision	60
Place	70
Census Tract	80
Block Group	90
Blocks	100

SF3 tables contain the total number of households in a given geography level and their distribution across demographic categories or classification. The data are differentiated for three types of households: 1) family households, 2) non-family households, and 3) group quarters. Family households are households with two or more related members, non-family households are households with unrelated persons (or living alone), and group quarters are dwellings such as dorms or prisons.

The GEN2 model uses the following demographics to control the population synthesis process at the block group level (summary level 90), namely:

- Age of householder (SF3 Segment 01, Table P13),
- Household size (SF3 Segment 01, Table P14),
- Household family income (SF3 Segment 07, Table P76), and
- Household non-family income (SF3 Segment 07, Table P79).

The SF3 classification system for these attributes is shown below:

Figure 21: 2000 SF3 Classification System for Table P13

```
P13.      HOUSEHOLD TYPE BY AGE OF HOUSEHOLDER [19]
          Universe:  Households
          Total:                                                  P013001   01 9
                Family households:                                P013002   01 9
                  Householder 15 to 24 years                      P013003   01 9
                  Householder 25 to 34 years                      P013004   01 9
                  Householder 35 to 44 years                      P013005   01 9
                  Householder 45 to 54 years                      P013006   01 9
                  Householder 55 to 64 years                      P013007   01 9
                  Householder 65 to 74 years                      P013008   01 9
                  Householder 75 to 84 years                      P013009   01 9
                  Householder 85 years and over                   P013010   01 9
                Nonfamily households:                             P013011   01 9
                  Householder 15 to 24 years                      P013012   01 9
                  Householder 25 to 34 years                      P013013   01 9
                  Householder 35 to 44 years                      P013014   01 9
                  Householder 45 to 54 years                      P013015   01 9
                  Householder 55 to 64 years                      P013016   01 9
                  Householder 65 to 74 years                      P013017   01 9
                  Householder 75 to 84 years                      P013018   01 9
                  Householder 85 years and over                   P013019   01 9
```

Figure 22: 2000 SF3 Classification System for Table P14

```
P14.      HOUSEHOLD TYPE BY HOUSEHOLD SIZE [16]
          Universe:  Households
          Total:                                                  P014001   01 9
                Family households:                                P014002   01 9
                  2-person household                              P014003   01 9
                  3-person household                              P014004   01 9
                  4-person household                              P014005   01 9
                  5-person household                              P014006   01 9
                  6-person household                              P014007   01 9
```

```
            7-or-more-person household                          P014008   01 9
          Nonfamily households:                                 P014009   01 9
            1-person household                                  P014010   01 9
            2-person household                                  P014011   01 9
            3-person household                                  P014012   01 9
            4-person household                                  P014013   01 9
            5-person household                                  P014014   01 9
            6-person household                                  P014015   01 9
            7-or-more-person household                          P014016   01 9
```

Figure 23: 2000 SF3 Classification System for Table P76

```
P76.      FAMILY INCOME IN 1999 [17]
          Universe:  Families
          Total:                                                P076001   07 9
                Less than $10,000                               P076002   07 9
                $10,000 to $14,999                              P076003   07 9
                $15,000 to $19,999                              P076004   07 9
                $20,000 to $24,999                              P076005   07 9
                $25,000 to $29,999                              P076006   07 9
                $30,000 to $34,999                              P076007   07 9
                $35,000 to $39,999                              P076008   07 9
                $40,000 to $44,999                              P076009   07 9
                $45,000 to $49,999                              P076010   07 9
                $50,000 to $59,999                              P076011   07 9
                $60,000 to $74,999                              P076012   07 9
                $75,000 to $99,999                              P076013   07 9
                $100,000 to $124,999                            P076014   07 9
                $125,000 to $149,999                            P076015   07 9
                $150,000 to $199,999                            P076016   07 9
                $200,000 or more                                P076017   07 9
```

Figure 24: 2000 SF3 Classification System for Table P79

```
P79.      NONFAMILY HOUSEHOLD INCOME IN 1999 [17]
          Universe:  Nonfamily households
          Total:                                                P079001   07 9
                Less than $10,000                               P079002   07 9
                $10,000 to $14,999                              P079003   07 9
                $15,000 to $19,999                              P079004   07 9
                $20,000 to $24,999                              P079005   07 9
                $25,000 to $29,999                              P079006   07 9
                $30,000 to $34,999                              P079007   07 9
                $35,000 to $39,999                              P079008   07 9
                $40,000 to $44,999                              P079009   07 9
                $45,000 to $49,999                              P079010   07 9
                $50,000 to $59,999                              P079011   07 9
                $60,000 to $74,999                              P079012   07 9
                $75,000 to $99,999                              P079013   07 9
                $100,000 to $124,999                            P079014   07 9
                $125,000 to $149,999                            P079015   07 9
                $150,000 to $199,999                            P079016   07 9
                $200,000 or more                                P079017   07 9
```

The SF3Prep program extracts the household summaries from SF3 files. Given the SF3 segment files and the field numbers or ranges corresponding to the controlled demographics, SF3Prep extracts the data for the desired summary level and state-PUMA combinations and saves the data into a new zone file to be later used in the population synthesis process. The SF3Prep control file used for this step is shown in Figure 25. The control keys map the field name ranges in the SF3 data to field names in the zone file. For example, P013003..10, F_HH_AGE1..8 copies eight fields from Table P13 to field names in the zone file (F_HH_AGE1, F_HH_AGE2, F_HH_AGE3, F_HH_AGE4, F_HH_AGE5, F_HH_AGE6, F_HH_AGE7, and F_HH_AGE8).

Figure 25: SF3Prep Control Keys

```
#---- Input ----

SF3_GEOGRAPHY_FILE_1          ..\..\SF3\orgeo.uf3
SF3_SEGMENT_FILE_1_1          ..\..\SF3\or00001.uf3
SF3_SEGMENT_FILE_1_7          ..\..\SF3\or00007.uf3
SF3_GEOGRAPHY_FILE_2          ..\..\SF3\wageo.uf3
SF3_SEGMENT_FILE_2_1          ..\..\SF3\wa00001.uf3
SF3_SEGMENT_FILE_2_7          ..\..\SF3\wa00007.uf3
INPUT_COORDINATE_SYSTEM       LATLONG, MILLION_DEGREES

#---- Output ----

NEW_ZONE_DATA_FILE            ..\..\SF3\2000.SF3.Zone_Data.txt
OUTPUT_COORDINATE_SYSTEM      UTM, 10N, METERS

#---- Parameters ----

ZONE_DATA_FIELD_RANGE_1       STATE
ZONE_DATA_FIELD_RANGE_2       COUNTY
ZONE_DATA_FIELD_RANGE_3       TRACT
ZONE_DATA_FIELD_RANGE_4       BLKGRP
ZONE_DATA_FIELD_RANGE_5       P013003..10, F_HH_AGE1..8
ZONE_DATA_FIELD_RANGE_6       P013012..19, NF_HH_AGE1..8
ZONE_DATA_FIELD_RANGE_7       P014003..8,  F_HH_SIZE2..7
ZONE_DATA_FIELD_RANGE_8       P014010..16, NF_HH_SIZE1..7
ZONE_DATA_FIELD_RANGE_9       P076002..17, F_HH_INC1..16
ZONE_DATA_FIELD_RANGE_10      P079002..17, NF_HH_INC1..16

SF3_SUMMARY_LEVEL             90
STATE_PUMA_LIST               4101301, 4101302, 4101303, 4101304, 4101305, 4101306,
                              4101307, 4101308, 4101309, 4101310, 4101311, 4101312,
                              4101313, 5302101, 5302102, 5302200
```

Step 6: Assemble Summary Tables of Household Demographics

The 1996 METRO marginals have a different classification scheme (Table 12) than the SF3 summary tables (see Figure 21, Figure 22, Figure 23 and Figure 24). The extracted control totals, therefore, need to be reclassified to match the METRO data. In addition, SF3 data do not provide a mapping between geography levels and PUMAs, which is needed to link PUMS data to block groups. This step links the block groups to the associated PUMA. Since block groups are not uniquely identified using their Census ID alone. A unique ID is created for each block group based on the following calculated value:

ZONE = STATE*100,000 + INDEX

where ZONE is the zone number associated with each block group, STATE is the state ID as given by the Census, and INDEX is the record number defined in the Census database.

The ZoneData program and a conversion script were used to calculate the unique block group ID and set the PUMA code. Figure 26 shows the ZoneData control file and Figure 27 contains the conversion script. A point-in-polygon search is used to select the ArcView data record used for setting the PUMA and block group indices.

Table 12: METRO 1996 Data Classification System

Category	*Attribute*		
	Household Size	*Income*	*Age of Householder*
1	1	$0 < $10000	15 < 25 years
2	2	$10000 < $15000	25 < 35 years
3	3	$15000 < $25000	35 < 45 years
4	4	$25000 < $35000	45 < 55 years
5	5	$35000 < $50000	55 < 65 years
6	6	$50000 < $100000	65 < 75 years
7	≥ 7	≥ $100000	≥ 75 years

Figure 26: ZoneData Control Keys

```
#---- Input ----

NET_ZONE_TABLE              ..\..\SF3\2000.SF3.Zone_Data.txt
BOUNDARY_POLYGON_1          ..\..\BoundaryFiles\p541_d00.shp
BOUNDARY_POLYGON_2          ..\..\BoundaryFiles\bg41_d00.shp
BOUNDARY_POLYGON_3          ..\..\BoundaryFiles\p553_d00.shp
BOUNDARY_POLYGON_4          ..\..\BoundaryFiles\bg53_d00.shp
CONVERSION_SCRIPT           ..\..\Scripts\BlockGroup_Script.txt
INPUT_COORDINATE_SYSTEM     LATLONG, DEGREES

#---- Output ----

NEW_ZONE_TABLE              ..\..\ZoneData\2000.Zone_Data.txt
OUTPUT_COORDINATE_SYSTEM    UTM, 10N, METERS

#---- Parameters ----

NEW_ZONE_FIELD_1            STATE, INTEGER, 3
NEW_ZONE_FIELD_2            COUNTY, INTEGER, 5
NEW_ZONE_FIELD_3            TRACT, INTEGER, 7
NEW_ZONE_FIELD_4            BLKGRP, INTEGER, 2
NEW_ZONE_FIELD_5            PUMA, INTEGER, 5
NEW_ZONE_FIELD_6            AGE1, INTEGER, 5
NEW_ZONE_FIELD_7            AGE2, INTEGER, 5
NEW_ZONE_FIELD_8            AGE3, INTEGER, 5
NEW_ZONE_FIELD_9            AGE4, INTEGER, 5
NEW_ZONE_FIELD_10           AGE5, INTEGER, 5
```

```
NEW_ZONE_FIELD_11          AGE6, INTEGER, 5
NEW_ZONE_FIELD_12          AGE7, INTEGER, 5
NEW_ZONE_FIELD_14          SIZE1, INTEGER, 5
NEW_ZONE_FIELD_15          SIZE2, INTEGER, 5
NEW_ZONE_FIELD_16          SIZE3, INTEGER, 5
NEW_ZONE_FIELD_17          SIZE4, INTEGER, 5
NEW_ZONE_FIELD_18          SIZE5, INTEGER, 5
NEW_ZONE_FIELD_19          SIZE6, INTEGER, 5
NEW_ZONE_FIELD_20          SIZE7, INTEGER, 5
NEW_ZONE_FIELD_21          INCOME1, INTEGER, 5
NEW_ZONE_FIELD_22          INCOME2, INTEGER, 5
NEW_ZONE_FIELD_23          INCOME3, INTEGER, 5
NEW_ZONE_FIELD_24          INCOME4, INTEGER, 5
NEW_ZONE_FIELD_25          INCOME5, INTEGER, 5
NEW_ZONE_FIELD_26          INCOME6, INTEGER, 5
NEW_ZONE_FIELD_27          INCOME7, INTEGER, 5
NEW_ZONE_FIELD_28          HOUSEHOLDS, INTEGER, 6
```

Figure 27: ZoneData Control Script

```
Out.ZONE = atoi (Polygon2.STATE) * 100000 + Polygon2.INDEX
Out.PUMA = atoi (Polygon1.PUMA5)
IF (Out.PUMA == 0 OR Out.ZONE == 0) THEN
    Out.ZONE = atoi (Polygon4.STATE) * 100000 + Polygon4.INDEX
    Out.PUMA = atoi (Polygon3.PUMA5)
ENDIF

Out.AGE1        = In.F_HH_AGE1 + In.NF_HH_AGE1
Out.AGE2        = In.F_HH_AGE2 + In.NF_HH_AGE2
Out.AGE3        = In.F_HH_AGE3 + In.NF_HH_AGE3
Out.AGE4        = In.F_HH_AGE4 + In.NF_HH_AGE4
Out.AGE5        = In.F_HH_AGE5 + In.NF_HH_AGE5
Out.AGE6        = In.F_HH_AGE6 + In.NF_HH_AGE6
Out.AGE7        = In.F_HH_AGE7 + In.F_HH_AGE8 + In.NF_HH_AGE7 + In.NF_HH_AGE7
Out.SIZE1       = In.NF_HH_SIZE1
Out.SIZE2       = In.F_HH_SIZE2 + In.NF_HH_SIZE2
Out.SIZE3       = In.F_HH_SIZE3 + In.NF_HH_SIZE3
Out.SIZE4       = In.F_HH_SIZE4 + In.NF_HH_SIZE4
Out.SIZE5       = In.F_HH_SIZE5 + In.NF_HH_SIZE5
Out.SIZE6       = In.F_HH_SIZE6 + In.NF_HH_SIZE6
Out.SIZE7       = In.F_HH_SIZE7 + In.NF_HH_SIZE7
Out.INCOME1     = In.F_HH_INC1 + In.NF_HH_INC1
Out.INCOME2     = In.F_HH_INC2 + In.NF_HH_INC2 + In.F_HH_INC3 + In.NF_HH_INC3
Out.INCOME3     = In.F_HH_INC4 + In.F_HH_INC5 + In.NF_HH_INC4 + In.NF_HH_INC5
Out.INCOME4    = In.F_HH_INC6 + In.F_HH_INC7 + In.NF_HH_INC6 + In.NF_HH_INC7
Out.INCOME5    = In.F_HH_INC8 + In.F_HH_INC9 + In.F_HH_INC10 + In.NF_HH_INC8 +
                 In.NF_HH_INC9 + In.NF_HH_INC10
Out.INCOME6    = In.F_HH_INC11 + In.F_HH_INC12 + In.F_HH_INC13 + In.NF_HH_INC11 +
                 In.NF_HH_INC12 + In.NF_HH_INC13
Out.INCOME7    = In.F_HH_INC14 + In.F_HH_INC15 + In.F_HH_INC16 + In.NF_HH_INC14 +
                 In.NF_HH_INC15 + In.NF_HH_INC16
Out.HOUSEHOLDS = Out.SIZE1 + Out.SIZE2 + Out.SIZE3 + Out.SIZE4 + Out.SIZE5 +
                 Out.SIZE6 + Out.SIZE7
```

Step 7: Assign PUMA and Block Group Codes to Each Activity Location

The final step in the data preparation process is to map each activity location to a block group and PUMA. This information is needed to distribute the synthesized households to specific activity locations within the TRANSIMS network. The LocationData program (Figure 28) performs a point-in-polygon search using the coordinates of each activity location and the Census boundary files to identify the PUMA and block group associated with each activity location. The conversion script (Figure 29) then calculates the unique IDs and populates other data fields of general interest.

Figure 28: LocationData Control Keys

```
#---- Input ----

NET_ACTIVITY_LOCATION_TABLE             ..\..\Network\Activity_Location
BOUNDARY_POLYGON_1                      ..\..\BoundaryFiles\p553_d00.shp
BOUNDARY_POLYGON_2                      ..\..\BoundaryFiles\bg53_d00.shp
BOUNDARY_POLYGON_3                      ..\..\BoundaryFiles\p541_d00.shp
BOUNDARY_POLYGON_4                      ..\..\BoundaryFiles\bg41_d00.shp
CONVERSION_SCRIPT                       ..\..\Scripts\LocationData_Script.txt
INPUT_COORDINATE_SYSTEM                 LATLONG, DEGREES

#---- Output ----

NEW_ACTIVITY_LOCATION_TABLE             ..\..\Network\2000.Activity.Location
LOCATIONDATA_REPORT_1                   CONVERSION_SCRIPT
OUTPUT_COORDINATE_SYSTEM                UTM, 10N, METERS

#---- Parameters ----

COPY_EXISTING_FIELDS                    YES
CREATE_NOTES_AND_NAME_FIELDS            YES
NEW_LOCATION_FIELD_1                    STATE, STRING, 2
NEW_LOCATION_FIELD_2                    PUMA, INTEGER, 10
NEW_LOCATION_FIELD_3                    BG_ID, INTEGER, 10
NEW_LOCATION_FIELD_4                    TRACT, INTEGER, 10
NEW_LOCATION_FIELD_5                    BLKGRP, INTEGER, 10
```

Figure 29: LocationData Conversion Script

```
IF (In.ID < 20839)  THEN
    Out.BG_ID = atoi (Polygon2.STATE) * 100000 + Polygon2.INDEX
    Out.PUMA = atoi (Polygon1.PUMA5)
    Out.STATE = Polygon2.STATE
    Out.TRACT = atoi(Polygon2.TRACT)
    Out.BLKGRP = atoi(Polygon2.BLKGROUP)
    IF (Out.BG_ID == 0) THEN
        Out.BG_ID = atoi (Polygon4.STATE) * 100000 + Polygon4.INDEX
        Out.PUMA = atoi (Polygon3.PUMA5)
        Out.STATE = Polygon4.STATE
        Out.TRACT = atoi(Polygon4.TRACT)
        Out.BLKGRP = atoi(Polygon4.BLKGROUP)
    ENDIF
ENDIF
```

Step 8: Create a Synthetic Population for the Year 2000

Given the PUMS data and the marginal distributions for each block group from the SF3 file, the PopSyn program can synthesize a population for the year 2000. The two-step IPF process first estimates the joint-distribution of the control attributes for each PUMA and then applies this distribution to each block group based on the block group control totals. As mentioned earlier, this can be done separately for family and non-family households or total households. In this case, total households were used.

PopSyn then uses the block group cross-classification tables to randomly assign synthetic households to the block group. These households are then randomly assigned to an activity location within the block group proportional to the location weight. The specified attributes for each household, person, and vehicle are copied from the PUMS record and written to the TRANSIMS data files. The vehicles are assigned to a parking lot connected to the household activity location.

The PopSyn control file is shown in Figure 30.

Figure 30: PopSyn Control Keys

```
TITLE                          Portland Year 2000 Population Synthesis
DEFAULT_FILE_FORMAT            TAB_DELIMITED
PROJECT_DIRECTORY              ../

#---- Input Files ----

NET_DIRECTORY                  ../network

NET_ACTIVITY_LOCATION_TABLE    Activity_Location
NET_PROCESS_LINK_TABLE         Process_Link

PUMS_HOUSEHOLD_FILE            demand/2000.PUMS.Household.txt
PUMS_POPULATION_FILE           demand/2000.PUMS.Population.txt

ZONE_DATA_FILE                 demand/2000.Zone_Data.txt

VEHICLE_TYPE_DISTRIBUTION      inputs/Vehicle_Distribution.txt

#---- Output Files ----

NEW_HOUSEHOLD_FILE             demand/2000.Household.txt
NEW_POPULATION_FILE            demand/2000.Population.txt
NEW_VEHICLE_FILE               demand/2000.Vehicle.txt
NEW_PROBLEM_FILE               results/2000.PopSyn_Problem.txt

POPSYN_REPORT_1                PUMS_HOUSEHOLD_SUMMARY
POPSYN_REPORT_2                PUMS_POPULATION_SUMMARY
POPSYN_REPORT_3                SYNTHETIC_HOUSEHOLD_SUMMARY
POPSYN_REPORT_4                SYNTHETIC_POPULATION_SUMMARY

#---- Parameters ----

STATE_PUMA_LIST                4101301, 4101302, 4101303, 4101304, 4101305, 4101306,
                               4101307, 4101308, 4101309, 4101310, 4101311, 4101312,
                               4101313, 5302101, 5302102, 5302200

PUMS_WEIGHT_FIELD              WEIGHT
##PUMS_VEHICLE_FIELD             VEHICLES
PUMS_AGE_FIELD                 AGE
```

```
ZONE_DATA_ID_FIELD              ZONE
LOCATION_ZONE_FIELD             BG_ID

ZONE_TOTAL_FIELD_1              HOUSEHOLDS

PUMS_ATTRIBUTE_FIELD_1_1        INCOME
PUMS_ATTRIBUTE_BREAKS_1_1       10000, 15000, 20000, 25000, 30000, 35000,
                                40000, 50000, 60000, 75000, 100000, 125000,
                                150000, 200000
ZONE_FIELD_GROUP_1_1            INCOME

PUMS_ATTRIBUTE_FIELD_1_2        HHAGE
PUMS_ATTRIBUTE_BREAKS_1_2       25, 35, 45, 55, 65, 75, 85
ZONE_FIELD_GROUP_1_2            AGE

PUMS_ATTRIBUTE_FIELD_1_3        PERSONS
PUMS_ATTRIBUTE_BREAKS_1_3       2, 3, 4, 5, 6, 7
ZONE_FIELD_GROUP_1_3            SIZE

STARTING_HOUSEHOLD_ID           1
STARTING_VEHICLE_ID             1

RANDOM_NUMBER_SEED              12332
MAXIMUM_IPF_ITERATIONS          10000
MAXIMUM_IPF_DIFFERENCE          0.0000001

OUTPUT_HOUSEHOLD_FIELDS         HHOLD, STATE, PUMA, PERSONS, VEHICLES, WORKERS,
                                INCOME, HHAGE, NUM_LT5, NUM_5TO15
OUTPUT_POPULATION_FIELDS        HHOLD, PERSON, GENDER, AGE, WORKER

LOCATION_WARNING_FLAG           FALSE
```

Step 9: Assemble the Marginal Totals for 1996

Once the baseline population for 2000 was synthesized, a backcast was performed to synthesize population for 1996 – the target year for the model calibration. As part of the original Portland study, METRO estimated marginal distributions of the three control attributes at the block group level for the year 1996. METRO did not make a distinction between family and non-family households. The distributions of household size, income, and age of householder were estimated using seven categories for each attribute. The definition of each category is shown in Table 12.

Unfortunately, METRO's 1996 estimates were based on the 1990 block group definitions. As shown in Table 13, the 2000 Census included numerous changes to PUMAs, Census Tracts, and block groups. ArcGIS was used to create a mapping between the 1990 block groups and the 2000 block groups. If the 1990 block group mapped to more than one 2000 block group, METRO's 1996 household attributes were split proportionally to the area of the year 2000 block groups. Year 2000 block groups that mapped to more than one 1990 block group were assigned the sum of the 1990 block group distributions.

Table 13: Difference between 1990 and 2000 Census Geography

Geographic Boundary	1990 Census			2000 Census		
	Oregon	Washington	Total	Oregon	Washington	Total
PUMA	6	2	6	13	3	16
Census Tracts	274	54	330	311	54	365
Block Groups	927	181	1108	925	232	1157

Step 10: Create the Synthetic Population for the Year 1996

The 1996 marginal distributions based on the year 2000 block group boundaries were then applied to the year 2000 synthetic population to estimate the 1996 synthetic population. Since the year 2000 population was already distributed to block groups, a one-step IPF process was used to adjust the household distributions to the 1996 control totals.

4.3 Results and Validation

The household control demographics are the age of head of household (HH-AGE), size (HH-SIZE), and income (HH-INCOME). The IPF procedures ensure that the marginal distributions of these control variables are met. It is the uncontrolled household and person demographics that need to be validated. In this study, the uncontrolled household demographics include vehicle ownership (HH-VEHICLES), number of workers (HH-WORKERS), number of children less than 5 years (HH-NUMLT5), number of children between the ages of 5 and 15 years (HH-NUM5TO15), and the number of adults between the ages of 26 and 45 years (HH-NUM26TO45).

The population synthesis results were compared against Portland METRO's independent estimates and the previous Portland PopSyn effort performed using TRANSIMS Version 3.x software and the 1990 Census data. As shown in Table 14, the population synthesizer generated 629,615 households, 1,627,869 persons and 807,792 workers. The number of households is slightly less than the previous estimate, but the population and number of workers are closer to METRO's estimates. The vehicle estimates are discussed in more detail below. Table 15 presents the range of results for other household attributes.

Table 14: Comparison of Regional Demographic Estimates

	METRO 1996 Estimate	1990→1996 PopSyn 3.x		2000→1996 PopSyn 4.0	
Households	639,090	636,531	-0.4%	629,615	-1.5%
Population	1,629,680	1,611,515	-1.1%	1,627,869	-0.1%
Workers	780,512	821,000	5.2%	807,792	3.5%
Vehicles	1,162,146	1,162,327	0.0%	1,184,451	1.9%

Table 15: Range of Synthetic Household Characteristics

Attribute	*Minimum*	*Maximum*	*Mean*	*Std Deviation*
HH-SIZE	1	17	2.58	1.42
HH-INCOME	0	748,000	40,871	38,707
HH-AGE	16	93	47.90	17.25
HH-VEHICLES	0	10	1.88	0.81
HH-WORKERS	0	8	1.28	0.90
HH-NUMLT5	0	5	0.20	0.50
HH-NUM5TO15	0	8	0.47	0.86
HH-NUM26TO45	0	7	0.85	0.85

4.3.1 Household Size and Income

The distributions of household size, income, and the age of head of household are controlled by the synthesis process, and hence they closely match the previous estimates and the 1996 marginal data. The previous model slightly under estimated household size (2.53 vs. 2.55) and the new model slightly over estimated household size (2.58 vs. 2.55). Figure 31 compares the household size distribution from the two models. In both models, about 60 percent of the households have two or less household members. Figure 32 compares the income distributions and shows that the new model replicates the METRO estimates more closely.

Figure 31: Distribution of Household by Household Size

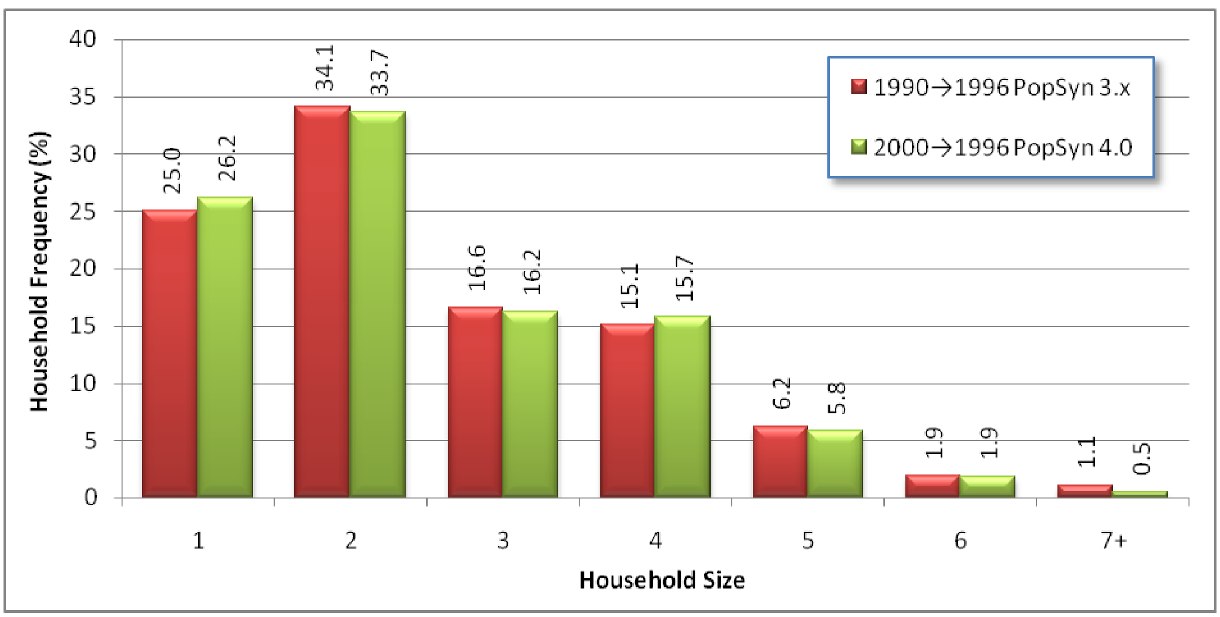
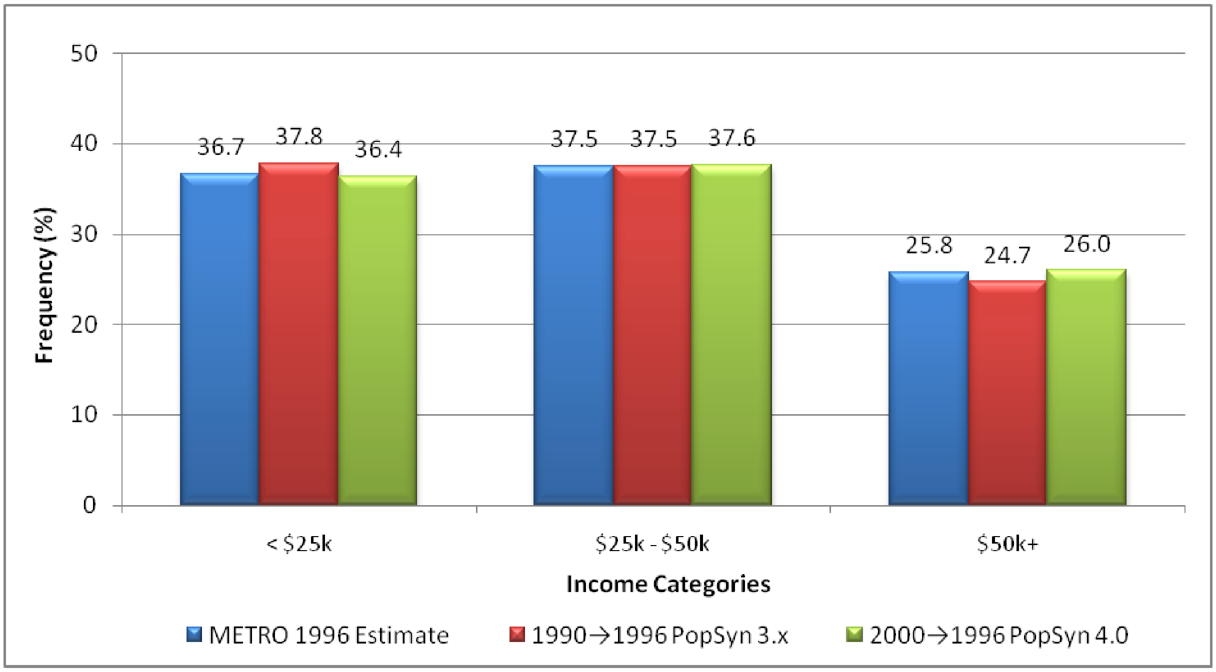


Figure 32: Distribution of Households by Income Category



4.3.2 Vehicles

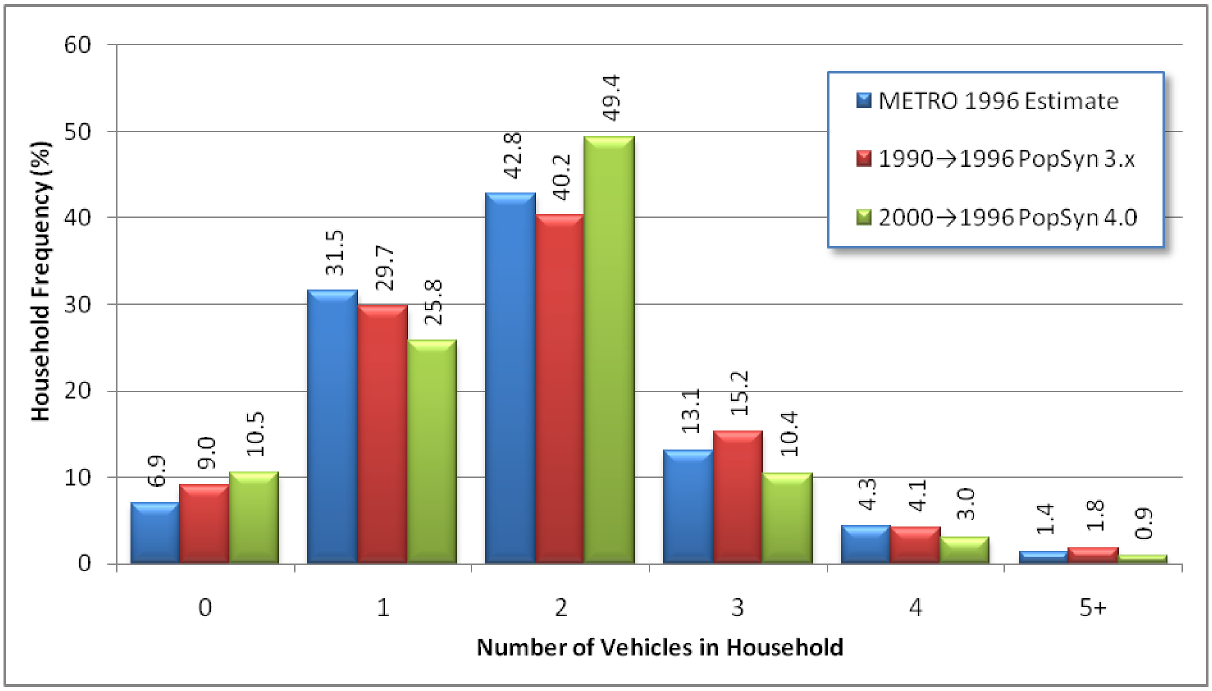
The PopSyn software provides two methods for generating household vehicles. The first method copies the vehicle ownership data from the PUMS household records. The second method adds a vehicle for each household member who is 16 years of age or older. The original Portland study selected the second method to simplify the switching of tours between highway and transit modes. This made a vehicle available if or when it was needed for a given trip or tour.

If PopSyn is executed using the PUMS-based approach, 1,012,280 vehicles or 1.61 vehicles per household are generated. If the age-based method is used, the estimate is 1,184,451 vehicles or 1.88 vehicles per household. This value is comparible to the 1,162,146 vehicles (1.82 veh/hh) estimated by METRO and the 1,162,327 vehicles (1.83 veh/hh) estimated by the previous model. Since age is an uncontrolled attribute, the difference in these estimates is related to the change in the age distribution between 1990 and 2000.

Properly accounting for households without access to vehicles is important for the activity generation and mode choice models. In this study, however, the mode specified for a trip or tour in the household activity survey was assumed for location and mode choice. In other words, vehicle availability is considered in selecting survey households and tour modes, and not based on the vehicles generated by the population synthesizer. If vehicle availability is calculated from the PUMS data, the previous model would estimate that 9.0 percent of the household do not own a car compared to 10.5 percent in the new model. This compares to 6.9 percent estimated by METRO.

A third approach to vehicle ownership is to preserve the households that own zero autos, but use the age-based approach for other households. This hybrid approach eliminates the option for a zero auto household to select drive as a trip or tour mode while at the same time simplifying the vehicle scheduling complications for other households. The impacts of this approach are shown in Figure 33. The method generated 1,087,475 vehicles (1.73 veh/hh) with 10.5 percent of households without a vehicle. It under estimates the number of households with only one vehicle and over estimates the number of households with two vehicles. It also under estimates the number of households with three or more vehicles. This implies that a number of households have more vehicles than licensed drivers.

Figure 33: Distribution of Households by Vehicle Ownership



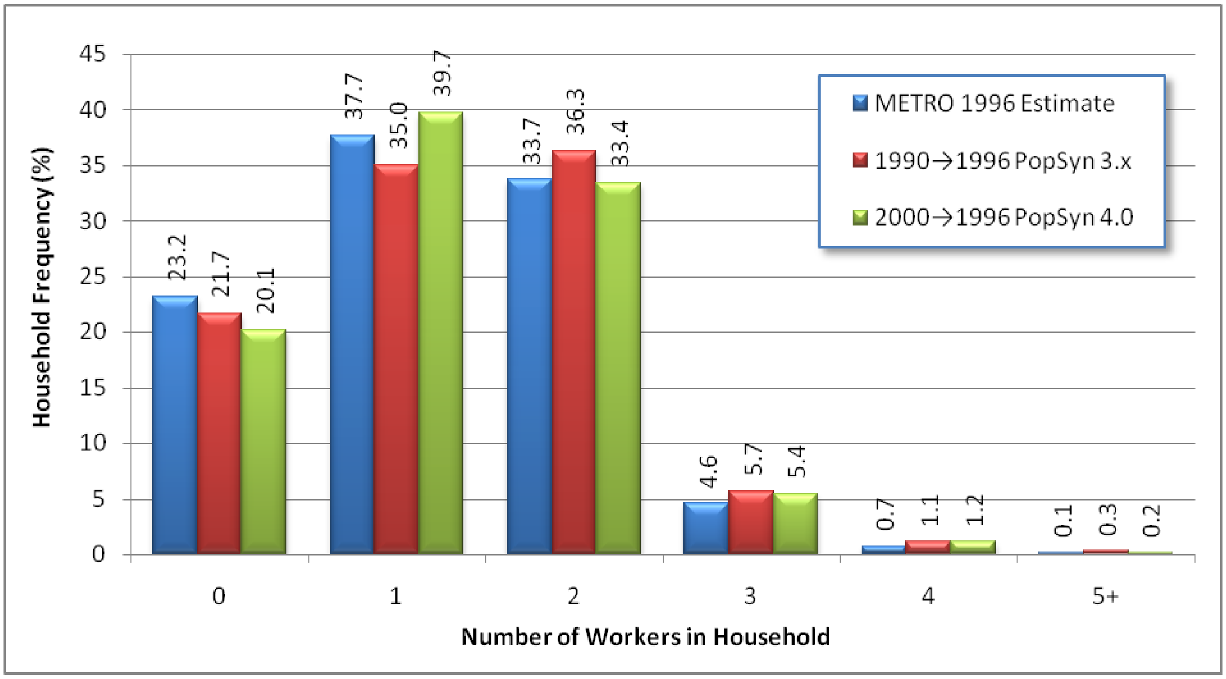
4.3.3 Workers

The population synthesizer estimates that the number of workers is 807,792 or 3.5 percent more than METRO's independent estimate of 780,521 workers. The PopSyn estimate represents 1.28 workers per household compared to the 1.27 workers per household generated by the original study. Table 16 shows that 49.6 percent of the population are workers, 24.3 percent are non-workers, and 26.1 percent are ineligible to work (16 years or younger). The distribution of workers shown in Figure 34 tracks reasonably well with both the METRO and previous estimates.

Table 16: Employment Status within the Population

Employment Status	Frequency	Percentage of Population
Ineligible to work	424,590	26.1
Workers	807,792	49.6
Non-workers	395,487	24.3
Total	**1,627,869**	**100.0**

Figure 34: Distribution of Households by Number of Workers



The number of workers is to a certain extent correlated to household income levels. For example, current results indicate that upwards of 77 percent of workers reside in households with middle and high income levels. By comparison, METRO estimated 70 percent and the previous model estimated 80 percent (Figure 35). The number of vehicles available to the household is strongly correlated to the number of workers as seen in Figure 36. The model over estimates the proportion of households where 1) workers do not have access to a vehicle and 2) the number workers is identical to the number of vehicles owned, and under estimates the number of households with more vehicle than workers. This is likely related to the observation that households appear to have more vehicles than licensed drivers.

Figure 35: Cross-Classification of Workers by Income Level

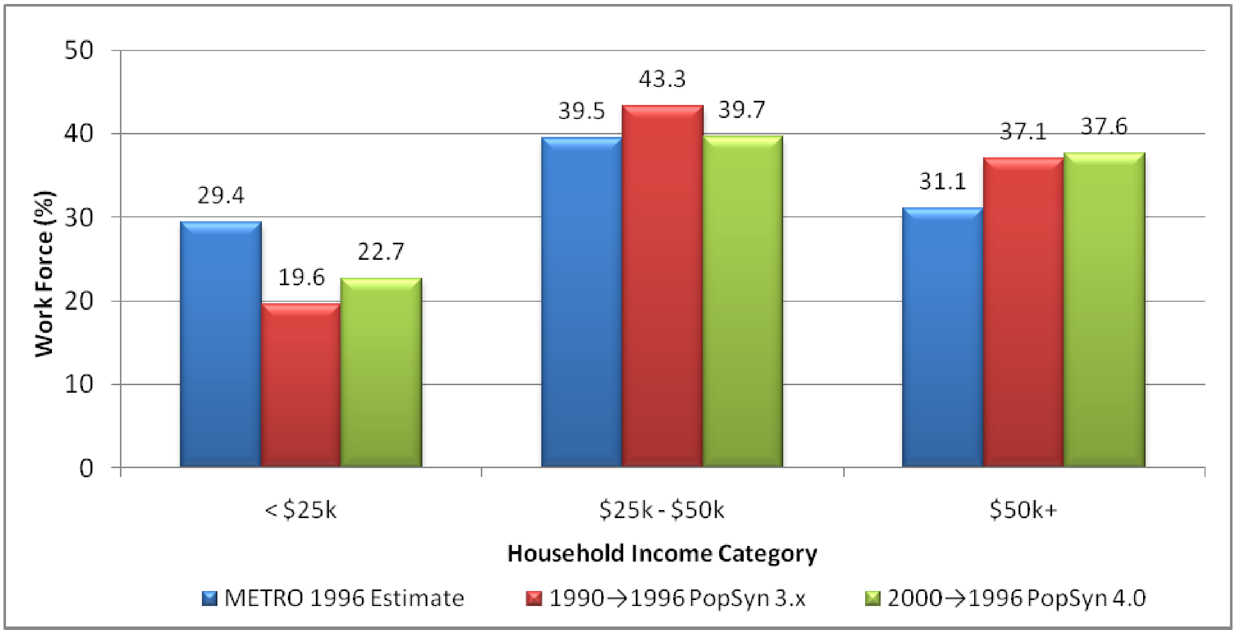
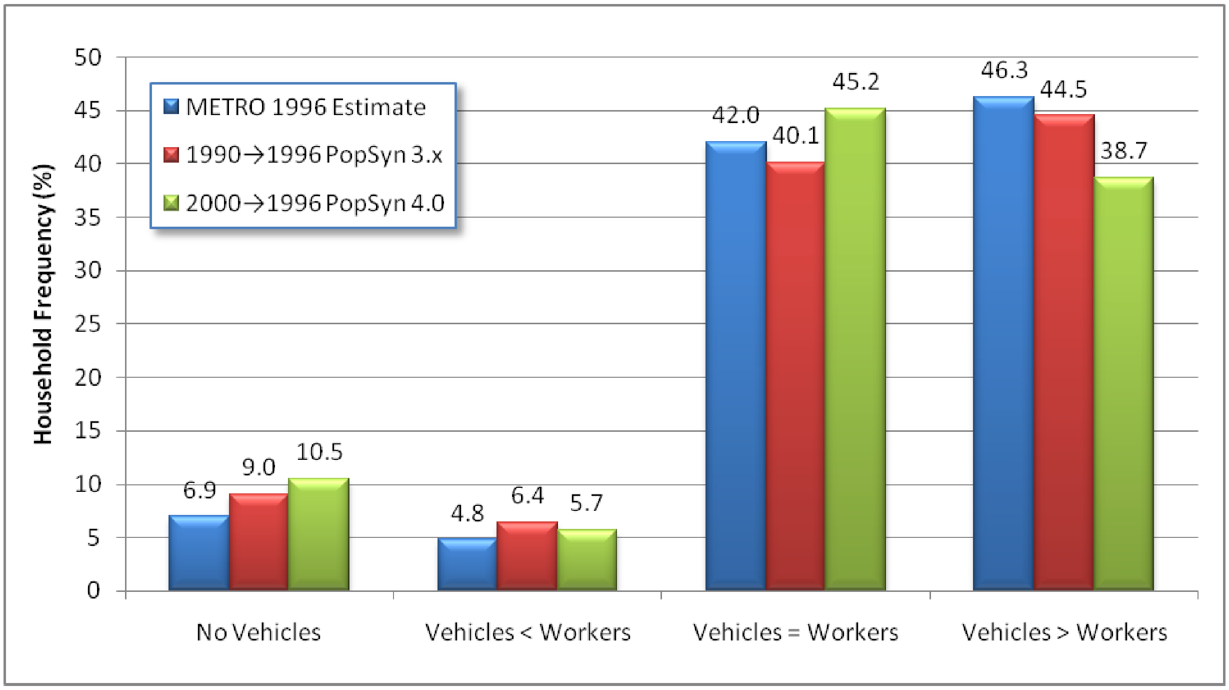


Figure 36: Cross-Classification of Workers by Vehicle Ownership



Chapter 5 Activity Generator

This chapter summarizes the steps taken to calibrate and validate the TRANSIMS activity generator for Portland. The population synthesizer and activity generator collectively form the activity-based demand model for the Portland region. Activity generation in TRANSIMS is achieved using the ActGen program, which has undergone intensive transformation and testing as it progressed from Version 3.x to Version 4.0 software. The ActGen program has three main functions:

1. Generate a daily schedule of activities for each person in the synthetic population,
2. Assign a location for each activity, and
3. Schedule the activities based on time constraints.

ActGen could be followed by mode choice to assign a travel mode to the trips within a tour. Alternatively, the mode could be accepted from the household survey or set within the activity generator script. For this application, the mode was accepted from the household activity survey.

Three major implementations of the ActGen model were performed as part of this study. All of these implementations involve, to varying extents, calibrating and validating the ActGen model using TRANSIMS 4.0 software and submitting the resulting activity file to the Router and Microsimulator to estimate the network performance impacts for feedback to ActGen. While these implementations employ the same location choice function, household classification tree, and input, they mostly differ in 1) how the location choice model is actually applied (side constraints), 2) how calibration is achieved, and 3) how integration with the Router-Microsimulator is designed.

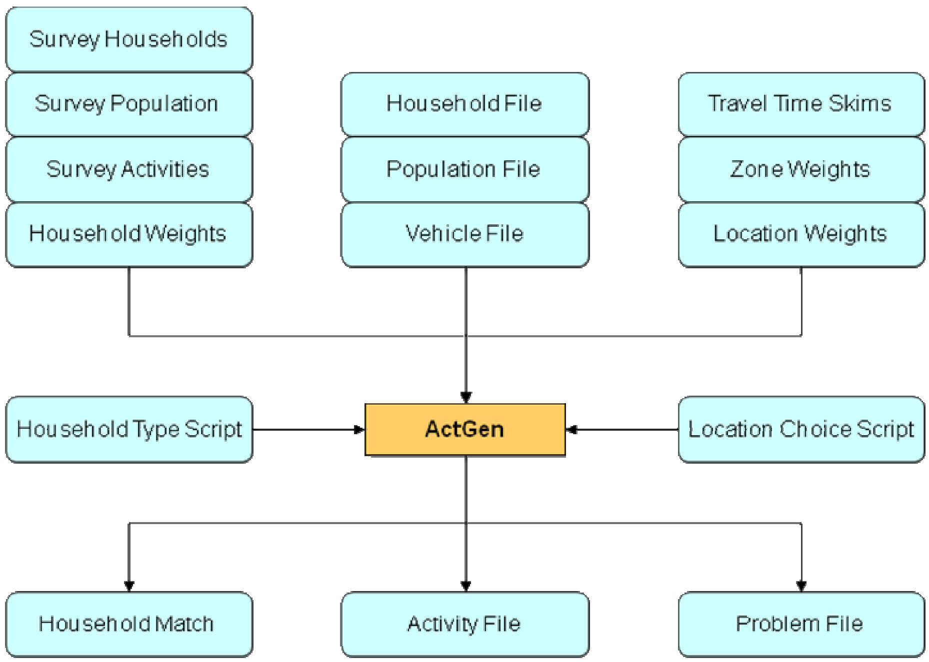
The first ActGen implementation is in many ways similar to the original GEN2 model developed during the first phase of the Portland project. The primary difference is the software conversion from Version 3.x to Version 4.0. The second implementation is marked by the introduction of time budgets, regeneration of problematic households, and the initial interaction between the activity-based demand models and the network performance simulation. The third implementation involved activity survey cleaning, the adoption of time budget penalties, the explicit splitting of the ActGen program into the ActivityPattern and LocationChoice modules, activity schedule adjustments, and the use of ActBalance for seamless calibration of the activity model.

The remainder of this chapter describes the general activity generation process and the components that are common to each of the three implementations. Subsequent chapters discuss the specific details for each of the three models.

5.1 ActGen Data Requirements

The input data requirements and primary outputs of the ActGen program are shown in Figure 37. These include four data files from the household activity survey – household records, household person (population) records, travel activity patterns for each person, and household weights used to adjust the probability that a household of a given type will be selected. It also reads the three files generated by the population synthesizer – synthetic households, persons within households (population), and the vehicles available to each household. The third set of input data are used by the location choice model. They include zone-to-zone travel time skims, attraction weights for each zone, and attraction weights for each activity location within the zone.

Figure 37: Activity Generator Inputs and Outputs



In addition to the basic data files, two or more modeling scripts are provided. The household type scripts perform logical tests on data fields included in the survey and synthetic household files to assign each household to a household classification type. This type is used to match a survey household record to each synthetic household record.   Location choice scripts are called for each trip within the tours assigned to a synthetic household member. They apply the location choice model to the travel time and attraction weight for each destination zone and return the utility used to calculate the probability that a given zone will be selected for the trip destination.

The output of the process is a household match file, activity file, and problem file. The household match file records the survey household and person that was matched to each synthetic household and person. The activity file contains the start time, end time, and duration of each activity performed by each household person plus the travel mode and vehicle that is to be used to travel to that activity location. The problem file lists the activities that the software was unable to accommodate and the reason it could not be included. It also lists activities the software took corrective actions to repair before adding it to the output activity file.

5.2 Household Classification Tree

The binary classification tree developed for the original GEN2 model was used for this effort. The classification tree groups households with a similar number of out-of-home activities. The grouping is based on seven household attributes:

- Household size [SIZE]
- Age of head of household [HHAGE]
- Number of workers [WORKERS]
- Income [INCOME]
- Number of children less than 5 [LT5]
- Number of children between 5 and 15 [5TO15]
- Number of adults between 26 and 45 [26TO45]

The classification was further stratified for three area types: very urban, urban, and suburban for a total of 55 different household types. The classification trees for very urban, urban, and suburban households are presented in Table 17, Table 18, and Table 19 respectively. The numbers reported between square brackets represent the household type.

Table 17: Classification Tree for Very Urban Households

WORKERS = 0	*WORKERS = 1*	*WORKERS > 1*
SIZE = 1	SIZE = 1 [5]	SIZE = 2 [7]
HHAGE <= 38 [1]	SIZE > 1 [6]	SIZE > 2 [8]
HHAGE > 38 [2]		
HHAGE <= 53 [3]		
HHAGE > 53 [4]		

Table 18: Classification Tree for Urban Households

WORKERS = 0	*WORKERS = 1*	*WORKERS = 2*	*SIZE <= 3*	*SIZE > 3*
SIZE = 1	SIZE = 1 [13]	SIZE = 2 [24]	[27]	[28]
HHAGE <= 38 [9]	SIZE = 2	SIZE = 3 []		
HHAGE > 38 [10]	5TO15 = 0 [14]	LT5 = 0 [25]		
HHAGE <= 53 [11]	5TO15 > 0 [15]	LT5 > 0 [26]		
HHAGE > 53 [12]	SIZE = 3			
	5TO15 <= 1 [16]			
	HHAGE <= 29 [17]			
	HHAGE > 29 [18]			
	5TO15 > 1 [19]			
	SIZE = 4			
	5TO15 <= 1 [20]			
	5TO15 > 1 [21]			
	5TO15 <= 2 [22]			
	5TO15 > 2 [23]			

Table 19: Classification Tree for Suburban Households

WORKERS = 0	***WORKERS = 1***	***WORKERS = 2***	***SIZE = 3***	***SIZE = 4***	***SIZE > 4***
SIZE = 1	SIZE = 1 [35]	SIZE = 2 [45]	[52]	HHAGE <= 46 [53]	[55]
HHAGE <= 38 [29]	SIZE = 2	SIZE = 3 [46]		HHAGE > 46 [54]	
HHAGE > 38 [30]	5TO15 = 0 [36]	SIZE = 4			
SIZE = 2	5TO15 > 0 [37]	C_LT5 = 0 [47]			
HHAGE <= 53	SIZE = 3	26TO45 = 0 [48]			
INCOME <= 5 [31]	5TO15 = 0	26TO45 > 0 [49]			
INCOME > 5 [32]	HHAGE <= 29 [38]	5TO15 <= 2 [50]			
HHAGE > 53 [33]	HHAGE > 29 [39]	5TO15 > 2 [51]			
SIZE > 2 [34]	5TO15 > 0 [40]				
	SIZE = 4				
	5TO15 <= 1 [41]				
	5TO15 > 1 [42]				
	5TO15 <= 2 [43]				
	5TO15 <= 2 [44]				

In the Version 3.x software, the classification tree was input through the relatively complex set of binary tests and line number references shown in Figure 38. This made the interpretation of the tree difficult to understand and subject to considerable coding error. The Version 4.0 software introduced a scripting language that can be used to construct a wide variety of logical tests for defining the household type. Alternatively, a household type field can be included in the input household file and/or the input household survey records. The classification script for this model is shown in Figure 39.

Since there are 4,562 survey households and 629,615 synthetic households, or an average of one survey household for every 138 synthetic households, it is important to verify that each of the 55 household types includes a reasonable number of survey records to provide some level of randomness to the activity patterns assigned to a given household type. Table 20 shows the number of synthetic households and survey records included in each household type. The types with ratios of synthetic households to survey households of 200 or greater are highlighted. These households typically have three or more workers, large families, and suburban addresses. The largest ratio (368:1), however, is zero worker households with three or more family members living in suburban areas. This household type was under sampled by the activity survey.

Figure 38: Binary Classification Tree used in the TRANSIMS 3.x Software

8	1.5	1
7	0.5	2
1	1.5	4
6	38.5	8
0	0.0	16
0	0.0	17
6	53.5	9
0	0.0	18
0	0.0	19
7	1.5	5
1	1.5	10
0	0.0	20
0	0.0	21
1	2.5	11
0	0.0	22
0	0.0	23
8	2.5	3
7	0.5	6
1	1.5	12
6	38.5	24
0	0.0	48
0	0.0	49
6	53.5	25
0	0.0	50
0	0.0	51
7	1.5	13
1	2.5	26
1	1.5	52
0	0.0	104
4	0.5	105
0	0.0	210
0	0.0	211
1	4.5	53
1	3.5	106
4	1.5	212
6	29.5	424
0	0.0	848
0	0.0	849
0	0.0	425
4	1.5	213
0	0.0	426
0	0.0	427
4	2.5	107
0	0.0	214
0	0.0	215
7	2.5	27
1	3.5	54
1	2.5	108
0	0.0	216
0	0.0	217
3	0.5	109
0	0.0	218
0	0.0	219
1	3.5	55
0	0.0	110
0	0.0	111
7	0.5	7
1	1.5	14
6	38.5	28
0	0.0	56
0	0.0	57
1	2.5	29
6	53.0	58
2	5.5	116
0	0.0	232
0	0.0	233
0	0.0	117
0	0.0	59
7	1.5	15
1	2.5	30
1	1.5	60
0	0.0	120
4	0.5	121
0	0.0	242
0	0.0	243
1	4.5	61
1	3.5	122
4	1.5	244
6	29.5	488
0	0.0	976
0	0.0	977
0	0.0	489
4	1.5	245
0	0.0	490
0	0.0	491
4	2.5	123
0	0.0	246
0	0.0	247
7	2.5	31
1	3.5	62
1	2.5	124
0	0.0	248
0	0.0	249
1	4.5	125
3	0.5	250
5	0.5	500
0	0.0	100
0	0.0	100
0	0.0	501
4	2.5	251
0	0.0	502
0	0.0	503
1	4.5	63
1	3.5	126
0	0.0	252
6	46.5	253
0	0.0	506
0	0.0	507
0	0.0	127

Figure 39: Household Classification Script used in the TRANSIMS 4.0 Software

```
#---- Very Urban ----
IF(Household.LOCATION==1)THEN
 IF(Household.WORKERS==0)THEN
  IF(Household.HHSIZE==1)THEN
   IF(Household.HHAGE<=38)THEN
    RETURN(1)
   ELSE
    RETURN(2)
   ENDIF
  ELSE
   IF(Household.HHAGE<=53)THEN
    RETURN(3)
   ELSE
    RETURN(4)
   ENDIF
  ENDIF
 ELSE
  IF(Household.WORKERS==1)THEN
   IF(Household.HHSIZE==1)THEN
    RETURN(5)
   ELSE
    RETURN(6)
   ENDIF
  ELSE
   IF(Household.HHSIZE==2)THEN
    RETURN(7)
   ELSE
    RETURN(8)
   ENDIF
  ENDIF
 ENDIF
#---- urban ----
ELSE
 IF(Household.LOCATION==2)THEN
  IF(Household.WORKERS==0)THEN
   IF(Household.HHSIZE==1)THEN
    IF(Household.HHAGE<38)THEN
     RETURN(9)
    ELSE
     RETURN(10)
    ENDIF
   ELSE
    IF(Household.HHAGE<=53)THEN
     RETURN(11)
    ELSE
     RETURN(12)
    ENDIF
   ENDIF
  ELSE
   IF(Household.WORKERS==1)THEN
    IF(Household.HHSIZE==1)THEN
     RETURN(13)
    ELSE
     IF(Household.HHSIZE==2)THEN
      IF(Household.A5TO15==0)THEN
       RETURN(14)
      ELSE
       RETURN(15)
      ENDIF
     ELSE
      IF(Household.HHSIZE==3)THEN
       IF(Household.A5TO15<=1)THEN
        IF(Household.HHAGE<=29)THEN
         RETURN(16)
        ELSE
         RETURN(17)
        ENDIF
       ELSE
        RETURN(18)
       ENDIF
      ELSE
       IF(Household.HHSIZE==4)THEN
        IF(Household.A5TO15<=1)THEN
         RETURN(19)
        ELSE
         RETURN(20)
        ENDIF
       ELSE
        IF(Household.A5TO15<=2)THEN
         RETURN(21)
        ELSE
         RETURN(22)
        ENDIF
       ENDIF
      ENDIF
     ENDIF
    ENDIF
   ELSE
    IF(Household.WORKERS==2)THEN
     IF(Household.HHSIZE==2)THEN
      RETURN(23)
     ELSE
      IF(Household.HHSIZE==3)THEN
       RETURN(24)
      ELSE
       IF(Household.ALT5==0)THEN
        RETURN(25)
       ELSE
        RETURN(26)
       ENDIF
      ENDIF
     ENDIF
    ELSE
     IF(Household.HHSIZE<=3)THEN
      RETURN(27)
     ELSE
      RETURN(28)
     ENDIF
    ENDIF
   ENDIF
  ENDIF
#---- suburban ----
 ELSE
  IF(Household.WORKERS==0)THEN
   IF(Household.HHSIZE==1)THEN
    IF(Household.HHAGE<=38)THEN
     RETURN(29)
    ELSE
     RETURN(30)
    ENDIF
   ELSE
    IF(Household.HHSIZE==2)THEN
     IF(Household.HHAGE<=53)THEN
      IF(Household.INCOME<=5)THEN
       RETURN(31)
      ELSE
       RETURN(32)
      ENDIF
     ELSE
      RETURN(33)
     ENDIF
    ELSE
     RETURN(34)
    ENDIF
   ENDIF
  ELSE
   IF(Household.WORKERS==1)THEN
    IF(Household.HHSIZE==1)THEN
     RETURN(35)
    ELSE
     IF(Household.HHSIZE==2)THEN
      IF(Household.A5TO15==0)THEN
       RETURN(36)
      ELSE
       RETURN(37)
      ENDIF
     ELSE
      IF(Household.HHSIZE==3)THEN
       IF(Household.A5TO15<=1)THEN
        IF(Household.HHAGE<=29)THEN
         RETURN(38)
        ELSE
         RETURN(39)
        ENDIF
       ELSE
        RETURN(40)
       ENDIF
      ELSE
       IF(Household.HHSIZE==4)THEN
        IF(Household.A5TO15<=1)THEN
         RETURN(41)
        ELSE
         RETURN(42)
        ENDIF
       ELSE
        IF(Household.A5TO15<=2)THEN
         RETURN(43)
        ELSE
         RETURN(44)
        ENDIF
       ENDIF
      ENDIF
     ENDIF
    ENDIF
   ELSE
    IF(Household.WORKERS==2)THEN
     IF(Household.HHSIZE==2)THEN
      RETURN(45)
     ELSE
      IF(Household.HHSIZE==3)THEN
       RETURN(46)
      ELSE
       IF(Household.HHSIZE==4)THEN
        IF(Household.ALT5==0)THEN
         IF(Household.A26TO45==0)THEN
          RETURN(47)
         ELSE
          RETURN(48)
         ENDIF
        ELSE
         RETURN(49)
        ENDIF
       ELSE
        IF(Household.A5TO15<=2)THEN
         RETURN(50)
        ELSE
         RETURN(51)
        ENDIF
       ENDIF
      ENDIF
     ENDIF
    ELSE
     IF(Household.HHSIZE==3)THEN
      RETURN(52)
     ELSE
      IF(Household.HHSIZE==4)THEN
       IF(Household.HHAGE<=46)THEN
        RETURN(53)
       ELSE
        RETURN(54)
       ENDIF
      ELSE
       RETURN(55)
      ENDIF
     ENDIF
    ENDIF
   ENDIF
  ENDIF
 ENDIF
ENDIF
```

Table 20: Ratio of Synthetic Households to Survey Households by Household Type

Type	Synthetic Households	Survey Households	Ratio	Type	Synthetic Households	Survey Households	Ratio
1	608	9	67.6	29	1,483	20	74.2
2	3,590	45	79.8	30	52,909	281	188.3
3	180	8	22.5	31	773	21	36.8
4	82	10	8.2	32	4,972	28	177.6
5	5,295	138	38.4	33	14,121	315	44.8
6	1,033	19	54.4	34	16,545	45	367.7
7	2,592	43	60.3	35	63,572	423	150.3
8	508	8	63.5	36	40,568	299	135.7
9	727	16	45.4	37	813	38	21.4
10	13,629	107	127.4	38	2,860	23	124.3
11	1,366	16	85.4	39	21,251	107	198.6
12	2,847	50	56.9	40	6,788	26	261.1
13	15,908	255	62.4	41	14,549	85	171.2
14	7,017	104	67.5	42	10,404	53	196.3
15	38	19	2.0	43	9,750	49	199.0
16	414	9	46.0	44	6,937	35	198.2
17	2,839	16	177.4	45	121,756	617	197.3
18	540	5	108.0	46	36,165	268	134.9
19	1,481	16	92.6	47	3,627	29	125.1
20	1,179	9	131.0	48	23,257	165	141.0
21	1,071	6	178.5	49	13,593	70	194.2
22	730	5	146.0	50	12,397	69	179.7
23	18,714	192	97.5	51	8,817	47	187.6
24	3,767	63	59.8	52	16,419	79	207.8
25	3,175	38	83.6	53	6,826	34	200.8
26	1,855	21	88.3	54	8,710	28	311.1
27	2,830	17	166.5	55	12,326	48	256.8
28	3,389	16	211.8				

5.3 Activity Pattern Matching

The basic concept behind the activity pattern model is to match each synthetic household to a survey household and then match each synthetic household member to a survey household member. The activities from the survey are then copied to the corresponding synthetic household member. The matching is done based on household characteristics and essentially provides a simple method for factoring up the survey activity patterns to the entire population.

The ActGen program requires four types of data to do the matching: a household classification script, a synthetic population, a survey population and survey activities. The household classification script was discussed in the previous section. The synthetic population includes household, person and vehicle files and is typically the result of a population synthesizer. The survey population includes household and person files as well as the complete daily activity diary for each survey person.

The steps involved in assigning the skeletal activity patterns to each person in the synthetic population are described below:

Step 1. Using a household classification script, classify synthetic and survey households by household type given their demographic attributes. There should be at least 10 survey households for each household type.

Step 2. Randomly assign a survey household to each synthetic household of the same type.

Step 3. Match each member of the synthetic household to an appropriate member of the survey household. The person-to-person matching is done based on age, work status and gender, in that order. Each person is assigned to an age group using the age ranges shown in Table 21.

Table 21: Age Group Definitions

Age Group	*Age Range*
1	0–4
2	5–11
3	12–15
4	16–20
5	21–64
6	65+

If the survey household does not have a person in the same age group, work status and gender as a person in the synthetic household, the second best match will be used. Children are only matched to children and adults are only matched with adults. If there are more children in the synthetic households than the survey household, the most appropriate survey household child will be cloned and matched to the synthetic household child. If there are more children in the survey household than the synthetic household, the least compatible children will be ignored. Adults are match in a similar way.

Step 4. After each household member is matched, the out-of-home activities of the corresponding survey person are copied to the synthetic household member. This includes the time schedule, purpose, and travel mode for all person tours. The location of the home activities are also set to the activity location of the synthetic household.

Step 5. The skeletal activity patterns are typically written to an Activity Pattern file for future reference and updates.

5.4 Location Choice Model

This study started with the same location choice model and zone-related data as the previous GEN2 model. The location choice model is a two-step process, whereby a zone is first selected based on inter-zonal impedances and zone attraction weights, and then an activity location is selected within that zone based on activity location weights and the distance between locations.

The model was originally developed by METRO as a trip-based destination choice model for the Portland region. It included nine trip purposes, namely home, work, shopping, recreation, visit, serve passenger, school, college, and other. Work purposes were further stratified by income level: low, middle, and high. The interchange impedances were based on logsums from the METRO mode choice model and the attraction weights included a variety of demographic and geographic attributes.

In converting the trip-based model to a tour-based model, a distinction was added between trips in an anchored tour and trips in a non-anchored tour. Anchored tours are tours that start and end at home and include work, school or college activities. If a tour includes multiple anchor activities, the location with the largest activity duration is selected as the anchor. Non-anchored tours start and end at home or start and end at the anchor (e.g., work) location. Table 22 presents the activities modeled in this study and their respective codes.

Table 22: Activity Purpose Codes

Activity Purpose	*Tour Type*	*Activity Code*
Home	Any	0
Work High Income	Anchored	1
Work Middle Income	Anchored	9
Work Low Income	Anchored	18
Shop	Anchored	2
Visit	Anchored	3
Recreation	Anchored	4
Other	Anchored	5
Serve Passenger	Anchored	6
School	Anchored	7
College	Anchored	8
Work High Income	Non-Anchored	11
Work Middle Income	Non-Anchored	17
Work Low Income	Non-Anchored	19
Shop	Non-Anchored	12
Visit	Non-Anchored	13
Recreation	Non-Anchored	14
Other	Non-Anchored	15
Serve Passenger	Non-Anchored	16

5.4.1 Locating Anchor Activities

The location choice model used to select the work, school or college activity location for an anchored tour is conditional on the home location defined by the population synthesizer. The model first selects a traffic analysis zone for the activity given the following logit-based model:

$$p_{(\text{anchor},j)|(\text{home},i)} = \frac{A_{\text{anchor},j} \times \exp\left[\beta_{\text{anchor}}\left(C_{ij}(ET_{\text{home}}) + C_{ji}(ET_{\text{anchor}})\right)\right]}{\sum_k A_{\text{anchor},k} \times \exp\left[\beta_{\text{anchor}}\left(C_{ik}(ET_{\text{home}}) + C_{ki}(ET_{\text{anchor}})\right)\right]} \tag{1}$$

Where $p_{(\text{anchor},j)|(\text{home},i)}$ is the probability that the anchor activity is located in zone j; $A_{\text{anchor},j}$ represents the attractiveness of zone j to host the anchor activity; ET_{home} is the end time for the home activity; β_{anchor} is a vector of mode-time factors for the anchor activity; and $C_{ij}(ET_{\text{home}})$ is the travel impedance or cost from zone i to j evaluated at ET_{home}.

The anchor activity is then assigned to an activity location within the zone based on the appropriate (work, school, or college) attraction weights of the activity locations within the zone and the straight-line distance between the home activity location and the anchor activity locations (see Equation (2)). This is intended to give preference to activity locations at opposite sides of nearby zones, but have no significant effect in selecting activity locations for far away zones.

$$p_{(\text{anchor},j)|(\text{home},i)} = \frac{A_{\text{anchor},j} \times D_{i,j}}{\sum_k A_{\text{anchor},k} \times D_{i,k}} \tag{2}$$

The skeletal activity pattern copied from the activity survey includes a location ID for each unique activity location visited by household members. If a traveler visits the same location more than once or if multiple household members travel to the same location, the activity pattern will contain the same location ID for each activity. This is particularly important for coordinating the activity locations for intra-household shared rides. So, when an activity location is selected, all other household activities with the same location ID are set to the specified location as well.

5.4.2 Locating Intermediate Stops

If the tour includes intermediate stops, a zone is selected for the intermediate stop based on the home and anchor activity locations. The order by which intermediate stops are located is based on the order they appear in the tour. The location choice model for selecting intermediate zones follows the same general form for the anchor activities except that the travel time impedances are based on the two closest located activities. The probability $p_{(n,z)}$ that an activity n is located in zone z given that home is located in zone i and the nearest previously located activity along the tour is in zone j is:

$$p_{(n,z)|(\text{home},i),(\text{nearest},j)} = \frac{A_{n,z} \times \exp\left[\beta_n \left(C_{iz}(ET_{\text{home}}) + C_{zj}(ET_n)\right)\right]}{\sum_k A_{n,k} \times \exp\left[\beta_n \left(C_{ik}(ET_{\text{home}}) + C_{kj}(ET_n)\right)\right]} \quad (3)$$

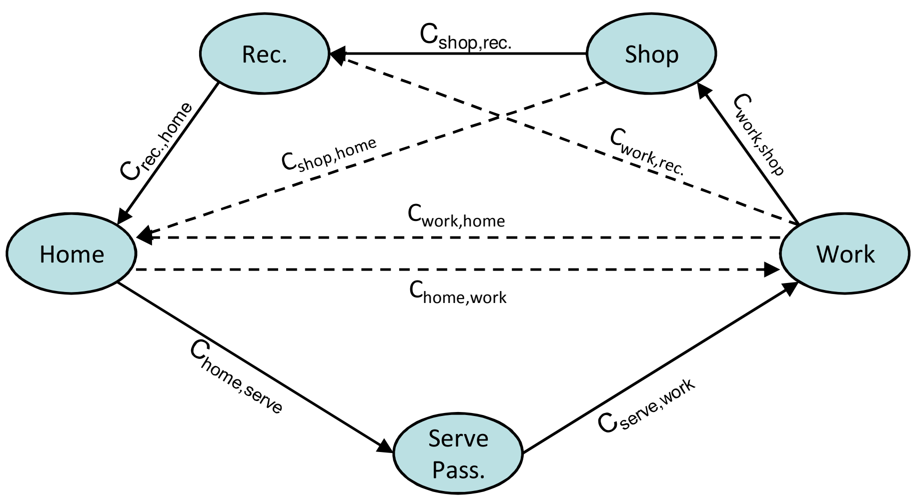
The intermediate stop in then assigned to an activity location within the selected zone based on the attraction weight for each activity location within the zone and the average distance between the two known locations and the activity location. Equation (4) gives preference to locations that are relatively far away from the two known locations.

$$p_{(n,z)|(\text{home},i),(\text{nearest, j})} = \frac{A_{n,j} \times (D_{i,n} + D_{n,j})/2}{\sum_k A_{n,k} \times (D_{i,n} + D_{n,j})/2} \quad (4)$$

5.4.3 Activity Location Illustration

To illustrate how the model locates activities, consider the tour depicted in Figure 40. The home location is fixed by the population synthesizer. ActGen first locates the work activity, because it is an anchor activity, using Equations (1) and (2). Then, the serve passenger activity is located given home and work locations (work is the nearest located activity other than home), using Equations (3) and (4). Shop is then located, given home and work locations, using Equations (3) and (4). Finally, these same equations are used to locate the recreation activity given the home and shop (nearest located activity other than home) locations.

Figure 40: Location Choice Example



The activity locations for destinations on non-anchored tours are implemented in the same way as intermediate stops on anchored tours. The process starts by locating the first out-of-home activity and then uses this activity location and home as the two known location points for selecting the next activity in the tour. The process continues until all of the activities are located.

5.5 Activity Scheduling and Vehicle Allocation

Once all of the activities are located, the activity generator adjusts the start and end times for each activity based on the estimated travel time between the activity locations. Travel times by mode and time of day are derived from the zone-to-zone skims provided by the user. The distance between the activity locations is compared to the distance between the zone centroids to calculate an adjustment factor for the zone-to-zone travel time. This time is then compared to the end time of the previous activity and the start time of the current activity to determine if a schedule adjustment is needed. If the adjusted travel time is greater than the available time budget, the activity schedule is modified to accommodate the estimated travel time. If the travel time is less than the available time budget, the activity schedule may be compressed.

The rules about how and where activity schedule adjustments are made depend on the schedule constraint code assigned to each activity. Most activities are typically unconstrained, meaning their time schedule is flexible and can be adjusted as needed. Other activities, such the work, may be constrained by the activity start time, end time, duration or any combination thereof. In these cases, the activity schedule is fixed by the time of the constrained activity and all activities before or after the constrained activity are rescheduled accordingly. For example, if the arrival time at work is fixed and the travel time indicates the trip will arrive 15 minutes late, the end time of the previous activity will be shifted 15 minutes earlier to ensure the traveler arrives at work on time.

Schedule adjustments in principal are relatively straightforward, but there are any number of complications that can arise. For example, a serve passenger activity is constrained by the schedule constraints of the passenger being served. If this passenger has a fixed arrival time at school and the driver has a fixed arrival time at work, the estimated travel times may make it impossible to satisfy both constraints. This type of problem can also arise within tours and between tours. If the travel times are long, multiple constraints can easily clash with one another and require reconciliation. ActGen includes several ways of compromising between conflicting activity schedules to accommodate the trips. Problem messages are generated for tours where a reasonable resolution cannot be found.

A similar process is used to schedule vehicle use. If the household has fewer vehicles than drivers, it is necessary to ensure that the vehicle assigned to a drive tour is available and at home when the tour starts. Obviously, if one household member drives the car to work in the morning, parks it at work for nine hours, and then drives the vehicle home in the evening, the vehicle is not available for another household member to use for a short tour in the middle of the day. The vehicle is not being used in the middle of the day, but it is also not parked at home.

Chapter 6 First ActGen Implementation

The first implementation of ActGen 4.0 resulted in an average of 14.4 activities per household or 5.5 activities per person per day. The distribution of activities by activity type is provided in Table 23. The home activity represents 33 percent of total activities, followed by work at 17 percent. If the home activities are removed, work represents 25.4 percent of the out-of-home activities. Note that about one third of all shop and recreation activities and two thirds of other activities are performed as part of an anchor-based tour. Serve passenger and visiting are evenly split.

Table 23: Distribution of Activities by Activity Type

Activity Purpose	Tour Type	Percentage	Non-Home %
Home	Any	33.0	--
Work High Income	Anchored	6.7	10.0
Work Middle Income	Anchored	6.2	9.3
Work Low Income	Anchored	4.1	6.1
Visit	Anchored	2.6	3.9
Shop	Anchored	3.9	5.8
Recreation	Anchored	3.9	5.8
Serve Passenger	Anchored	4.0	6.0
Other	Anchored	6.3	9.4
College	Anchored	2.0	3.0
School	Anchored	2.6	3.9
Work High Income	Non-Anchored	1.0	1.5
Work Middle Income	Non-Anchored	0.6	0.9
Work Low Income	Non-Anchored	0.2	0.3
Visit	Non-Anchored	2.8	4.2
Shop	Non-Anchored	6.0	9.0
Recreation	Non-Anchored	6.3	9.4
Serve Passenger	Non-Anchored	3.5	5.2
Other	Non-Anchored	4.0	6.0

6.1 Generation Problems

ActGen ran and produced logical results, but it also flagged numerous problems. Almost half of the households had at least one problem and 18 percent of them had household matching problems. Matching problems are not considered fatal (all activities will be generated), but they do point to a general inconsistency in the way households are grouped. Table 24 shows that the majority of matching problems related to work status followed by gender. One might argue that the types of trips made by males and females are not substantially different, but it is hard to see how the activity patterns of workers and non-workers can be interchangeable. Since the number of workers is a primary criterion in the household classification system, it was difficult to see how this many match problems could be generated.

Table 24: Distribution of Household Matching Problems

Problem	Percentage
Persons matched to a different age group	19.8%
Persons matched to a different work status	46.1%
Persons matched to a different gender	34.1%

Other categories of problems identified by ActGen are shown in Table 25. The majority of problems (59 percent) are vehicle passenger problems. This means that a trip with a travel mode of auto passenger was unable to identify a driver who was traveling from the specified origin to the specified destination at the specified time of day. This problem can be caused by coding errors in the activity survey or by household matching that could not assign the driver to a synthetic household member. ActGen corrects this problem by changing the travel mode to magic move. This means that the passenger will automatically be moved from their origin to their destination at the appropriate time of day. From the passenger's perspective, this has no impact on the simulation. If the driver is truly missing from the activity file, the impact is more significant.

Table 25: ActGen Problem Distribution

Problem	Percentage
Vehicle Access	3.8%
Activity Location	13.8%
Vehicle Passenger	58.8%
Activity Duration	23.6%

Activity duration problems are also not fatal. These problems identify situations where ActGen needed to compress the time allocated for a given activity in order to fit the travel time estimates into the tour constraints. In other words, the activity (work, school, shopping, recreation, etc.) is modeled as taking less time than was reported in the activity survey. For activities such as work or school the ActGen adjustment may be unacceptable and other types of corrective actions would be preferable. These problems are reported to provide the user with the data needed to implement other adjustment strategies.

The two problem types listed in Table 25 that are fatal are vehicle access and activity location problems. A vehicle access problem means that a car is not available at home when a driver needs to start their tour. This can be caused by coding problems in the activity survey or by duplicating a driver in the survey household to match an extra adult in the synthetic household. The result of this problem is that the whole tour is not included in the activity file.

Activity location problems mean that a location could not be found for the activity by the location choice model. This could mean that there is no accessibility (skim) to or from a given zone by the specified mode or the attraction weights for the zones or the activity locations are zero. It could also means that the location choice parameters are too restrictive for the location in question. Since 14 percent of the households had location problems, the source of the problems were worth investigating.

6.2 Problem Analysis

A total of 88 percent of all activities were generated by ActGen despite the fact that about 50 percent of all households had some type of problem with their tours. To investigate the source of these problems, a set of 20 problematic households encompassing all types of problems was randomly selected from the ActGen problem file. To support this analysis, the ActGen software was modified to generate detailed messages and data dumps of the location choice calculations.

6.2.1 Vehicle Passenger Problems

Vehicle passenger problems in ActGen generally result from shared rides. To illustrate, consider the example in Figure 41. In this case persons 2 and 3 in household 11 have a vehicle passenger problem (problem 28) for the 4th and 5th activities. Household 11 was matched to survey household 202992 and the household match report (Figure 42) shows that persons 2 and 3 for household 11 were both matched to person 3 in survey household 202992. This in itself is indicative of another type of problem because person 3 in the survey (a worker) is matched to two non-workers in the synthetic household. In fact the household type (46) assigned to the synthetic household is incorrect. Type 46 should be a suburban household with two workers and three persons. Since the synthetic household does not include two workers, there must be a problem with the household type script.

Figure 41: ActGen Problem File for Household 11

HHOLD	PERSON	ACTIVITY	MODE	PROBLEM	START	END	LOCATION	VEHICLE	SURVEY
11	2	4	2	28	18:00	21:00	7093	-2	202992
11	2	5	2	28	22:00	1@3:00	11807	-2	202992
11	3	4	2	28	18:00	21:00	7093	-2	202992
11	3	5	2	28	22:00	1@3:00	11807	-2	202992

Figure 42: Household Match Report for Household 11

	------------- Household					----------------	-------------- Survey					-----------------
Type	Household	Person	Age	Group	Worker	Gender	Household	Person	Age	Group	Worker	Gender
46	11	1	43	5	1	0	202992	1	48	5	1	2
46	11	2	42	5	0	1	202992	3	51	5	1	1
46	11	3	69	6	0	1	202992	3	51	5	1	1

Fixing the household type script is necessary and important, but even if it were correct, this household would still have a problem. The activity data for this household (Figure 43) indicates that person 1 is using auto (mode 2) to drive (vehicle 1) between activities 8, 9 and 10. Person 3 is also using auto (mode 2) as a passenger (vehicle 2) between activities 4, 5 and 6. Both tours have the same schedule and start and end at home, but the activity location is different (5 vs. 13) and the activity purpose is different (recreation vs. visiting). This appears to be a simple coding error in the activity survey that assigned a different location ID to the two people. Since the matching process used person 3 twice, the error was duplicated for synthetic persons 2 and 3.

Figure 43: Survey Activity Pattern for Household 202992

HHOLD	Person	Activity	Purpose	Start	End	Duration	Mode	Vehicle	Loc	Pass
202992	1	1	0 (HOME)	0.00	8.50	8.50	1	0	1	0
202992	1	2	14 (REC)	8.50	9.25	0.75	2	1	2	0
202992	1	3	0 (HOME)	9.33	10.75	1.42	2	1	1	0
202992	1	4	3 (VISIT)	10.83	11.25	0.42	2	1	3	0
202992	1	5	1 (WORK)	11.33	13.00	1.67	2	1	4	0
202992	1	6	0 (HOME)	13.08	13.58	0.50	2	1	1	0
202992	1	7	1 (WORK)	13.67	16.83	3.17	2	1	4	0
202992	1	8	0 (HOME)	16.92	17.00	0.08	2	1	1	0
202992	1	9	14 (REC)	18.00	21.00	3.00	2	1	5	2
202992	1	10	0 (HOME)	22.00	23.00	1.00	2	1	1	2
202992	1	11	0 (HOME)	23.00	27.00	4.00	1	0	1	0
202992	2	1	0 (HOME)	0.00	8.08	8.08	1	0	1	0
202992	2	2	8 (COL)	8.08	10.00	1.92	2	1	6	0
202992	2	3	8 (COL)	10.17	12.00	1.83	2	1	7	0
202992	2	4	0 (HOME)	12.08	12.50	0.42	2	1	1	0
202992	2	5	8 (COL)	12.58	15.00	2.42	2	1	8	0
202992	2	6	0 (HOME)	15.08	15.50	0.42	2	1	1	0
202992	2	7	2 (SHOP)	15.58	16.00	0.42	2	1	9	0
202992	2	8	8 (COL)	16.08	16.17	0.08	2	1	10	0
202992	2	9	4 (REC)	17.17	19.50	2.33	9	0	11	0
202992	2	10	4 (REC)	19.50	21.00	1.50	1	0	11	0
202992	2	11	5 (OTHER)	21.00	22.00	1.00	9	0	12	0
202992	2	12	0 (HOME)	22.08	23.00	0.92	2	1	1	0
202992	2	13	0 (HOME)	23.00	27.00	4.00	1	0	1	0
202992	3	1	0 (HOME)	0.00	8.33	8.33	1	0	1	0
202992	3	2	1 (WORK)	8.33	16.83	8.50	2	1	4	0
202992	3	3	0 (HOME)	16.92	17.00	0.08	2	2	1	2
202992	3	4	13 (VISIT)	18.00	21.00	3.00	2	2	13	2
202992	3	5	0 (HOME)	22.00	23.00	1.00	2	2	1	2
202992	3	6	0 (HOME)	23.00	27.00	4.00	1	0	1	0

As shown in Figure 44, ActGen replaces the shared ride with a magic move (mode 8) which has no significant impact on the tours for persons 2 and 3. It does, however, ignore the schedule coordination issues for person 1. Person 1 is still considered a two person carpool from a routing and simulation perspective, but the passenger trips are not impacted by person 1's travel time. One of the outcomes of this analysis was to spend more time cleaning the survey records. The household matching algorithm in ActGen was also modified to minimize the likelihood that a given person in the survey is selected multiple times while other household members are skipped.

Figure 44: Activity Records for Synthetic Household 11

HHOLD	PER	ACTIVITY	PURPOSE	START	END	DURATION	MODE	VEH	LOC	PASS
11	1	1	0	0:00	8:34:28	8:34:28	1	0	11807	0
11	1	2	14	8:34:44	9:19:44	0:45	2	16	15066	0
11	1	3	0	9:20	10:54:29	1:34:29	2	16	11807	0
11	1	4	3	10:54:40	11:19:40	0:25	2	16	6442	0
11	1	5	1	11:20	13:00	1:40	2	16	7949	0
11	1	6	0	13:00:24	13:39:33	0:39:09	2	16	11807	0
11	1	7	1	13:40	16:50	3:10	2	16	7949	0
11	1	8	0	16:50:28	17:59:42	1:09:14	2	16	11807	0
11	1	9	14	18:00	21:00	3:00	2	16	14840	2
11	1	10	0	22:00	1@3:00	5:00	2	16	11807	2
11	2	1	0	0:00	8:19:34	8:19:34	1	0	11807	0
11	2	2	1	8:20	16:50	8:30	2	15	7949	0
11	2	3	0	16:55	17:00	0:05	10	16	11807	2
11	2	4	13	18:00	21:00	3:00	8	0	7093	2
11	2	5	0	22:00	1@3:00	5:00	8	0	11807	2
11	3	1	0	0:00	8:19:34	8:19:34	1	0	11807	0
11	3	2	1	8:20	16:50	8:30	2	14	7949	0
11	3	3	0	16:55	17:00	0:05	10	16	11807	2
11	3	4	13	18:00	21:00	3:00	8	0	7093	2
11	3	5	0	22:00	1@3:00	5:00	8	0	11807	2

6.2.2 Location Problems

The analysis showed that the majority of location problems are due to walk-distance problems. This is caused by the 2,000 meter (1.25 mile) limit imposed on walk distances in the location choice script for one-way trips. Therefore, when an activity that is to be accessed using the walk mode needs to be located, the script is unable to find an activity location within the 2,000 meter limit. Moreover, the script was also unable to locate that activity in the same zone of the current activity location because the walk skim generated by PlanSum lacked intra-zonal entries. When intra-zonal entries were included most of the walk problems were resolved.

Another source of location problems was incomplete tours. The Portland travel survey has a significant number of tours that do not return home at the end of the day. This created situations where ActGen did not have a defined anchor at the end of the tour for locating intermediate activity locations. ActGen was modified to locate intermediate stops based on the anchor at the other end of the trip when the end-of-tour location is undefined.

A third source for location problems is input data inconsistency regarding school and college activities. Some zones with positive attraction weights for school and college activities had no activity locations with a positive weight. This prevented ActGen from locating school or college activities despite successfully selecting a zone location.

6.2.3 Activity Duration Problems

Activity duration problems were investigated next. Analysis showed that over 39 percent of activity duration problems were due to selecting primary anchor locations in far away zones that made it impossible to satisfy the time constraints associated with the activities. To address this issue, the location choice script was modified to limit the choice of destinations to those that fall within a round trip travel time of 2 hours (as determined from travel time skims). The new script did improve the proximity of activity locations (destinations are now closer to the home location), however, time duration problems did not disappear completely. Further investigation revealed that these problems were associated with multiple home-based tours. ActGen sets the start time of home activities based on the end time of the previous tour (or zero at the start of the day). If the second tour has time constraints that require the tour to leave home before the end time of the previous tour, a duration problem is generated. The model needs to consider rescheduling the first tour to accommodate the time constraints of the second tour.

6.2.4 Other Problems

Another issue was the presence of similar problematic activities for several households. The root of the problem was the limited number of survey households available to a given household type and the fact that several of these survey households had coding errors in their activity records. One of these survey households was replicated a hundred times which resulted in a hundred instances of the coding error in the regional population. In this case, if the survey record could not be corrected, it was deleted to avoid propagating the error throughout the process.

Addressing these four problem types increased the activity generation by 40 percent and increased the vehicle trips from 2.5 million to 3.6 million trips or 45 percent.

6.3 Location Choice Calibration

In the original GEN2 model, the zone-to-zone impedances were based on logsums from the METRO mode choice model. The plan was to ultimately replace or recalculate these logsums using travel time data generated by the Microsimulator. For this implementation, the location choice model was calibrated using zone-to-zone travel times by time of day from the Track 1 work rather than the logsums from the METRO model. No changes were made to the attraction weights prepared for the original model.

There are two parameters in the location choice model, β_n and $A_{n,j}$ that need to be calibrated. These parameters are normally mode-specific, but for this study only auto trips were considered. The calibration process adjusted the β_n vector to match the average trip distances from the travel survey. The process started by assuming $\beta_n = -1$ for all purposes. ActGen was run on a random sample of regional households and the trip distance report was analyzed for all trip purposes. If the resulting average trip distance was higher than the target, β_n was decreased by 25 percent, and vice versa. The process was repeated until the target trip lengths were achieved for each trip purpose.

6.3.1 Travel Time Sensitivity Tests

While this calibration strategy did in fact match average trip distances, the resulting β_n vector was much too low (in the order of -0.001) to have any sensitivity to the travel time differences of competing alternatives. The resulting trip distance distributions were not at all reasonable. The reason for the low β_n values was found to be related to using only travel time skims rather than logsums or composite skims in the logit model. The original zone-to-zone impedances included river crossing flags and distance costs in addition to travel times. Excluding these components from the logit function distorted the model sensitivities. To illustrate this point, consider the following example. Assume the model is selecting a work destination zone out of 5 candidates with equal attraction weights and the travel times listed in Table 26.

Table 26: Travel Times to Candidate Destinations

Destination Zone	*Travel Time (minutes)*
A	20
B	25
C	30
D	35
E	40

Assuming $\beta = -1$ and using Equation (1), zone A would have an almost 100 percent probability of being selected. Clearly, zone A is expected to be the most attractive, but it is also reasonable to expect some trips to select zone B and to a lesser extent zone C.

Table 27: Choice Probabilities with a Beta of -1.0

Destination Zone	*Choice Probability*
A	100.0
B	0.0
C	0.0
D	0.0
E	0.0

By using only travel time skims in a logit-based probability function, the closest destination will always dominate further destinations. Moreover, due to the nature of the logit function, only differences in travel times between the alternatives are important. For example one would get the same probabilities using the travel times shown in Table 28.

Table 28: New Travel Times to Candidate Destinations

Destination Zone	*Travel Times (minutes)*
A	80
B	85
C	90
D	95
E	100

Decreasing the time factor β_n vector does improve the probabilities for the remaining destinations, but at the expense of travel time sensitivity. For example, $\beta_n = -0.001$ would assume an almost equal spread among all options. One can calibrate average trip distances by having a relatively low β_n values, but the trip distance distributions will not be logical.

Table 29: Choice Distributions for Candidate Locations using Difference Beta Values

Destination Zone	*β = - 1.0*	*β = -0.5*	*β = - 0.25*	*β = - 0.1*	*β = - 0.01*	*β = - 0.001*
A	100.00	99.33	91.79	63.64	24.19	20.40
B	0.00	0.67	7.53	23.41	21.88	20.20
C	0.00	0.00	0.62	8.61	19.80	20.00
D	0.00	0.00	0.05	3.17	17.92	19.80
E	0.00	0.00	0.00	1.17	16.21	19.60

6.3.2 LogSum Impedance Tests

This issue was further investigated by replacing the travel time skims with the logsums calculations from the METRO mode choice model. These logsums are differentiated by trip type and by income group for work-related purposes. River crossing factors were added to the TRANSIMS travel time skims to enable the location choice script to calculate the logsums directly. At first, only the logsums pertaining to the auto and shared ride modes were calculated. Later it became apparent that the walk mode was frequently used (Table 30), so the logsums were modified to include walk and bike alternatives. The walk and bike travel times were estimated from straight-line distances between zones and an assumed average speed of 3 mph and 10 mph, respectively.

Table 30: Percent Mode Utilization by Area Type

Area Type	*Walk*	*Drive*	*Bus*	*Rail*	*Park & Ride*	*Bike*	*Magic*	*School Bus*
Very Urban	8.91	84.02	1.60	0.10	0.20	0.06	2.80	2.30
Urban	10.40	81.90	0.80	0.12	0.04	0.70	3.60	2.44
Suburban	10.33	81.40	0.90	0.15	0.44	0.73	3.49	2.57

Figure 45 shows a portion of the location choice script used to calculate logsums. It also includes logic to limit walks to 2.5 miles and round trip travel times to 120 minutes.

Figure 45: Excerpt from the Location Choice Script used in the First ActGen Implementation

```
#### DECLARATIONS ####
REAL cost, Time1, Time2, Time, Dist1, Dist2, Dist, DistCost  ENDDEF
REAL WalkTime, BikeTime, CBDWalkTime, NonCBDWalkTime     ENDDEF
INTEGER Delta, stat  ENDDEF

#### VARIABLE DEFINITIONS ####
Time1          = Skim2A.TIME/60             ### Travel time in minutes
Time2          = Skim2B.TIME/60             ### Travel time in minutes
Time           = Time1 + Time2              ### Total travel time in minutes
Dist1          = Tour.DISTANCE1/1609        ### Distance converted from meters to miles
Dist2          = Tour.DISTANCE2/1609        ### Distance converted from meters to miles
Dist           = Dist1 + Dist2              ### Total distance in miles
DistCost       = 0.091*(Dist1 + Dist2)      ### Total out of pocket cost $0.91/mile
WalkTime       = 60*Dist/3                  ### walk speed is assumed to be 3 miles/hr
BikeTime       = 60*Dist/10                 ### bike speed is assumed to be 10 mph
CBDWalkTime    = 5                          ### Walk time from parking to work in minutes
NonCBDWalkTime = 2                          ### Walk time from parking to work in minutes

#### Utility Calculations ####
IF(Time1 > 120 OR Time2 > 120)THEN
   RETURN (0)
ENDIF

IF(Tour.MODE == 1)THEN
   IF(Dist > 2.5)THEN
      RETURN(0)
   ELSE
      IF(Time > 120)THEN
         RETURN(0)
      ELSE
         Tour.UTILITY = Tour.UTILITY * EXP(Time)
         RETURN(1)
      ENDIF
   ENDIF
ENDIF

IF(Tour.PURPOSE == 1 OR Tour.PURPOSE == 11)THEN                 ### High Income Work trips
   cost = 0.079*Time + 0.097*CBDWalkTime + 0.35*DistCost        ### Drive alone
        + 3.37 + 0.079*Time + 0.097*CBDWalkTime + 0.35*DistCost  ### Drive with pass
        + 3.84 + 0.079*Time + 0.097*CBDWalkTime + 0.35*DistCost  ### Auto passenger
   IF(Dist < 10)THEN                                             ### Bike
      cost = cost+5.15 + 0.069*BikeTime
   ENDIF
   IF(Dist < 2.5)THEN                                            ### Walk
      cost = cost + 3.46 + 0.064*WalkTime
   ENDIF
   cost = cost + 1.054*(Skim2A.WAS_OR_FLG    + Skim2B.WAS_OR_FLG)
               + 1.366*(Skim2A.OR_WAS_FLG    + Skim2B.OR_WAS_FLG)
               + 0.370*(Skim2A.WEST_EAST_FLG + Skim2B.WEST_EAST_FLG)
               - 0.265*(Skim2A.EAST_WEST_FLG + Skim2B.EAST_WEST_FLG)
   Tour.UTILITY = Tour.UTILITY*EXP(-0.9*cost)
   stat = 1
ENDIF
RETURN(stat)
```

6.3.3 Calibrated Coefficients

Table 31 shows the calibrated β values by activity purpose for the GEN2 location choice model. The values calibrated for the original GEN2 model are also shown for comparison purposes. Notice that the calibrated coefficients are considerably different, but the "effective rescaling" coefficients are much closer. Effective rescaling normalizes the calibration coefficients based on the global correction needed to convert a trip-based destination choice model to a tour-based model. It also highlights differences in the way the Version 3.x activity generator located anchor activities versus the way the Version 4.0 software is implemented. ActGen 4.0 uses the round trip impedances where the Version 3.x software used only the outbound impedances. This algorithm change appears to have addressed the reason for rescaling the coefficients.

Table 31: Comparison of Calibrated Beta Values

Activity Purpose	*Tour Type*	*Version 3.x*		*Version 4.0 Coefficients*
		Calibrated Coefficients	*Effective Rescaling*	
Work High Income	Anchored	- 0.0368	- 1.0	- 0.9
Work Middle Income	Anchored	- 0.0368	- 1.0	- 0.9
Work Low Income	Anchored	- 0.0368	- 1.0	- 0.9
Visit	Anchored	- 0.0345	- 0.9375	- 1.0
Shop	Anchored	- 0.0345	- 0.9375	- 1.0
Recreation	Anchored	- 0.0345	- 0.9375	- 1.0
Serve Passenger	Anchored	- 0.0345	- 0.9375	- 1.0
Other	Anchored	- 0.0345	- 0.9375	- 1.0
College	Anchored	- 0.0368	- 1.0	- 2.0
School	Anchored	- 0.0368	- 1.0	- 0.5
Work High Income	Non-Anchored	- 0.0368	- 1.0	- 0.9
Work Middle Income	Non-Anchored	- 0.0368	- 1.0	- 0.9
Work Low Income	Non-Anchored	- 0.0368	- 1.0	- 0.9
Visit	Non-Anchored	- 0.0040	- 0.1087	- 0.5
Shop	Non-Anchored	- 0.0040	- 0.1087	- 0.5
Recreation	Non-Anchored	- 0.0040	- 0.1087	- 0.75
Serve Passenger	Non-Anchored	- 0.0040	- 0.1087	- 0.5
Other	Non-Anchored	- 0.0040	- 0.1087	- 0.75

6.4 Trip Length Results

The average trip length targets for validation were based on trip and tour leg types obtained from the METRO survey. METRO lumped all activities other than home, work, school, and college into a general discretionary category. This means that the tour leg targets for recreation, visit, serve passenger, shop, and other activity purposes are all based on the discretionary (D) average. No grouping was necessary for trip length targets by trip purpose since METRO used a wider range of activity purposes.

Figure 46 presents the average trip length by trip purpose. The results show a good agreement with METRO's results, falling within 0.4 miles of target distances, except for non-home-non-work and college, where the difference is greater than 0.6 miles. In general, the average length for all trips is within four percent of the target. Figure 47 presents the average length by tour leg type. The results show a good agreement with METRO's results, falling within 0.4 miles of target distances, except for work-work, work-discretionary, and discretionary-work leg types, where the difference is greater than a mile. The work-work type is at least better than the Version 3.x value.

Figure 46: Average Trip Length by Trip Purpose

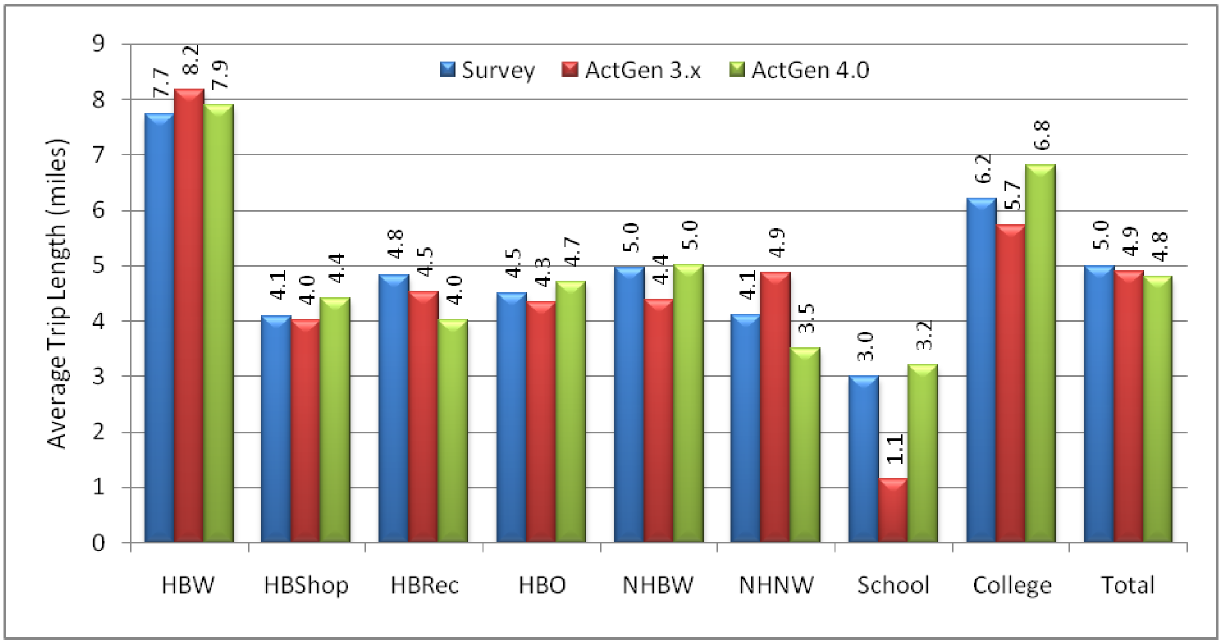
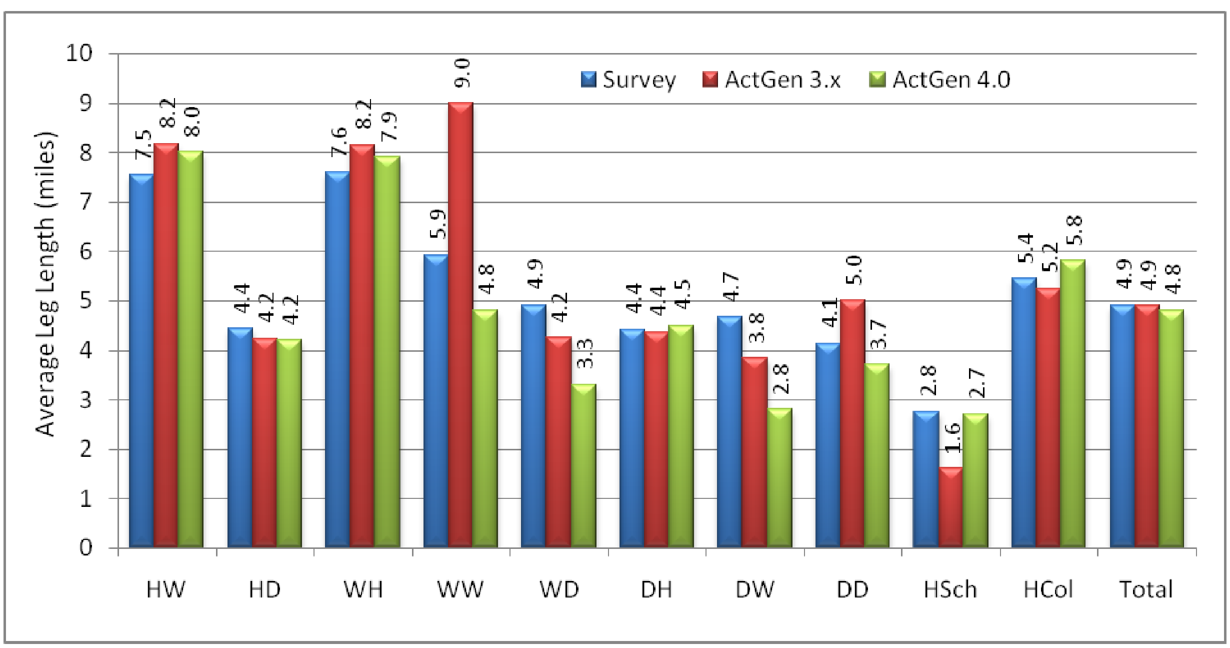


Figure 47: Average Leg Length by Tour Leg Type



6.5 Router-Simulation Results

The output of the activity generation process is an activity file that specifies the time of day each activity starts and ends and the travel mode and vehicle that should be used for travel to the activity. A trip is therefore defined as starting at the time of day when the previous activity ends and ending at the time of day when the current activity starts. It starts at the previous activity location and ends at the current activity location. If a vehicle is specified, the vehicle must be parked at or near the previous activity location.

The activity file is then read by the Router to build a path or travel plan between each activity. If a path between the two activity locations cannot be constructed in the allotted time at the specified time of day, a problem message is generated and repairs are attempted or the remaining activities planned for the traveler are skipped. Problematic activities may be addressed through feedback to the activity generator or by adjustments to the activity schedule based on more recent travel time estimates.

The routed activities are merged with non-household-related vehicle traffic for input to the Microsimulator. The Microsimulator tracks the vehicle movements and interactions on a second-by-second basis and summarizes the resulting network performance every 15 minutes. Link travel times and turning movement delays are fed back to the Router to reconsider and adjust the travel paths and trip times. Only a fraction of the trips are adjusted at any given iteration in order to permit the system to respond gradually to new travel conditions rather than simulate major changes that can easily result in unreasonable queues and excessive congestion. Once the network performance stabilizes, zone-to-zone travel times can be summarized for feedback to activity generation.

6.5.1 Itinerant Traffic

The METRO regional model includes approximately 223,000 non-household-related vehicle trips per day. These are internal truck and commercial vehicle trips and external trips that 1) originate outside the region and end inside the region, 2) originate inside the region and end outside the region, or 3) travel through the region without stopping. All of these trips need to be added to the activities generated by synthetic households in order to simulate the full vehicle load on the roadway network.

The non-household trips are traditionally called “itinerant” trips in a TRANSIMS model. Trip tables from the METRO model are converted to TRANSIMS format using the ConvertTrips process developed for Track 1. This process assigns each itinerant trip a unique household and vehicle ID in order to manage and model the trip as if it were a standard household-related trip. When the itinerant households are added to the synthetic households, the Portland GEN2 model simulates 852,000 households or 1.8 million travelers.

6.5.2 Partitioning and Router Buildup

In order to process and manage the 852,000 households more efficiently, each household was randomly assigned to one of six partitions. In this case a partition is simply a list of the household IDs stored in a file with a sequential partition extension AA, AB, AC, AD, AE, or AF. The HHList program was used to distribute the household IDs to the partition lists. The Router and plan processing tools can then use the household list file to identify the records in the activity or trip files that should be processed during a given run or iteration. This distributes the computer processing load over multiple computers or CPUs and thereby improves the overall processing time.

An incremental loading approach was used to initially distribute the regional trips to travel paths. The household list for each partition was randomly assigned to 10 subsets for Router buildup purposes. The Router for each partition builds paths for the first 10 percent of the households and then combines the volumes from the six partitions into a single link volume that is used to estimate link travel times using traditional volume-delay functions. The travel times are then used to build paths for the next 10 percent of the households until all trips are loaded.

Once all of the trips are loaded, several Router feedback iterations are used to help refine the travel plans and address network bottlenecks. Travelers with routing problems or who travel through links with high volume-capacity ratios are given an opportunity to select a new path based on the most recent link travel times. At this point the results are reviewed to determine what needs to be done to address major inconsistencies between the travel plans and the estimated network performance.

6.5.3 Initial Router Results

On average, the Router had problems with about 14 percent of the vehicle trips. As the sample output in Figure 48 demonstrates, the number of problems generated even under free-flow conditions was significant. These trips were mostly vehicle access problems (48 percent) in which a path building problem for an earlier trip meant that the vehicle was not parked in the parking lot where a subsequent trip expected to find it. The time schedule problems (29 percent) and walk distance problems (10 percent) suggest discrepancies between the travel times used for activity generation and the path options between the specified activity locations.

Figure 48: Router Problems from the First 10 Percent Loading of Partition AA

```
Number of Vehicle Trips Saved = 56346
Number of Output Plans = 261345
Number of Output Records = 317691
Number of Output Travelers = 27660
Number of Output Trips = 56346

Total Number of Problems = 7943

Number of Time Schedule Problems = 2286 (28.8%)
Number of Zero Node Problems = 856 (10.8%)
Number of Path Circuity Problems = 100 (1.3%)
Number of Vehicle Access Problems = 3788 (47.7%)
Number of Walk Distance Problems = 787 (9.9%)
Number of Walk Access Problems = 121 (1.5%)
Number of Path Size Problems = 1 (0.0%)
Number of Access Restriction Problems = 4 (0.1%)
```

In total, the Router generated travel plans for 3.6 million vehicle trips as compared to the 5.0 million vehicles loaded to the network in Track 1. If all of the Router problems were resolved, this would still mean that the activity generator under estimated vehicle travel by almost 20 percent. This is the general level of under reporting METRO identified in the household survey and attempted to correct through selection bias factors in the original GEN2 model. Analysis was pursued to ensure that the factors developed for the Version 3.x software were properly translated and applied in the Version 4.0 software.

Despite the apparent under reporting of total vehicle trips, the initial validation checks using the routed volumes revealed additional concerns. The results (Figure 49) show that the Router over estimated the link volumes by 29 percent with a root mean squared error (RMSE) of 107 percent. Considering that the Router produces 1.4 million less vehicle trips than Track 1, this clearly indicates that the paths are sub-optimal or the trips are significantly longer or distributed very differently. Since the freeways and expressways are over estimated by 45 percent and 63 percent, respectively, there is some chance that additional feedback iterations will improve the results. The screenline results (Figure 50), however, clearly demonstrate that the problems are not purely path-based. The Columbia River crossings are grossly over-estimated and there is little chance path refinements can change this fact. This must be addressed through location choice.

Figure 49: Initial Router Validation Results by Facility Type

Facility Type	Num. Obs.	Volume Estimate	Volume Observed	Difference Volume	Difference %	Abs.Error Avg.	Abs.Error %	Std. Dev.	% RMSE	R Sq.
Freeway	51	4188117	2874828	1313289	45.7	28474	50.5	24095	65.9	0.118
Expressway	24	670521	409706	260815	63.7	11393	66.7	10042	88.1	0.307
Major Arterial	104	1311884	1143236	168648	14.8	6246	56.8	6078	79.1	0.163
Minor Arterial	82	553597	507048	46549	9.2	4765	77.1	4404	104.6	0.078
Collector	130	236317	462788	-226471	-48.9	2260	63.5	1699	79.3	0.064
Other	18	390512	286236	104276	36.4	8859	55.7	7146	70.8	0.450
TOTAL	409	7350948	5683842	1667106	29.3	7871	56.6	12630	107.0	0.806

Figure 50: Initial Router Validation Results by Screenline

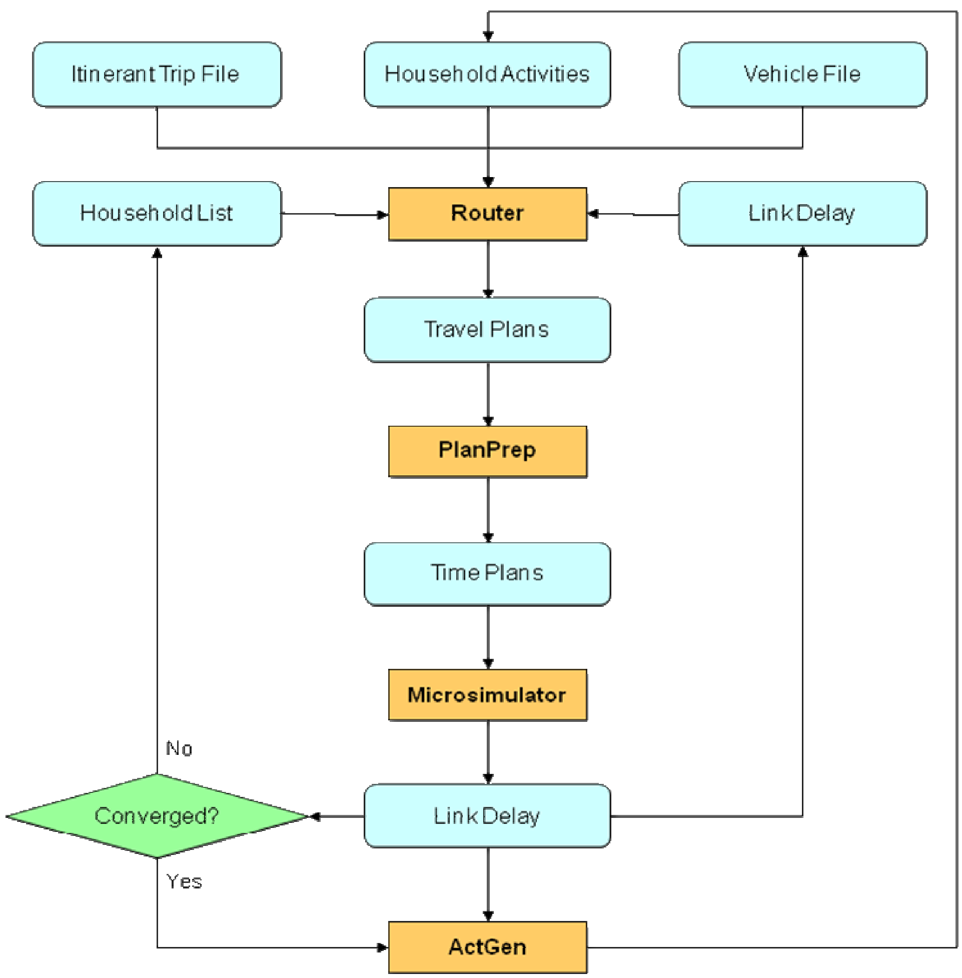
Link Group	Num. Obs.	Volume Estimate	Volume Observed	Difference Volume	Difference %	Abs.Error Avg.	Abs.Error %	Std. Dev.	% RMSE	R Sq.
R-1 St. John's Br	2	51937	22610	29327	129.7	14664	129.7	8683	140.6	0.000
R-2, William. River Bridg	16	636377	464966	171411	36.9	14411	49.6	15940	72.7	0.830
R-3, Sellwood Bridge	2	33972	31382	2590	8.3	6234	39.7	1831	40.6	0.000
R-4 OrCty 205 Br	4	175864	109571	66293	60.5	16573	60.5	10509	69.0	0.995
R-5, I-5 Bridge	2	264341	113838	150503	132.2	75252	132.2	23682	135.4	0.000
R-6 I-5 Hayden	2	266556	128300	138256	107.8	69128	107.8	23664	110.9	0.000
R-7 I205	2	224825	104074	120751	116.0	60376	116.0	20354	119.3	0.000
E-9 W of 122nd	34	301790	348370	-46580	-13.4	3933	38.4	3749	52.7	0.874
E-21 W of Tacoma	20	249267	272032	-22765	-8.4	2442	18.0	1580	21.2	0.984
W-7 Westhills	10	263987	172226	91761	53.3	10529	61.1	15220	103.8	0.925
W-09 Tigard Int	6	179337	154962	24375	15.7	7068	27.4	7137	37.2	0.990
W-16 NW of Tigard	10	212395	140922	71473	50.7	9514	67.5	16069	127.5	0.965
TOTAL	110	2860648	2063253	797395	38.6	10669	56.9	17784	110.2	0.852

6.5.4 Microsimulator Iterations

Given the large discrepancies in the number of vehicle trips and the travel patterns highlighted by the Router analysis, the likelihood that the Microsimulator could improve the results or even provide meaningful information to feed back to the activity generator was quite small. Nevertheless, a Router-Microsimulator convergence process was developed to establish procedures and evaluate the robustness of the Microsimulator under extremely congested conditions.

Figure 51 presents the high level structure of the Router-Microsimulator feedback process developed by the Track 1 effort. The Router builds paths for a subset of households based on the link delays from the previous simulation. These plans are merged with the full regional plan set, sorted by start time, and simulated. New link travel times are generated, the path travel times are compared to the previous estimates, and a decision is made to continue or exit the process.

Figure 51: Router-Microsimulator Convergence Process



The problems generated by the first Microsimulator application are shown in Figure 52. Of the 3.6 million trips that were routed, 3.4 million trips were loaded to the network and 2.85 million or 79 percent completed their trip. Departure time and vehicle access problems were the primary reasons vehicles were not loaded to the network. Time schedule, arrival time and wait time problems are the primary reasons vehicles do not complete their trips. Arrival time and wait time problems are primarily congestion related. Time schedule problems most likely point to incompatibilities between the travel times used by the Router and/or ActGen and the simulated times.

Figure 52: Initial Microsimulator Problems

```
Number of Vehicle Trips Processed = 3608163
Number of Vehicle Trips Started = 3461411
Number of Vehicle Trips Completed = 2854701

Number of Travelers with Problems = 383090 (34.6%)

Total Number of Problems = 1183894
Number of Time Schedule Problems = 441315 (37.3%)
Number of Vehicle Access Problems = 117582 (9.9%)
Number of Wait Time Problems = 289909 (24.5%)
Number of Departure Time Problems = 28711 (2.4%)
Number of Arrival Time Problems = 305206 (25.8%)
Number of Link Access Problems = 18 (0.0%)
Number of Vehicle Spacing Problems = 564 (0.0%)
Number of Access Restriction Problems = 589 (0.0%)
```

After 20 iterations of the Router-Microsimulator process using various methods of selecting households to update, the Microsimulator problems (Figure 53) were substantially reduced, but still significant. Slightly greater than 3 million trips (compared to 2.85 million) were now completed. Time schedule problems remained the most predominant problem at 43 percent, followed by wait time problems at 31 percent. The overall number of problems, however, was reduced from 1,180,000 to 510,000 and the number of households with problems was reduced from 34.6 percent to 20.7 percent.

Figure 53: Microsimulator Problems after 20 Feedback Iterations

```
Number of Vehicle Trips Processed = 3608163
Number of Vehicle Trips Started = 3461411
Number of Vehicle Trips Completed = 3088722

Number of Travelers with Problems = 226870 (20.7%)

Total Number of Problems = 510384
Number of Time Schedule Problems = 219798 (43.1%)
Number of Vehicle Access Problems = 55667 (10.9%)
Number of Wait Time Problems = 160007 (31.4%)
Number of Departure Time Problems = 13256 (2.6%)
Number of Arrival Time Problems = 61374 (12.0%)
Number of Link Access Problems = 7 (0.0%)
Number of Vehicle Spacing Problems = 186 (0.0%)
Number of Access Restriction Problems = 89 (0.0%)
```

Validation of the Microsimulator results (Figure 54) showed that the volumes were over estimated by 12 percent with an RMSE of 72 percent. The freeway and expressway volumes were 19 percent and 34 percent high. The screenline summaries (Figure 55) continued to show that the Columbia River bridges were oversubscribed. All of these values are markedly better than the initial Router validation statistics presented above, but this is largely due to the fact that the Microsimulator was unable to complete 400,000 vehicle trips. Since the distribution of errors by facility type and screenline are generally consistent, the issues and problems are uniformly distributed throughout the region and not localized to any particular area. The only noticeable exception to this rule is the I-205 Columbia River Bridge which appears to have somewhat more reserve capacity than the rest of the region.

Figure 54: Facility Type Validation after 20 Feedback Iterations

Facility Type	Num. Obs.	------Volume------ Estimate	Observed	---Difference--- Volume	%	--Abs.Error-- Avg.	%	Std. Dev.	% RMSE	R Sq.
Freeway	51	3427500	2874828	552672	19.2	18211	32.3	16831	43.8	0.144
Expressway	24	547895	409706	138189	33.7	8260	48.4	4792	55.6	0.364
Major Arterial	104	1227495	1143236	84259	7.4	4488	40.8	4058	54.9	0.235
Minor Arterial	82	536747	507048	29699	5.9	3864	62.5	3377	82.8	0.115
Collector	130	281804	462788	-180984	-39.1	2117	59.5	1536	73.4	0.100
Other	18	348213	286236	61977	21.7	5660	35.6	3306	40.9	0.629
TOTAL	409	6369654	5683842	685812	12.1	5593	40.2	8290	71.9	0.821

Figure 55: Screenline Validation after 20 Feedback Iterations

Link Group	Num. Obs.	------Volume------ Estimate	Observed	---Difference--- Volume	%	--Abs.Error-- Avg.	%	Std. Dev.	% RMSE	R Sq.
R-1 St. John's Br	2	37989	22610	15379	68.0	7690	68.0	2514	69.8	0.000
R-2, William. River Bridg	16	554190	464966	89224	19.2	7320	25.2	11321	45.4	0.859
R-3, Sellwood Bridge	2	29839	31382	-1543	-4.9	6056	38.6	1091	38.9	0.000
R-4 OrCty 205 Br	4	151060	109571	41489	37.9	10372	37.9	7002	43.9	0.949
R-5, I-5 Bridge	2	215961	113838	102123	89.7	51062	89.7	22572	94.0	0.000
R-6 I-5 Hayden	2	217997	128300	89697	69.9	44848	69.9	22509	74.2	0.000
R-7 I205	2	199467	104074	95393	91.7	47696	91.7	15062	93.9	0.000
E-9 W of 122nd	34	294001	348370	-54369	-15.6	3137	30.6	3101	42.7	0.860
E-21 W of Tacoma	20	226988	272032	-45044	-16.6	3577	26.3	3747	37.6	0.963
W-7 Westhills	10	231489	172226	59263	34.4	6530	37.9	10259	68.1	0.925
W-09 Tigard Int	6	159519	154962	4557	2.9	2716	10.5	2824	14.5	0.983
W-16 NW of Tigard	10	180637	140922	39715	28.2	5974	42.4	8654	72.0	0.979
TOTAL	110	2499137	2063253	435884	21.1	7208	38.4	12317	75.8	0.858

Chapter 7 Second Actgen Implementation

It was clear from the first ActGen implementation and integration with the Router and Microsimulator that there was a major disconnect between the vehicle trips and travel patterns generated by ActGen and the traffic counts observed in the real world. The fact that there were fewer vehicles generating significantly move volume on the network means the trip lengths were way too long. Since the average trip lengths validated reasonably well, there must be significant differences in the trip length distributions or errors in the trip length calculations.

7.1 Track 1 Trip Length Distributions

Trip length distribution comparisons were made as part of the original GEN2 model calibration for home-based work, home-based other, and non-home-based trips. In each case the distributions were similar to the activity generator results though in all cases the survey included greater percentages of trips of three miles or less.

In an effort to generate this level of data for other trip purposes and tour leg types, an attempt was made to estimate trip lengths from the X-Y coordinates recorded in the travel survey. Trip lengths were estimated based on the straight-line distance between the X-Y coordinates and compared to the travel time estimated by subtracting the start time of the current activity from the end time of the previous activity. Analysis showed that the time distribution was not continuous with most values rounded up to the nearest multiple of 5 minutes. Moreover, the distance and time values were not very consistent. Around 11 percent of the activity records indicated a zero-distance with a positive time and vice-versa. Still, even when discounting such records, the results were of little value because the survey expansion factors were lost. The household weights posted on the survey records were replaced by bias factors for selecting households within a household type. As such they could no longer be used to expand the survey trip lengths to the full population.

The alternative was to extract trip length distributions from the Track 1 trip tables. The TripSum program was used to generate trip length and travel time reports for HBW, NHBW, NHNW, HBO, School, and College trips. Figure 56 presents the distributions for each trip type. Table 32 compares the average trip lengths from the Track 1 data with the value calculated from the travel survey. On average the survey trip lengths are 14 percent higher than the Track 1 estimates which complicates the calibration. It was decided to place a premium on matching the “shape” of the distribution curve rather than the distribution itself.

Figure 56: Track 1 Trip Length Distributions by Trip Purpose

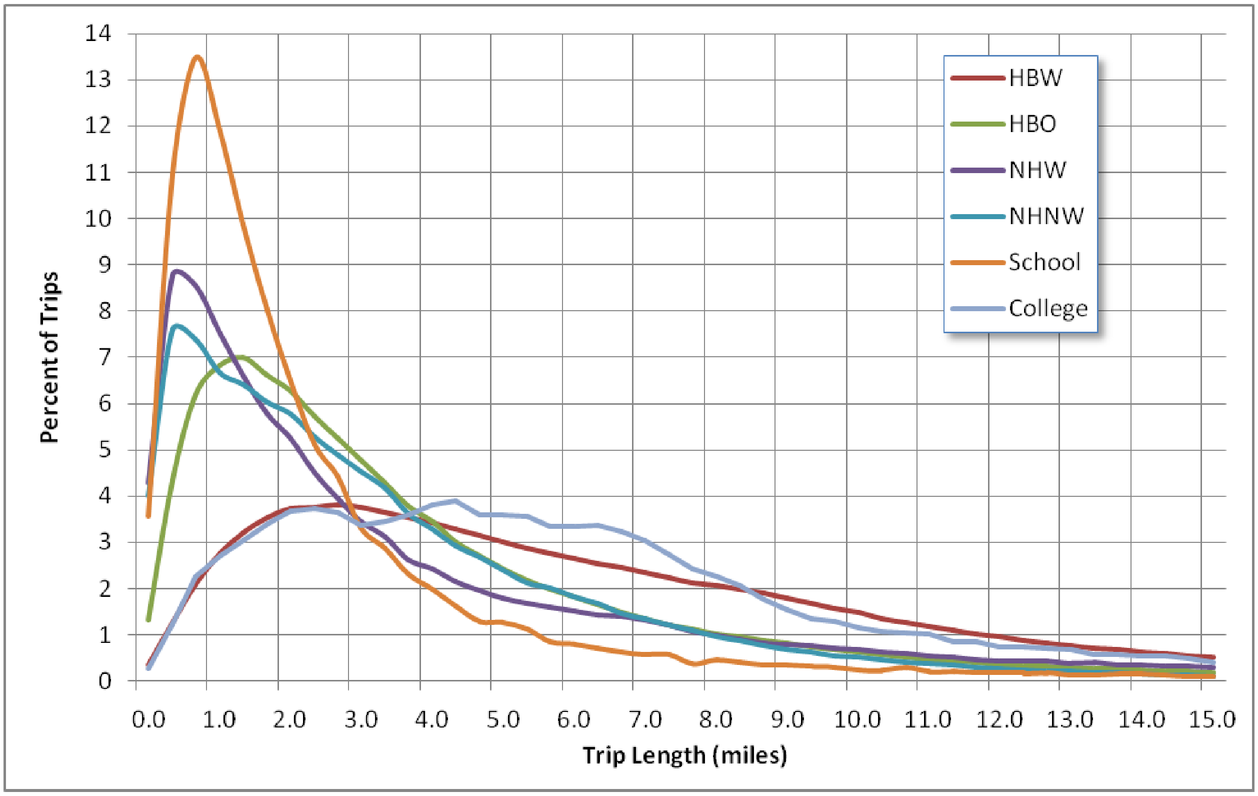


Table 32: Track 1 Average Trip Lengths by Trip Purpose

Trip Purpose	*Track 1 Average Trip Length (miles)*	*Survey Average Trip Length (miles)*
HBW	6.3	7.7
HBO	4.0	N/A
NHW	4.0	5.0
NHNW	3.4	4.1
School	2.5	3.0
College	5.7	6.2
Total	**4.4**	**5.0**

7.1 Time Budgets

It was evident from the time schedule problems in the Router and Microsimulator that the locations selected by the first ActGen implementation could not be reached within the allotted time. The second ActGen implementation attempted to address this problem by adding time-budget constraints to the location choice model. Time budgets measure time available for travel based on the activity schedules defined in the household survey.

Time-budget constraints limit the choice of zones available for selection as a destination by the location choice model. The difference between the start time of the current activity and the end time of the previous activity defines the overall time available for travel. A lower and upper bound is applied to the expected travel time to define the range of travel times that will be considered. If a candidate destination zone does not fall within the specified time range, the zone is ignored. This makes the destination that is selected compatible with the activity schedule and less likely to generated time schedule problems in ActGen or the Router-Microsimulator process. Time budget constraints do increase the number of activity location problems, but this may be rectified to a certain extent by generating more accurate inter-zonal and intra-zonal travel time skims.

For locating anchor activities, the time budget can be calculated as follows:

$$\Gamma_{\text{home,anchor}} = ST(\text{anchor}) - ET(\text{home}) - \sum_{k \in \text{home} \rightarrow \text{anchor}} AD(k) \tag{5}$$

For locating intermediate stops given the home and anchor locations, two time budgets are needed depending on whether the intermediate stop is in the first or second half of the tour. For locating activities in the first half of the tour, the outbound travel time budget is

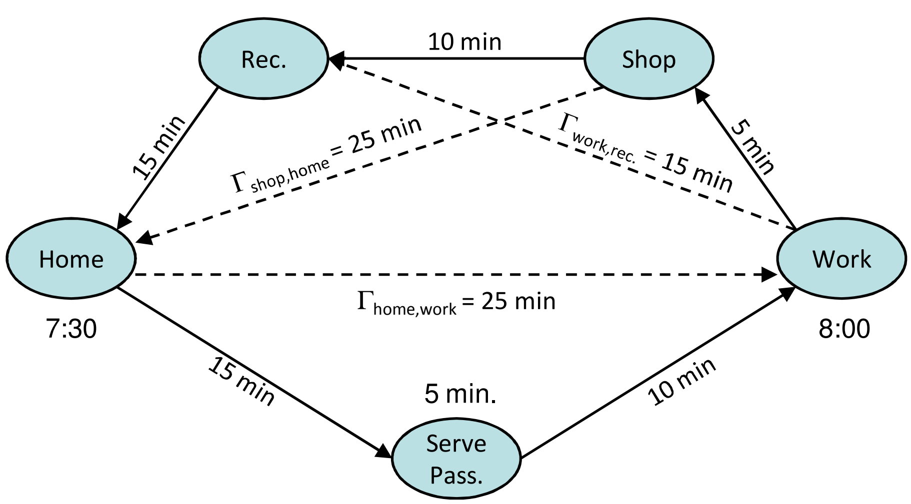
$$\Gamma_{\text{home},j} = ST(j) - ET(\text{home}) - \sum_{k \in \text{home} \rightarrow j} AD(k) \tag{6}$$

and the inbound travel time budget is

$$\Gamma_{j,\text{anchor}} = ST(\text{anchor}) - ET(j) - \sum_{k \in j \rightarrow \text{anchor}} AD(k) \tag{7}$$

To illustrate this concept, consider the location of work given the travel pattern shown in Figure 57.

Figure 57: Time Budget Example



The pattern indicates that the available time allocated for travel to work from home using Equation (5) is 25 minutes (8:00 – 7:30 – 5 minutes) = 25 minutes. Similarly, the available travel time to recreation from work is calculated using Equation (7) and is 5+10 = 15 min. The remaining time budgets are simply the travel times between these activities.

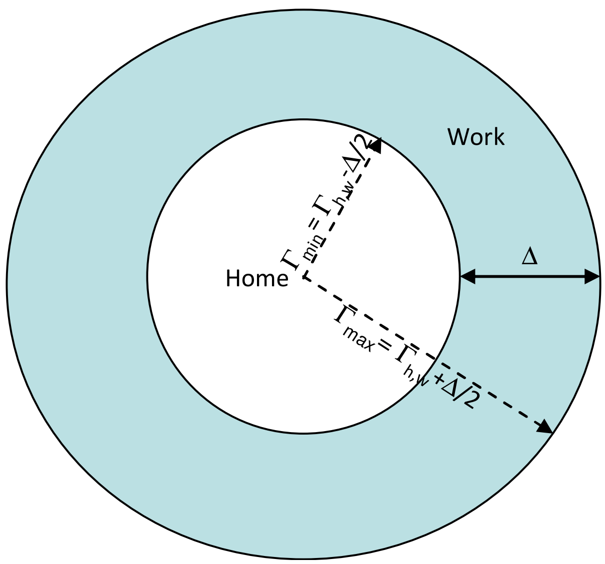
However, due to the use of travel time skims and not point-to-point travel times, it is close to impossible to pin-point a work location exactly 25 minutes away from the home location. Therefore, a time window, Δ, is used around the calculated time budgets to form a time range or "donut" that encompasses eligible destination zones.

$$\Gamma_{\text{home,anchor}}^{\min} = \Gamma_{\text{home,anchor}} - \Delta / 2 \tag{8}$$

$$\Gamma_{\text{home,anchor}}^{\max} = \Gamma_{\text{home,anchor}} + \Delta / 2 \tag{9}$$

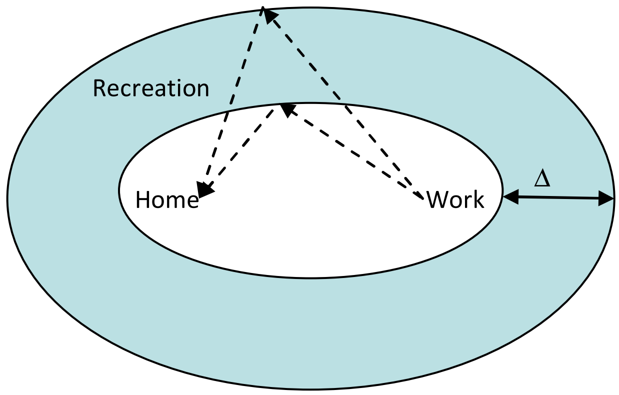
The location choice model will therefore disregards locations that do not fit within the specified range of these two time constraints.

Figure 58: Time Buffer for an Anchor Destination



Locating intermediate stops is treated a bit differently. The buffer is not applied to the inbound and outbound time budgets individually, but rather to their sum. This preserves the VHT within the network and protects against extreme situations. Such a method simply creates an elliptical donut due to the "rubber banding" effect.

Figure 59: Time Buffer for an Intermediate Destination



$$\Gamma^{\min}_{\text{work,rec,home}} = \Gamma_{\text{work,rec}} + \Gamma_{\text{rec,home}} - \Delta/2 \tag{10}$$

$$\Gamma^{\max}_{\text{work,rec,home}} = \Gamma_{\text{work,rec}} + \Gamma_{\text{rec,home}} + \Delta/2 \tag{11}$$

The location choice model will therefore disregard locations that do not fit within the specified range of these two time constraints. The location choice model has been updated to incorporate time-budget constraints:

$$p_{(\text{anchor},j)|(\text{home},i)} = \frac{A_{\text{anchor},j} \times \exp\left[\beta_{\text{anchor}}\left(C_{ij}(ET_{\text{home}}) + C_{ji}(ET_{\text{anchor}})\right)\right]}{\sum_k A_{\text{anchor},k} \times \exp\left[\beta_{\text{anchor}}\left(C_{ik}(ET_{\text{home}}) + C_{ki}(ET_{\text{anchor}})\right)\right]}$$
$$\text{such that:} \begin{cases} T_{ij}(ET_{\text{home}}) \le \Gamma^{\max}_{\text{anchor,home}} \\ T_{ij}(ET_{\text{home}}) \ge \Gamma^{\min}_{\text{anchor,home}} \end{cases} \tag{12}$$

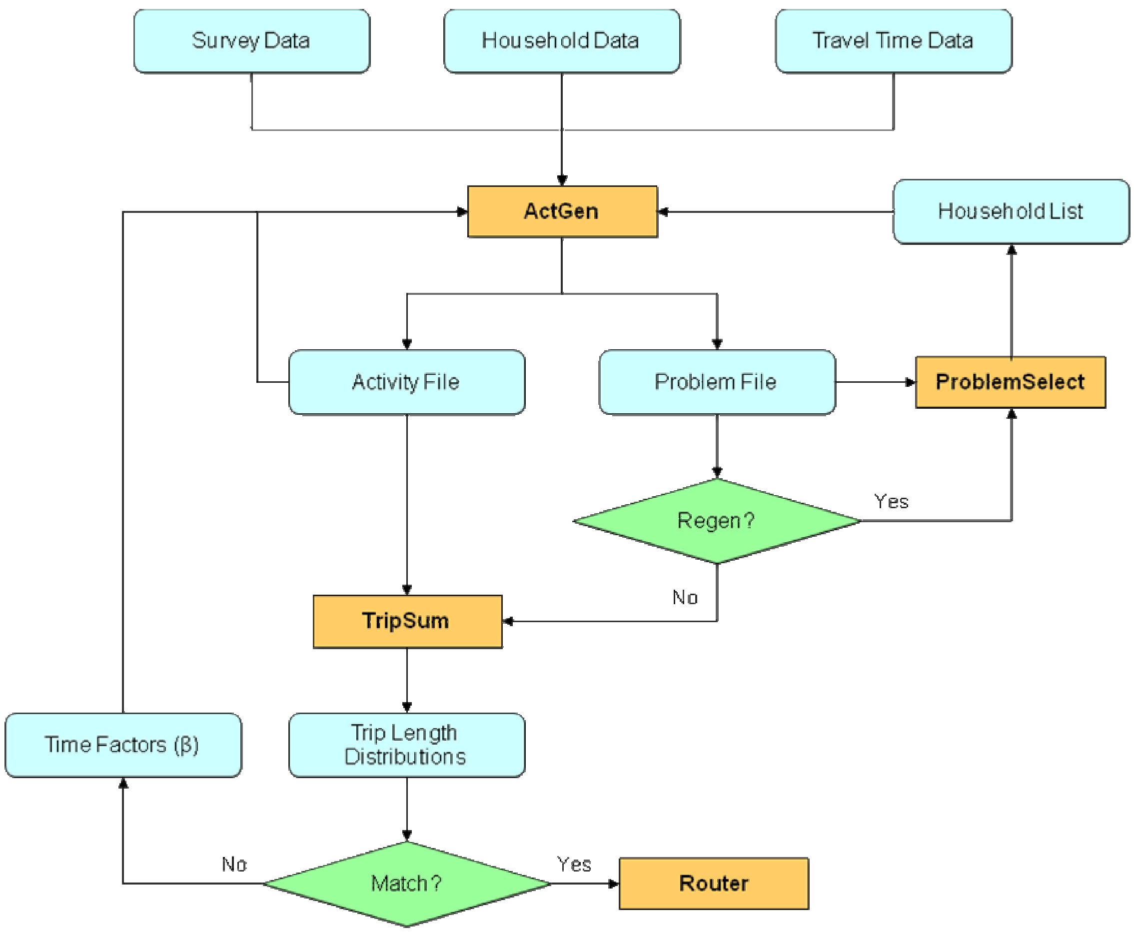
Where $T_{ij}(ET_{\text{home}})$ is the travel time skim between zones i and j evaluated at ET_{home}. Similarly, for intermediate zones,

$$p_{(n,z)|(\text{home},i),(\text{nearest},j)} = \frac{A_{n,z} \times \exp\left[\beta_n\left(C_{iz}(ET_{\text{home}}) + C_{zj}(ET_n)\right)\right]}{\sum_k A_{n,k} \times \exp\left[\beta_n\left(C_{ik}(ET_{\text{home}}) + C_{kj}(ET_n)\right)\right]}$$
$$\text{such that:} \begin{cases} T_{ik}(ET_{\text{home}}) + T_{kj}(ET_n) \le \Gamma_{\text{work,rec}} + \Gamma_{\text{rec,home}} + \Delta/2 \\ T_{ik}(ET_{\text{home}}) + T_{kj}(ET_n) \ge \Gamma_{\text{work,rec}} + \Gamma_{\text{rec,home}} - \Delta/2 \end{cases} \tag{13}$$

7.3 Activity Regeneration and Calibration

It was also evident from the previous ActGen implementation effort, that ActGen produced far too few vehicle trips compared to Track 1, 1.4 million to be exact. This is partially due to the fact that ActGen only accommodated 88 percent of the household activities. Activity regeneration was implemented to improve the overall generation rate. The basic idea is to have households with problematic tours matched to a different survey household in the hope that a different set of activities would be a better fit, spatially and temporally. The ProblemSelect program creates a list of households for regeneration and then ActGen is re-run with the household list and a new random number seed. The process is repeated until most of the problems are resolved. In this case, the stopping criteria were 95 percent of the activities generated or a maximum of five iterations. Afterwards, the average trips lengths were calibrated by adjusting the time factors (β). This process is depicted in Figure 60.

Figure 60: Activity Model Calibration Framework



The calibration effort itself proved to be more complicated than anticipated due to the use of time-budget constraints, which adds another degree of freedom in calibrating the model, in addition to the time factors. To this end, a bi-level approach was used for this calibration effort, whereby time-budget constraints were applied in the first level, and time factors were calibrated in the second level.

There are two parameters to be calibrated for this exercise. The first is the time buffer Δ associated with time budgets, and the time factors β associated with the location choice model. It was assumed that Δ would be applicable across the board whereas β would be purpose-specific. The calibration process started with an initial guess for the values for Δ and β. ActGen was then executed and distances were checked against calibration targets. These parameters were then alternately adjusted to avoid over-correction pitfalls. The β was corrected based on average trip lengths and Δ was corrected based on the trip length distributions. If the resulting average trip length was shorter than calibration targets, β was increased and vice-versa. If the resulting distribution was too-tight, then Δ was increased, and vice-versa. The process was iterated until a reasonable solution was achieved.

For this application, the calibration values for β from the first ActGen implementation were assumed for the initial solution. A ten minute buffer was used as an initial value for Δ. After seven calibration iterations, the coefficients shown in Table 33 were estimated. These values are significantly lower than the parameters calibrated for the first ActGen implementation. This is a direct result of the time budget constraints which control the maximum and minimum range of trip lengths that are eligible for selection. The lower coefficients imply that locations with a broader range of trip lengths, but still within the time budget, have a higher probability of being selected.

Table 33: Calibrated Coefficients for the Second ActGen Implementation

Activity Purpose	*Tour Type*	β	Δ *(min)*
Work High Income	Anchored	-0.20	5
Work Middle Income	Anchored	-0.20	5
Work Low Income	Anchored	-0.20	5
Visit	Anchored	-0.20	5
Shop	Anchored	-0.28	5
Recreation	Anchored	-0.40	5
Serve Passenger	Anchored	-0.30	5
Other	Anchored	-0.30	5
College	Anchored	-0.09	5
School	Anchored	-0.13	5
Work High Income	Non-Anchored	-0.90	5
Work Middle Income	Non-Anchored	-0.90	5
Work Low Income	Non-Anchored	-0.90	5
Visit	Non-Anchored	-0.20	5
Shop	Non-Anchored	-0.28	5
Recreation	Non-Anchored	-0.40	5
Serve Passenger	Non-Anchored	-0.30	5
Other	Non-Anchored	-0.30	5

7.3.1 Trip Length Distribution

The coefficient calibration process matched the average trips lengths by trip purpose, but also attempted to match the shapes of trip length distributions derived from the Track 1 trip tables. The resulting trip length distributions for several trip purposes are shown in Figure 61, Figure 62, Figure 63, and Figure 64. None of these curves match perfectly. All except home-based work match reasonably well at longer trip lengths and all except home-based other show significant differences in the shorter trip lengths. The fact that the home-based work data shows much higher percentages for trips over 15 miles long and a significant under reporting of trips between 2 and 12 miles long suggest that over estimation of volumes is likely to persist.

Figure 61: Home-Based Work Trip Length Distribution Results

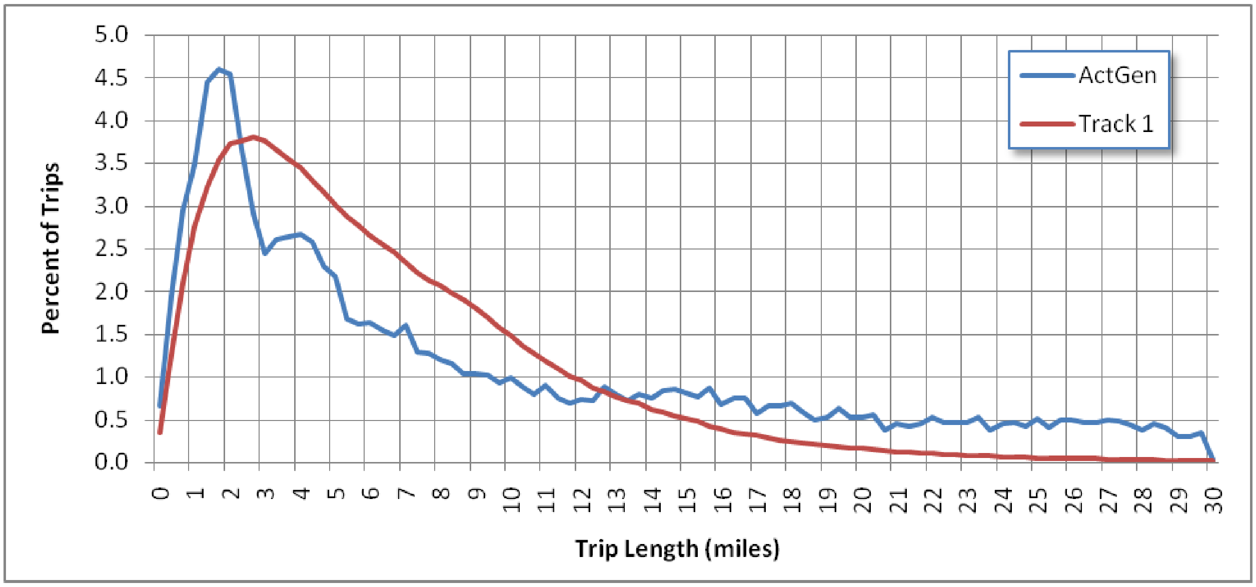


Figure 62: Non-Home Work Trip Length Distribution Results

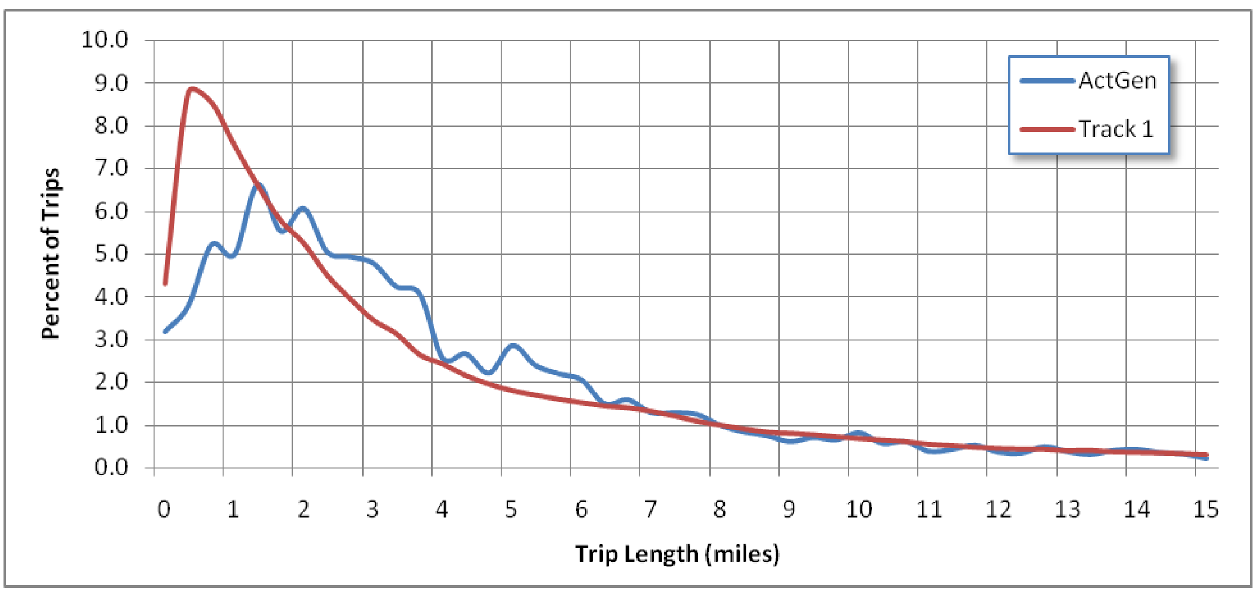


Figure 63: Home-Based Ohter Trip Length Distribution Results

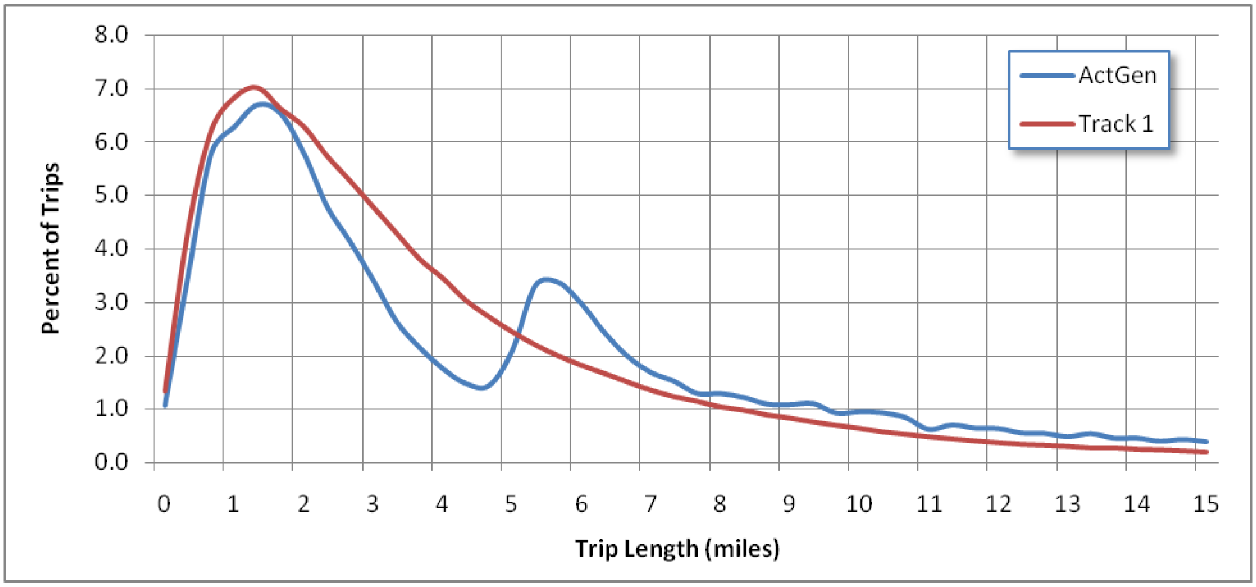
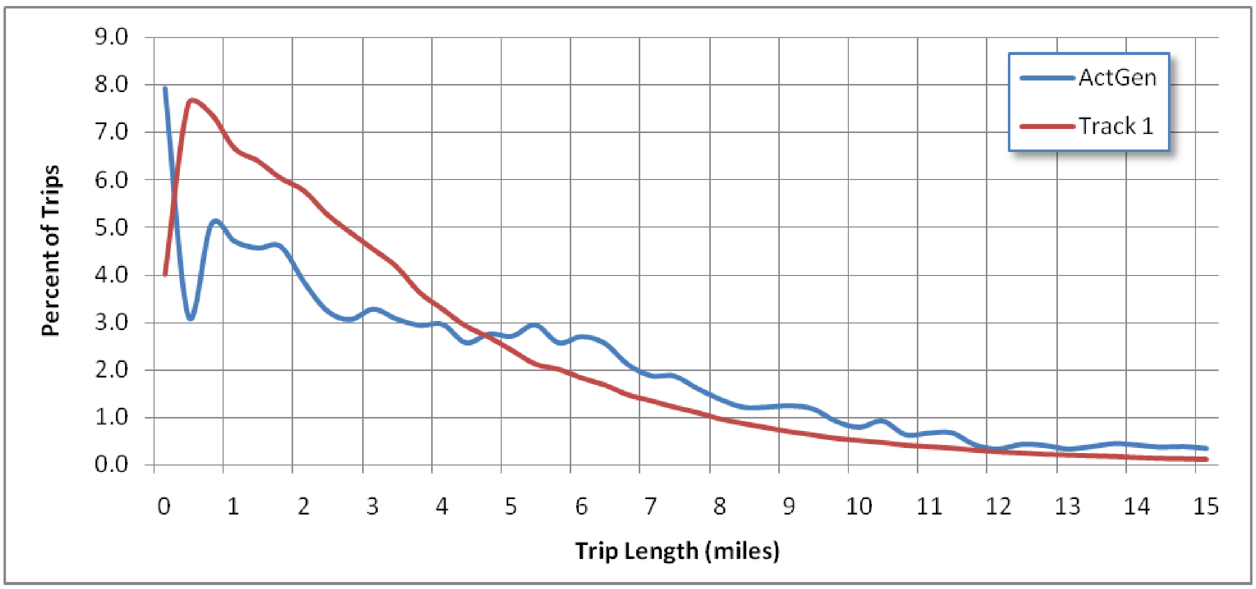


Figure 64: Non-Home-Non-Work Trip Length Distribution Results

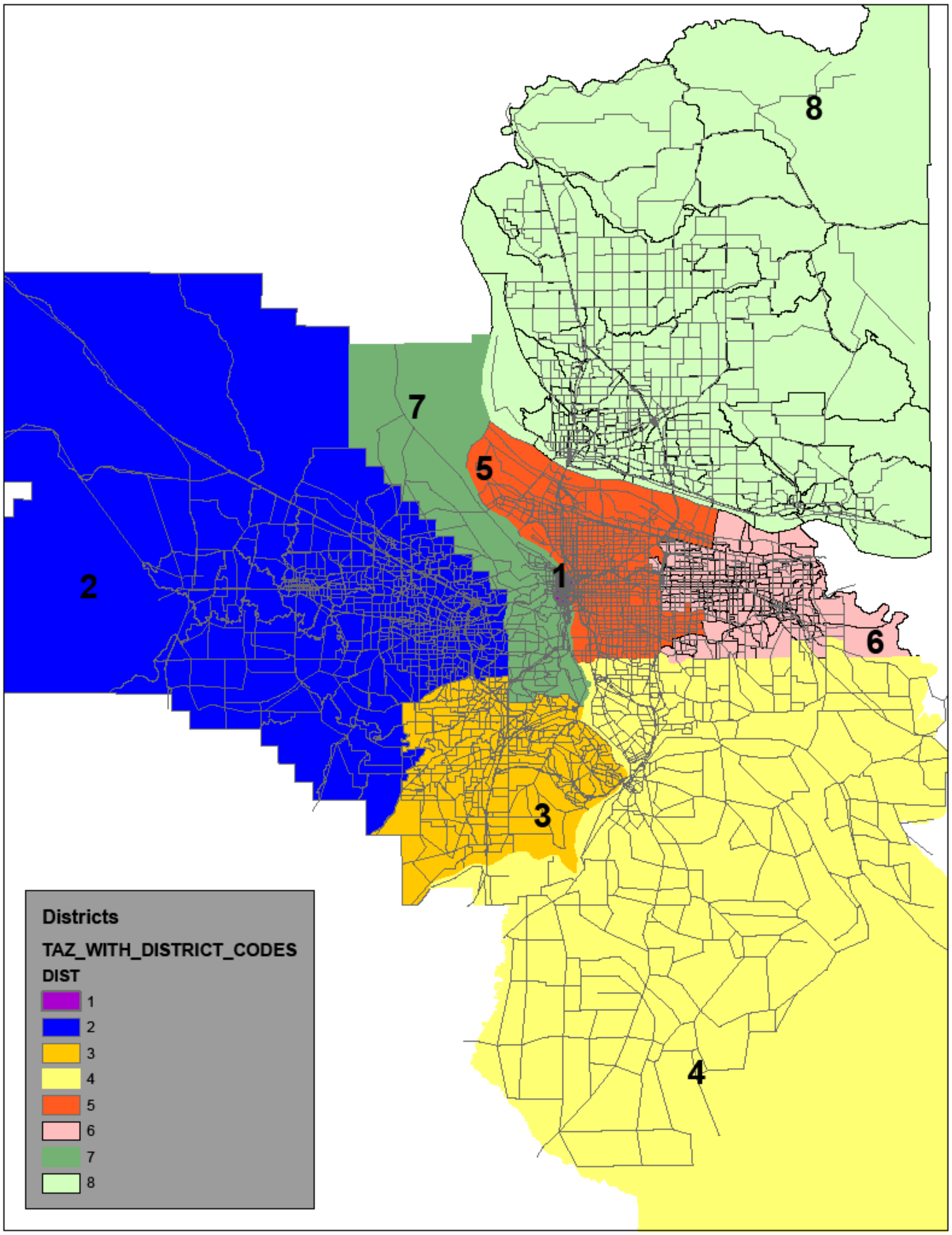


The re-generation feature of the ActGen program proved effective in feeding back travel times from the Microsimulator to activity generation. Activities for a small percentage of the households were re-generated rather than all of the households. The households selected for the first few rounds of feedback were selected because they experienced problems in the activity generator, Router or Microsimulator. The synthetic households were re-matched to a survey household and the activities generated. Minor adjustments to the location choice model were made at this point to help improve the overall average trip lengths. The new activities were added to the regional activity set and re-assigned through the Router-Microsimulator process.

7.4 District-to-District Work Trip Table

In addition to calibrating the location choice model to the trip length distribution by trip purpose, the activity patterns produced by the activity generator were compared / validated against the Track 1 trip tables by trip purpose. The trips in the activity file were summarized by METRO's eight analysis districts shown in Figure 65 to compare the district-to-district movements in METRO's trip tables.

Figure 65: METRO's Eight Analysis Districts



Due to the discrepancy in the number of vehicle trips generated by Track 1 and GEN2, the distribution of work trips by production district was studied. The comparison between the Track 1 distribution shown in Table 34 and the GEN2 distribution shown in Table 35 are reasonably close.

Table 34: Distribution of Track 1 Work Trip Productions by Destination District

Origin District	Destination District							
	1	2	3	4	5	6	7	8
1	15.7	13.3	5.5	5.8	31.8	4.7	18.0	5.3
2	2.5	76.9	8.5	1.6	3.0	0.5	6.5	0.5
3	2.0	16.1	55.2	13.7	3.7	1.0	8.0	0.4
4	1.3	1.9	8.7	64.0	11.8	8.8	2.3	1.4
5	5.0	2.6	1.7	8.5	55.4	14.2	7.7	5.0
6	1.4	0.7	0.8	11.8	27.1	54.2	1.7	2.4
7	8.5	16.5	10.5	4.8	22.7	2.6	32.1	2.4
8	1.2	0.5	0.2	1.3	6.4	1.6	1.1	87.8

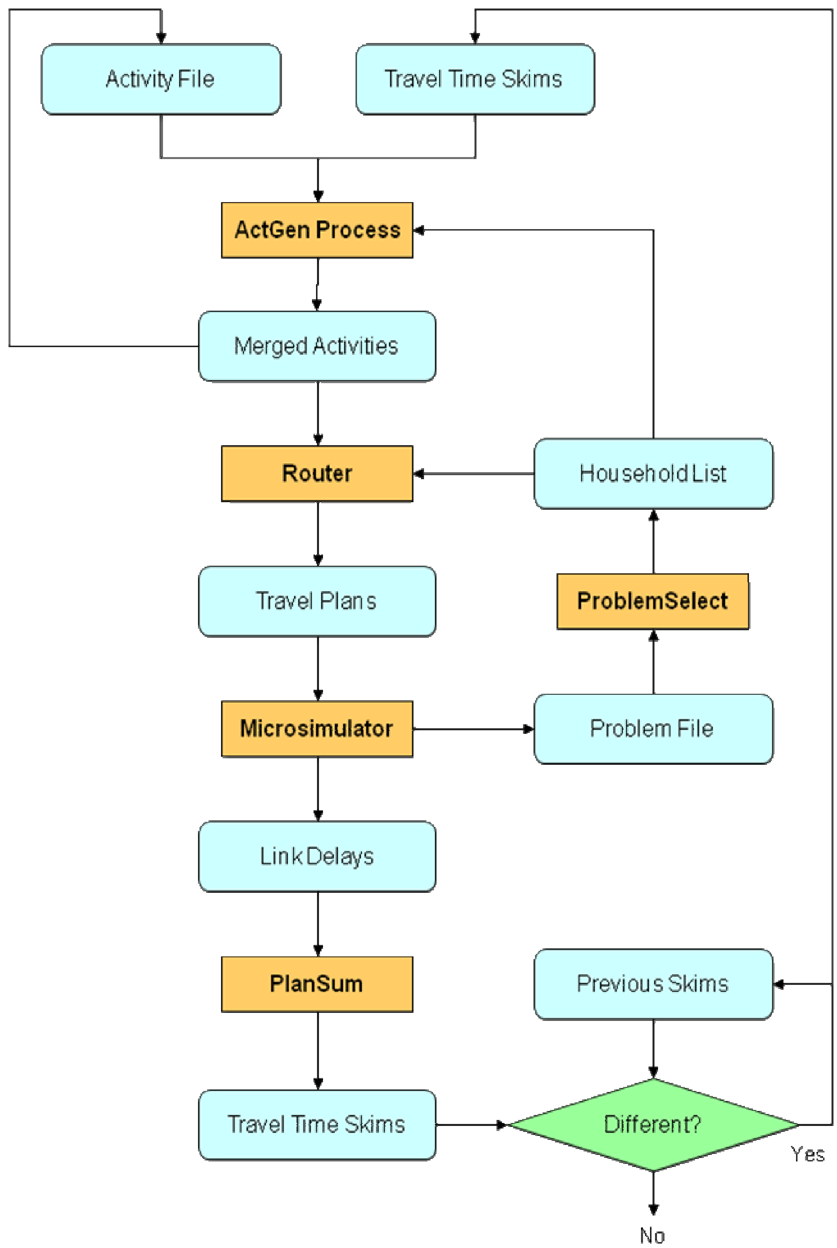
Table 35: Distribution of GEN2 Work Trip Productions by Destination District

Origin District	Destination District							
	1	2	3	4	5	6	7	8
1	13.8	10.8	6.7	7.5	23.6	7.9	20.2	9.4
2	1.0	73.1	5.8	3.6	5.1	2.2	5.7	3.4
3	0.9	11.7	58.2	10.9	4.8	3.2	6.4	4.0
4	0.6	4.5	6.6	60.9	10.6	7.9	3.3	5.6
5	1.7	5.6	2.6	9.2	55.3	8.9	5.4	11.3
6	1.0	4.1	2.8	10.7	14.8	53.3	2.0	11.4
7	4.1	14.7	8.5	7.7	14.5	3.3	42.1	5.1
8	0.8	4.2	2.4	5.4	12.1	7.3	1.9	65.8

7.5 Router-Microsimulator Runs

For this implementation a slight change was made in the way the Router and Microsimulator were integrated with the activity generation process (see Figure 66). The activities are generated using the process described above and then the Router-Microsimulator iterations are executed to stabilize the travel paths. At this point the zone-to-zone travel time skims are compared with the skims from the previous iteration. If travel time skims are unchanged, the overall modeling process has converged. Otherwise, the updated skims and a subset of households are fed back to ActGen for regeneration. The households may be selected based on routing or simulation problems or as a random sample to evaluate the impacts of the updated travel times on the location choice calibration. Minor adjustments could be made to the calibrated parameters to improve the overall average trip lengths. The regenerated activities for the selected households replace the original activities for those households and additional Router-Microsimulator iterations are applied.

Figure 66: Feedback Process for the Second ActGen Implementation



The key improvement over the first implementation revolves around the selection of problematic households for 1) regeneration in the activity model, 2) re-routing in the Router process, and 3) feedback to ActGen from the Microsimulator. Such a selection strategy is designed to minimize Microsimulator problems while simultaneously achieve stabilization. Problematic households manifest the inconsistency between the travel time skims and the realized point-to-point travel times and therefore regenerating these households is expected to reconcile the inconsistency as quickly as possible.

7.5.1 Initial Router Results

The results from the initial Router application showed the same level of over estimation as the first implementation. The link volumes were over estimated by 50 percent with an RMSE of 120 percent (Figure 67). The Columbia River screenlines (Figure 68) were even higher than the first implementation indicating that the Washington-Oregon bias constants were still not working properly.

Figure 67: Initial Router Validation by Facility Type

Facility Type	Num. Obs.	Volume Estimate	Volume Observed	Difference Volume	Difference %	Abs.Error Avg.	Abs.Error %	Std. Dev.	% RMSE	R Sq.
Freeway	51	4513110	2874828	1638282	57.0	32331	57.4	26675	74.1	0.383
Expressway	24	763657	409706	353951	86.4	14748	86.4	9085	100.9	0.822
Major Arterial	104	1861396	1143236	718160	62.8	7878	71.7	6779	94.4	0.513
Minor Arterial	82	611708	507048	104660	20.6	3399	55.0	3034	73.5	0.316
Collector	130	310006	462788	-152782	-33.0	1774	49.8	1410	63.6	0.209
Bridge/Other	18	450872	286236	164636	57.5	10014	63.0	12348	98.3	0.623
TOTAL	409	8510749	5683842	2826907	49.7	8586	61.8	14331	120.1	0.870

Figure 68: Initial Router Validation by Screenline

Link Group	Num. Obs.	Volume Estimate	Volume Observed	Difference Volume	Difference %	Abs.Error Avg.	Abs.Error %	Std. Dev.	% RMSE	R Sq..
R-1 St. John's Br	2	35609	22610	12999	57.5	6500	57.5	1240	58.0	0.000
R-2, William. River Bridg	16	730049	464966	265083	57.0	17530	60.3	16986	82.7	0.846
R-3, Sellwood Bridge	2	33203	31382	1821	5.8	1022	6.5	1288	8.7	0.000
R-4 OrCty 205 Br	4	132999	109571	23428	21.4	6850	25.0	5096	29.7	0.926
R-5, I-5 Bridge	2	283762	113838	169924	149.3	84962	149.3	22492	151.9	0.000
R-6 I-5 Hayden	2	285541	128300	157241	122.6	78620	122.6	22511	125.0	0.000
R-7 I205	2	279308	104074	175234	168.4	87617	168.4	23492	171.4	0.000
E-9 W of 122nd	34	537729	348370	189359	54.4	6701	65.4	8745	106.5	0.933
E-21 W of Tacoma	20	401693	272032	129661	47.7	8223	60.5	11905	104.6	0.947
W-7 Westhills	10	256024	172226	83798	48.7	10044	58.3	16469	107.8	0.997
W-09 Tigard Int	6	213547	154962	58585	37.8	12265	47.5	12595	65.1	0.915
W-16 NW of Tigard	10	155209	140922	14287	10.1	1938	13.7	1479	17.0	0.994
TOTAL	110	3344673	2063253	1281420	62.1	12827	68.4	21097	131.2	0.859

7.5.2 Microsimulator Iterations

Since the Router validation demonstrated that the network demand still far exceeds the observed traffic counts, a more informative application of the Microsimulator was to incrementally load a fraction of the total trips to monitor the system performance and estimate the overall network capacity. Table 36 shows the Microsimulator was able to handle the traffic loads up to the 60 percent mark with less than one percent wait time problems, after which tours started to break down. At 60 percent of households, 2.15 million trips were loaded and 97 percent of the trips were completed with a total of 30,000 wait time problems. At 70 percent, 92 percent of the trips were completed and the wait time problems jumped to 217,000. At the 100 percent level, only 63 percent of the 4 million vehicle trips were completed.

Table 36: Incremental Loading Results

Measure of Effectiveness	60% Demand	70% Demand	80% Demand	90% Demand	100% Demand
Fraction trips started	100.0%	98.9%	98.3%	97.9%	94.9%
Fraction completed	97.2%	92.1%	88.4%	85.3%	63.0%
Wait time problems	0.8%	41.9%	43.0%	44.2%	937,683 (42.2%)
Departure time problems	0.0%	0.3%	0.4%	0.3%	12,557 (0.6%)
Arrival time problems	0.2%	0.8%	10.1%	9.4%	436,721 (19.7%)
Time schedule problems	80.5%	40.5%	37.4%	37.1%	624,690 (28.1%)

7.5.3 Stabilization Runs

The link delays from the Microsimulator were used to re-route up to 5 percent of the total households with routing or simulation problems in the AM Peak period (6-8 AM). The plans were merged with the remaining plans and re-simulated. The results after nine feedback iterations are reported below (Figure 69 and Figure 70). The number of completed trips increased from 63 percent to 81 percent, but the links were still significantly over loaded. Correcting this problem requires travel demand changes.

Figure 69: Facility Type Validation after Nine Feedback Iterations

Facility Type	Num. Obs.	------Volume------ Estimate	Observed	---Difference--- Volume	%	--Abs.Error-- Avg.	%	Std. Dev.	% RMSE	R Sq.
Freeway	51	4107028	2874828	1232200	42.9	24345	43.2	21628	57.5	0.442
Expressway	24	666604	409706	256898	62.7	10940	64.1	6637	74.5	0.812
Major Arterial	104	1883273	1143236	740037	64.7	7515	68.4	4267	78.5	0.463
Minor Arterial	82	741979	507048	234931	46.3	4085	66.1	3278	84.5	0.310
Collector	130	553186	462788	90398	19.5	1935	54.4	1900	76.0	0.215
Bridge/Other	18	493510	286236	207274	72.4	11515	72.4	7480	85.6	0.752
TOTAL	409	8445580	5683842	2761738	48.6	7530	54.2	10913	95.3	0.889

Figure 70: Screenline Validation after Nine Feedback Iterations

Link Group	Num. Obs.	------Volume------ Estimate	Observed	---Difference--- Volume	%	--Abs.Error-- Avg.	%	Std. Dev.	% RMSE	R Sq.
R-1 St. John's Br	2	42918	22610	20308	89.8	10154	89.8	3704	92.8	0.000
R-2, William. River Bridg	16	737747	464966	272781	58.7	17049	58.7	9258	66.3	0.952
R-3, Sellwood Bridge	2	36149	31382	4767	15.2	2384	15.2	870	15.7	0.000
R-4 OrCty 205 Br	4	142989	109571	33418	30.5	8354	30.5	6407	36.6	0.930
R-5, I-5 Bridge	2	277709	113838	163871	144.0	81936	144.0	5925	144.1	0.000
R-6 I-5 Hayden	2	281259	128300	152959	119.2	76480	119.2	5947	119.4	0.000
R-7 I205	2	233765	104074	129691	124.6	64846	124.6	6994	125.0	0.000
E-9 W of 122nd	34	546694	348370	198324	56.9	6300	61.5	6742	89.3	0.921
E-21 W of Tacoma	20	381820	272032	109788	40.4	6198	45.6	7784	72.0	0.942
W-7 Westhills	10	266171	172226	93945	54.5	9670	56.1	11079	82.9	0.969
W-09 Tigard Int	6	201728	154962	46766	30.2	9439	36.5	11021	53.4	0.884
W-16 NW of Tigard	10	167932	140922	27010	19.2	3070	21.8	3765	33.4	0.983
TOTAL	110	3316881	2063253	1253628	60.8	11818	63.0	17409	111.8	0.867

Chapter 8 Third ActGen Implementation

Adding time budgets helped with activity generation, but it did not address the problem of trip distribution. Some locations attracted far too many trips while other areas attracted too few. There were also river crossing impedances that were not adequately discouraging trips on various interchanges. It did appear that the granularity of the travel time data (zone-to-zone skims) was a bit too course for the time budget approach to be as effective as it might. This issue was further complicated by a significant number of inaccuracies and inconsistencies in the time allocated for travel in the household activity survey. Since each error in the survey is magnified about 100 times when applied to the synthetic population, correcting the survey records became a priority.

8.1 Survey Cleaning

A typical ActGen run for Portland produced problem messages for 11 percent of the trips (or tour legs) , with the problems spread over 41 percent of households. The effect of the 11 percent, however, is more pronounced, since a failed trip eventually leads to a failed tours that cannot be simulated even with the aid of magic moves.

Analysis of the problematic trips shows that 45 percent of the problems were due to vehicle passenger problems (passengers not finding a ride at the time of the activity), which were directly the result of coding or reporting inaccuracies in the household activity survey. The CheckSurvey program was created to identify inconsistencies in the household activity survey and automatically correct the problems or flag the records for manual cleaning and reconciliation.

In the example shown in Figure 71, 12 of the 22 trips recorded for this household have minor coding errors that are flagged as problems by the activity generator. The first problem CheckSurvey identified is labeled “No Driver” in the comment field of person 1 activity 7. This is followed by “Origin Location” in activity 8. Person 1 expected to leave home at 19:30 to arrive at location #6 at 19:45 as a passenger in a car. The person then expects a ride from location #6 at 20:30 to return home at 20:45. Simple inspection of the other household activity records shows that person 2 is the most likely driver for this trip, but the activity records for this person indicate the driver leaves home at 19:00 rather than 19:30 and travels to location #12 rather than #6. The return trip has the correct times and the same number of passengers. Person 4 appears to be the third person in the car. These records have the same times as person 1, but the location code matches the driver.

The resolution of these errors is shown in Figure 72. Since two out of the three travelers claim location #12 as the destination, the destination for person 1 was changed to location #12. Also two out of the three travelers claim the trip started at 19:30 rather than 19:00. The end time of activity 7 of person 2 was accordingly changed to 19:30. Note that 19:30 would have been chosen even if there were only two travelers. In this case the time allocated to the outbound trip would have been compared to the travel time for the inbound trip to decide that the outbound trip was only 15 minutes long rather than 45 minutes long. From a travel time budget and trip length perspective, this is a significant difference.

Figure 71: Typical Data Problems Identified by CheckSurvey

HHOLD	PER	ACT	PUR	START	END	DUR	MODE	VEH	LOC	PASS	CON	NOTES
201204	1	1	0	0:00	8:16	8:16	1	0	1	0	0	
201204	1	2	2	8:25	8:35	0:10	2	1	2	0	0	
201204	1	3	1	9:05	16:00	6:55	2	1	3	0	0	
201204	1	4	6	16:10	16:11	0:01	2	1	4	0	0	
201204	1	5	2	16:22	16:42	0:20	2	1	5	1	0	
201204	1	6	0	17:00	19:30	2:30	2	1	1	1	0	
201204	1	7	1	19:45	20:30	0:45	2	2	6	2	0	No Driver
201204	1	8	0	20:45	1@3:00	6:15	2	2	1	2	0	Origin Location
201204	2	1	0	0:00	7:45	7:45	1	0	1	0	0	
201204	2	2	3	8:00	9:00	1:00	2	1	7	1	0	
201204	2	3	1	9:05	11:50	2:45	2	1	8	0	0	
201204	2	4	4	12:10	12:12	0:02	2	1	9	0	0	
201204	2	5	5	12:35	13:15	0:40	2	1	10	0	0	
201204	2	6	1	13:20	17:20	4:00	2	1	11	0	0	
201204	2	7	0	18:00	19:00	1:00	2	1	1	0	0	
201204	2	8	1	19:45	20:30	0:45	2	1	12	2	0	
201204	2	9	0	20:45	1@3:00	6:15	2	1	1	2	0	
201204	3	1	0	0:00	7:15	7:15	1	0	1	0	0	
201204	3	2	1	8:00	9:00	1:00	2	2	7	1	0	Origin Time
201204	3	3	0	9:10	1@3:00	17:50	8	0	1	0	0	
201204	4	1	0	0:00	19:30	19:30	1	0	1	0	0	
201204	4	2	1	19:45	20:30	0:45	2	2	12	2	0	Origin Time
201204	4	3	0	20:45	1@3:00	6:15	2	2	1	2	0	
201204	5	1	0	0:00	7:35	7:35	1	0	1	0	0	
201204	5	2	7	8:05	16:10	8:05	9	0	13	0	0	
201204	5	3	2	16:22	16:42	0:20	2	2	14	1	0	No Driver
201204	5	4	0	17:00	1@3:00	10:00	2	2	1	1	0	Origin Location
201204	6	1	0	0:00	1@3:00	1@3:00	1	0	1	0	0	

The next problem is highlighted for activity 3 of person 2. The "Origin Time" message implies there is a driver leaving the same origin and traveling to the same destination that arrives at the same time, but the end time of the previous activity is different. Person 2 leaves home at 7:45 to travel to location #7 while person 3 claims to have left at 7:15. In this case, the two people have the same activity duration, but they travel to different places by different modes after the activity is completed. Person 2 got a ride home from a friend (mode 8) and the trip took 10 minutes. This would imply that the 7:45 departure time from home is more likely the correct time. If this information were not available, the travel time estimate of the driver would be assumed to be more accurate.

The third set of problems was flagged for person 5 activities 3 and 4. In this case, the trip times and purposes suggest person 1 is the driver for this trip, but the activities should be at locations #4 and #5 rather than #13 and #14. The small discrepancy in the end time of the previous activity relates to the fact that the driver's activity purpose is serve passenger (6). A serve passenger activity is assumed to take a minimum of one minute in TRANSIMS. The end time of the passenger's previous activity was modified by one minute to match the driver's serve passenger activity.

Figure 72: Results of the Survey Cleaning Process

HHOLD	PER	ACT	PUR	START	END	DUR	MODE	VEH	LOC	PASS	CON	NOTES
201204	1	1	0	0:00	8:16	8:16	1	0	1	0	0	
201204	1	2	2	8:25	8:35	0:10	2	1	2	0	0	
201204	1	3	1	9:05	16:00	6:55	2	1	3	0	0	
201204	1	4	6	16:10	16:11	0:01	2	1	4	0	0	
201204	1	5	2	16:22	16:42	0:20	2	1	5	1	0	
201204	1	6	0	17:00	19:30	2:30	2	1	1	1	0	
201204	1	7	12	19:45	20:30	0:45	2	2	12	2	0	
201204	1	8	0	20:45	1@3:00	6:15	2	2	1	2	0	
201204	2	1	0	0:00	7:45	7:45	1	0	1	0	0	
201204	2	2	3	8:00	9:00	1:00	2	1	7	1	0	
201204	2	3	11	9:05	11:50	2:45	2	1	8	0	0	
201204	2	4	4	12:10	12:12	0:02	2	1	9	0	0	
201204	2	5	5	12:35	13:15	0:40	2	1	10	0	0	
201204	2	6	1	13:20	17:20	4:00	2	1	11	0	0	
201204	2	7	0	18:00	19:30	1:30	2	1	1	0	0	
201204	2	8	12	19:45	20:30	0:45	2	1	12	2	0	
201204	2	9	0	20:45	1@3:00	6:15	2	1	1	2	0	
201204	3	1	0	0:00	7:45	7:45	1	0	1	0	0	
201204	3	2	13	8:00	9:00	1:00	2	2	7	1	0	
201204	3	3	0	9:10	1@3:00	17:50	8	0	1	0	0	
201204	4	1	0	0:00	19:30	19:30	1	0	1	0	0	
201204	4	2	12	19:45	20:30	0:45	2	2	12	2	0	
201204	4	3	0	20:45	1@3:00	6:15	2	2	1	2	0	
201204	5	1	0	0:00	7:35	7:35	1	0	1	0	0	
201204	5	2	7	8:05	16:11	8:06	9	0	4	0	0	
201204	5	3	2	16:22	16:42	0:20	2	2	5	1	0	
201204	5	4	0	17:00	1@3:00	10:00	2	2	1	1	0	
201204	6	1	0	0:00	1@3:00	1@3:00	1	0	1	0	0	

After cleaning the survey, ActGen was re-run and 82 percent of the vehicle passenger problems and 49 percent of the total problems were eliminated (Table 37). Consequently, the number of households with problems dropped 42 percent. In a clean survey, the internal consistency reduces the number of activity problems and increases the number of tours that can be successfully simulated.

Table 37: Impact of Survey Cleaning on Activity Generation Problems

Problem Type	*Original Survey*	*Cleaned Survey*	*% Difference*
Households with Problems	260,804	152,184	- 42%
Total problems	1,013,719	515,377	- 49%
Vehicle access problems	7,974	8,932	+ 12%
Vehicle passenger problems	432,780	76,132	- 82%
Time schedule problems	78,074	84,570	+ 8%
Activity location problems	474,005	319,651	-33%
Activity duration problems	28,886	26,092	- 10%

8.2 Activity Location Problems

The survey cleaning addressed the majority of vehicle passenger problems and reduced most other problems (Table 37). There were small increases in vehicle access problems and time schedule problems, but these changes were small compared to the number of activity location problems. Activity location problems were reduced by 33 percent, but now represented 47 percent of all the problems. It was suspected that these problems might be caused by the time budget constraints in the location choice model (i.e., not being able to fit travel times into the person's activity pattern). Further investigation, however, showed that only 25 percent of the location problems are due to time budget constraints, with 74 percent due to walk distances exceeding 2,000 meters. Three new problem types were added to ActGen to differentiate location problems by mode of travel (auto, walk, bike, and transit).

Since the majority of the problems were walk-related, the impact of network density and zone-based skims became more important. The METRO regional network only includes collector and above facility types. This means there are fewer streets in the network that can be used for walk trips. This impact is compounded by the fact that zone-to-zone travel time and distance skims do not accurately measure the travel impedance of short trips. Methods of refining the travel impedance for short trips would improve the trip length distributions for both auto and walk trips. This could also help reduce the number of long trips that caused the volumes to exceed the counts.

8.3 Time Budgets Revisited

Several location choice models using time budgets and distance ranges were tested and evaluated. This process worked well for a single activity between two anchor activity locations. If there are multiple activities between the anchor locations, the process was considerably more difficult. In these cases, some consideration for the diversion from the minimum distance path to locate intermediate activities needed to be considered. An attempt was made to estimate the diversion using the Portland survey data, but inconsistencies between the survey activity locations and reported travel times made this difficult.

Several location choice models were tested that use time budgets to locate multiple activities between two anchor locations. The approach that generated the most favorable results limits the diversion from the shortest path between the anchor locations to five minutes per intermediate activity. A schedule penalty is computed for each candidate zone based on the type of constraint, expected travel time (from skims) and the survey schedule. This penalty was added to the total utility for each zone to select the zone that best fits the survey schedule. The process worked well and generated about 97 percent of all possible activities. Previous location choice models never crossed the 90 percent mark. The advantage of this "best fit" location choice model is that it requires much less regenerations and circumvents the need to progressively adjust the diversion factor.

The impacts of this approach on activity generation problems are shown in Table 38. All of the location-based problems improved over the other options. The percentage of problems that relate to time schedules and duration increased. This is partially due to the fact that the total number of problems reduced, but it is also indicative of the fact that locating the intermediate stops along the path between the anchor locations did not address the time schedule problems. Options for improving time scheduling are considered in the next section.

Table 38: The Impact of Various Time Budget Methods on ActGen Problems

	No Time Budget	*Time Budget*	*Penalty-Based Time Budget*
Activities with Problems	17.4%	13.6%	10.4%
Activities Written	82.8%	90.7%	96.0%
Households with Problems	65.1%	57.1%	43.5%
Time Schedule Problems	9.7%	9.9%	14.0%
Activity Location Problems	27.7%	37.5%	8.5%
Vehicle Passenger Problems	6.9%	0%	0%
Activity Duration Problems	39.1%	21.5%	35.1%

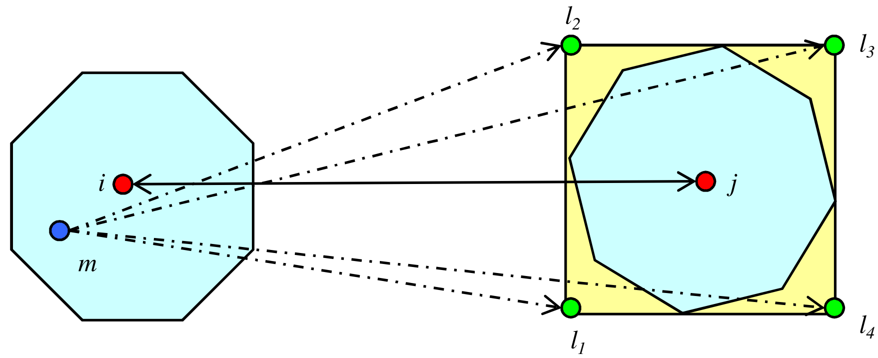
8.4 Adjusting Travel Time Skims

The location choice model uses zone-to-zone travel time skims to locate zones first and then activity locations later. While this methodology is both computationally and memory efficient, it's fidelity with respect to travel time constraints and time budgets leaves much to be desired. Moreover, computing point-to-point travel times is practically infeasible. For example, the Portland network would entail solving for a shortest path tree 26,000*26,000 times for each time slice. Assuming eight time periods, this would result in 250 gigabytes of data and require 9,600 hours to generate on a standard computer.

A key reason zone-to-zone travel time skims are problematic is that they represent an average travel time without any account for the variance. Since skims measure centroid-to-centroid travel times, the value is reasonably accurate for activity locations in close proximity to the zone centroid. As zone size increases or the distance between zones decreases, the travel time skims will not accurately represent the travel times between random activity locations within the zones. The skims will overestimate or underestimate actual travel times. This biases the location choice probabilities and introduces activity scheduling errors.

To address this problem, zone size adjustment factors were added to the location choice model. The software uses the X-Y coordinates of each activity location to calculate the minimum and maximum X and Y range for each zone. When the location choice model considers a given destination zone, it calculates the straight-line distance between the current activity location and each of the corner points of the bounding box defined by the X-Y range (see Figure 73). The distance to the closest corner point is divided by the distance between the zone centroids and saved as the minimum adjustment factor. The distance to the farthest corner is divided by the centroid distance to estimate the maximum adjustment factor. These adjustment factors can then be applied to the zone-to-zone skim to estimate a range of travel times between the current location and potential activity locations in the destination zone.

Figure 73: Computer Lower and Upper Travel Time Bounds



Mathematically, the lower and upper bounds are computed as follows:

$$LB_{ij} = C_{ij} \times \frac{\underset{l}{Min}\{d_{ml}\}}{d_{ij}} \tag{14}$$

$$UB_{ij} = C_{ij} \times \frac{\underset{l}{Max}\{d_{ml}\}}{d_{ij}} \tag{15}$$

where LB_{ij} and UB_{ij} are the lower and upper travel time bounds for C_{ij}, respectively. The zone bounds are used to more accurately identify which zones fall within the required travel time budgets of a given activity. If a zone is only partially enclosed by the time budget range, the utility value is factored by a percent coverage or overlap factor.

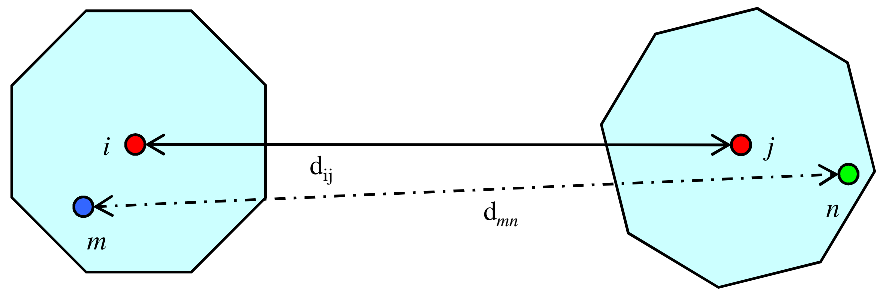
Once the zone is selected, an activity location within the zone is chosen based on the distance between the activity locations. This distance is also divided by the distance between the zone centroids and multiplied by the zone-to-zone travel time skim to estimate the travel time between the activity locations. This travel time is used to schedule the activity start and end times.

The activity location travel time is computed as follows:

$$\hat{C}_{ij} = C_{ij} \times \frac{d_{mn}}{d_{ij}} \tag{16}$$

where $\hat{C}_{ij}$ is the adjusted zone-to-zone travel time skim between zones i and j, d_{ij} is the straight-line distance between the zones i and j, and d_{mn} is the straight-line distance between activity locations m and n. Figure 74 presents a schematic for this type of adjustment.

Figure 74: Adjusting Zone-to-Zone Travel Time Skims



8.5 Calibrating the Activity-Based Model

The second ActGen implementation introduced travel time budgets in the location choice model and re-generation of problematic households to increase the likelihood that a given household will be able to complete the full set of activities assigned to each household member. Unfortunately, such an operation led to more questions than answers especially in the context of calibrating the activity generation model. In particular, the regeneration process gave problematic households a new set of activities as well as new locations. This had two major implications. First, the distance calibration step was less effective because the activity pattern changes made the new location choice coefficients somewhat obsolete. Second, travel time budgets depend on stable and accurate travel time data and stable travel time data depends on stable travel patterns.

This implementation reduced the degrees of freedom in the calibration process by separating the feedback iterations into four sequential steps. These steps:

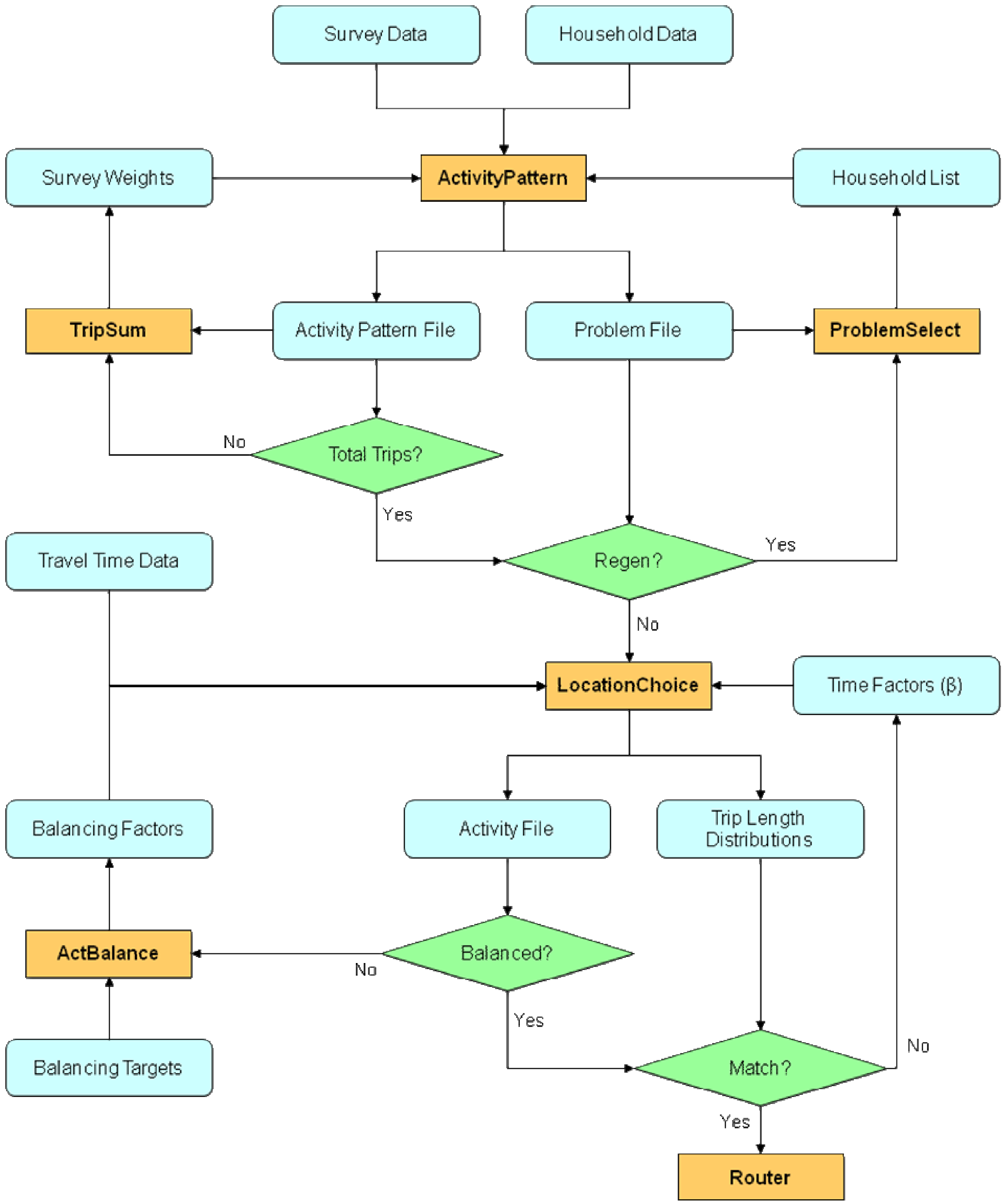
1. Adjust the survey weights to replicate total trips;
2. Address problems with household activity patterns;
3. Balance attraction totals to activity capacity; and
4. Adjust location choice coefficients to replicate trip lengths.

The ActGen program was split into two programs in order to calibrate household activity patterns separately from the location choice model. The main output of the ActivityPattern program is a skeletal activity pattern file (activities without locations) for each synthetic household. The LocationChoice program then uses the activity pattern file to assign a location for each activity. The ActBalance program was written to facilitate the balancing of activity attraction totals to activity targets by zone.

Figure 75 presents the calibration framework for this implementation. Note that ActivityPattern generates a skeletal activity pattern for each synthetic household and then loops through several regeneration iterations until the total number of trips and most of the activity pattern problems are resolved. At this point the activity patterns for a household are fixed and the location choice process begins. This process applies a set of location choice coefficients and travel time skims to the activity patterns using time budgets, constrained intermediate stop deviations, and zone range factors. The total activities attracted to a given area are compared to the attraction targets by trip purpose.

Attraction balancing factors are fed back to the location choice model until a reasonable balance is achieved. At this point the average trip length and trip length distribution by trip purpose are compared to determine if adjustments to the location choice coefficients are warranted. Finally, trip movements between the eight planning districts are compared to METRO trip tables to validate the overall process.

Figure 75: Activity Generation Process with Activity Patterns and Attraction Balancing



8.5.1 Total Vehicle Trips in the Region

The 1994 Portland regional model includes 5.0 million daily vehicle trips. 223,411 of these trips are non-household-related (itinerant) trips that need to be added to any activity-based model to replicated total traffic. This means the activity model will need to generate approximately 4.8 million daily vehicle trips to reproduce the METRO and Track 1 travel demand. As noted in the original study, expanding the household activity survey to the full population generates less than 4.1 million vehicle trips. A closer look into the trip making of survey households revealed that about 12 percent of the households do not make any vehicle trips. Assigning a selection weight of zero to non-trip making households enabled the activity generator to produce 4.4 million vehicle trips.

To achieve the 4.8 million target, the survey household weights needed to be calibrated to generate eight percent more vehicle trips. This was done by assigning a higher weight to households that made more trips. The probability of a household i of a given type N being selected in the ActivityPattern program is given by:

$$p_{(i)} = \frac{w_i}{\sum_{j \in N} w_j}, \qquad i \in N \tag{17}$$

where $p_{(i)}$ is the probability that household i is selected and w_i is the weight associated with household i. In the calibration process, a new set of weights were computed such that the weighted number of trips per household of type N will yield an additional $h\%$ over the old set of weights. The objective is to grow the number of trips generated uniformly across all household types to preserve the original distribution of trip generation by household type.

Let $w_i^{(old)}$ and $w_i^{(new)}$ be the old and new sets of household weights, respectively. The new weights must satisfy the following equality:

$$\frac{\sum_{i \in N} T_i w_i^{(new)}}{\sum_{i \in N} w_i^{(new)}} = (1 + 0.01h) \frac{\sum_{i \in N} T_i w_i^{(old)}}{\sum_{i \in N} w_i^{(old)}}, \qquad \forall N \tag{18}$$

where T_i is the number of vehicle trips generated by household i. The left-hand side of this equality represents the weighted average number of trips generated by a household of type N using the new set of weights. In order to produce more trips, households making more trips must have larger weights than households making fewer trips. This can be done by assuming the following weighting function:

$$w_i^{(new)} = T_i^{\lambda_N} \tag{19}$$

where λ_N is a calibration parameter. The weighting function renders Equation (18) to:

$$\frac{\sum_{i\in N} T_i^{(\lambda_N+1)}}{\sum_{i\in N} T_i^{(\lambda_N)}} = (1+0.01h)\frac{\sum_{i\in N} T_i w_i^{(old)}}{\sum_{i\in N} w_i^{(old)}}, \qquad \forall N, \lambda_N \geq 0 \tag{20}$$

The right-hand-side of Equation (20) is known. This equality is solved numerically (trial-and-error) to find the λ_N for each household type N. For h = 8 percent, the computed values of λ_N varied from a low of 0.3 to a high of 1.1, generating 4.8 million daily vehicle trips in the process. When added to the itinerant trips, the activity file includes 4.98 million vehicle trips or 0.4 percent less than the total trips generated by the Track 1 process.

8.5.2 Trip Attraction Balancing

The trip attraction balancing was performed for work, discretionary, college, and school trip purposes. The attraction targets were obtained from METRO (Track 1) trip tables. Since TRANSIMS uses a tour-based location choice model, return trips to home were not considered in the attraction targets. In other words, only the production-attraction trip tables were used. The ActBalance program calculates attraction balancing factors by trip purpose that are applied to the zone attraction weights in the location choice model to increase or decrease the attractiveness of a given zone as a activity location.

The key complication in this process was establishing the attraction targets for tour-based destination types. The METRO model includes home-based work and non-home work trip tables containing 567,000 and 408,000 daily trips, respectively, or a total of 975,000 trips. Work trips in the activity model are reported under home-based tours for high, medium, and low income and non-home work sub-tours for high, medium, and low income households for a total of 952,000 trips, or a 2.8 percent difference.

Discretionary trips are somewhat tricky to analyze due to the differences between trip and tour modeling conventions. Track 1 has 979,000 home-based other trips and 460,000 non-home-non-work trips for a total of 1,440,000 daily discretionary trips. The activity model, on the other hand, has several additional trip purposes such as recreation, visit, shop, other, and serve passenger that are also discretionary in nature. If these trips are included, the total discretionary trips would be 1,610,000, or 11 percent higher than the Track 1 total.

The activity model generated 53,000 college trips which compares well with the 49,000 college trips in the Track 1 model. School trips, on the other hand, were significantly different. The Track 1 process includes 21,000 school trips, whereas the activity model generated less than 6,000 trips. Neither number appears to be very realistic given the number of children in the synthetic population. Walk and school bus trips may not be properly included in either dataset. The shortcoming of the activity model can be attributed to the limited number of survey households with school trips, which stands at five percent. On the other hand, a significant number of the serve passenger trips included in the discretionary total are parents driving their children to school. Unfortunately, there is no easy way to identify the purpose of the passenger in a serve passenger trip.

With the exception of school trips, the activity model attraction totals compare reasonably well to the trip purpose totals generated by METRO's regional model. Table 39 shows the difference in trips between METRO (Track 1) and ActGen. The overall activity model is about 5.6 percent higher than the trip model, indicating more multi-stop tours.

Table 39: Comparison of Total Trips by Trip Purpose

Trip Purpose	*METRO*	*ActGen 4.0*	*Percent Difference*
Work	975,258	951,700	-2.4
Discretionary	1,438,948	1,609,040	11.8
College	49,321	53,171	7.8
School	20,748	5,705	-72.5
Total	**2,484,275**	**2,619,616**	**5.6**

The attraction totals by trip purpose were then distributed to the eight planning districts shown in Figure 65 based on the zone attraction shares from the METRO (Track 1) trip tables. ActBalance uses the zone attraction targets by trip purpose to calculate a balancing factor for each planning district. All zones within a given district have their attraction weights modified by this factor. The mathematical expression for computing such a factor is provided below:

$$A_{n,j}^{new} = A_{n,j}^{old} \frac{\dfrac{T_{n,d}^{*}}{\sum\limits_{k=1-8} T_{n,k}^{*}}}{\dfrac{T_{n,d}}{\sum\limits_{k=1-8} T_{n,k}}} = \qquad \forall n, d, j \in d \tag{21}$$

where $A_{n,j}^{new}$ is the new attraction weight for zone j for hosting activity purpose n, $T_{n,d}^{*}$ is the total trips of purpose n attracted to district d as computed by Track 1, and $T_{n,d}$ is the total trips of purpose n attracted to district d as computed by the activity model. The term $T_{n,d}^{*} / \sum\limits_{k} T_{n,k}^{*}$ represents the fraction of trips of purpose n attracted to district d.

METRO's trip model zone attraction weights were used as the initial zone weights. Before balancing, the location choice model sent 70 percent and 54 percent more work trips to districts 5 and 1, respectively; and 81 percent fewer trips to district 7 (Figure 76). The gaps were considerably reduced after balancing. District 1 received 8.5 percent more trips than Track 1, while districts 5 and 7 receiving 18 percent and 8 percent less than Track 1, respectively.

Similarly, the location choice model sent 53 percent more discretionary trips to district 5 and slightly less trips to the remaining districts (Figure 77). After balancing, the gaps in all districts, including district 5, were virtually zero. The distribution of college trips across the planning districts were initially reasonable except for districts 2, where 48 percent more trips were assigned, and district 7 where 29 percent fewer trips were attracted (Figure 78). Little improvement was noticed after balancing mainly due to the fact that relatively few zones have college activities.

Figure 76: Work Trip Attractions by Planning District

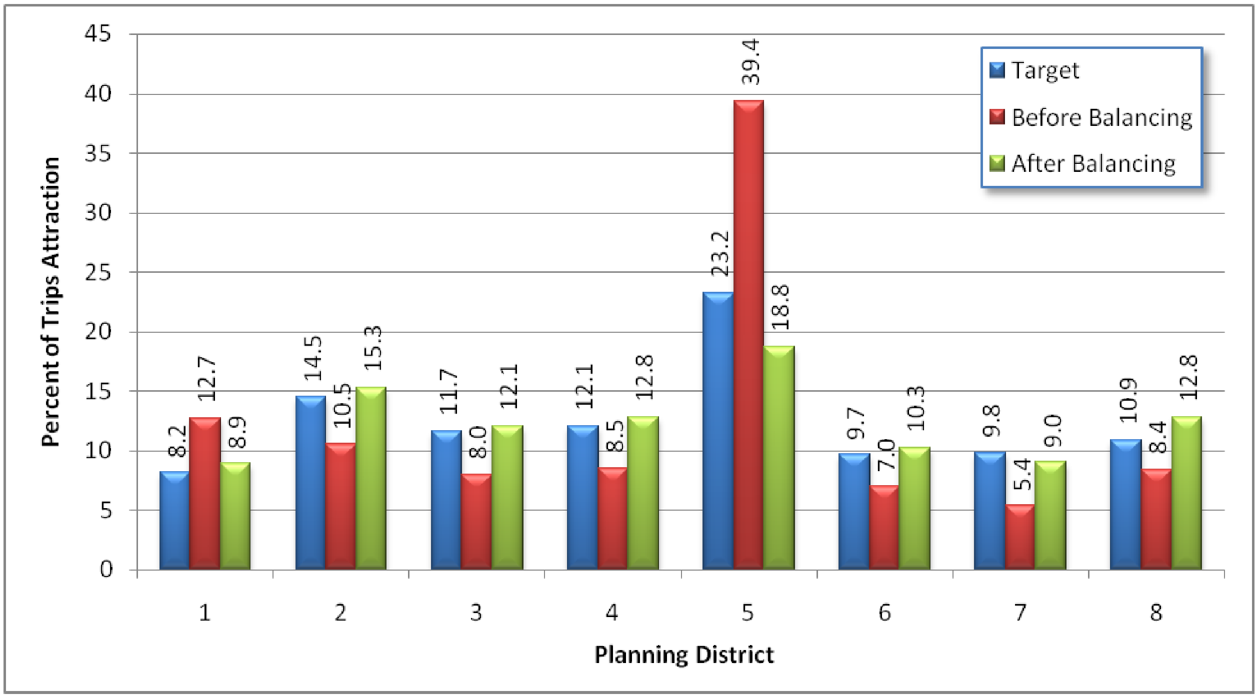


Figure 77: Discretionary Trip Attractions by Planning District

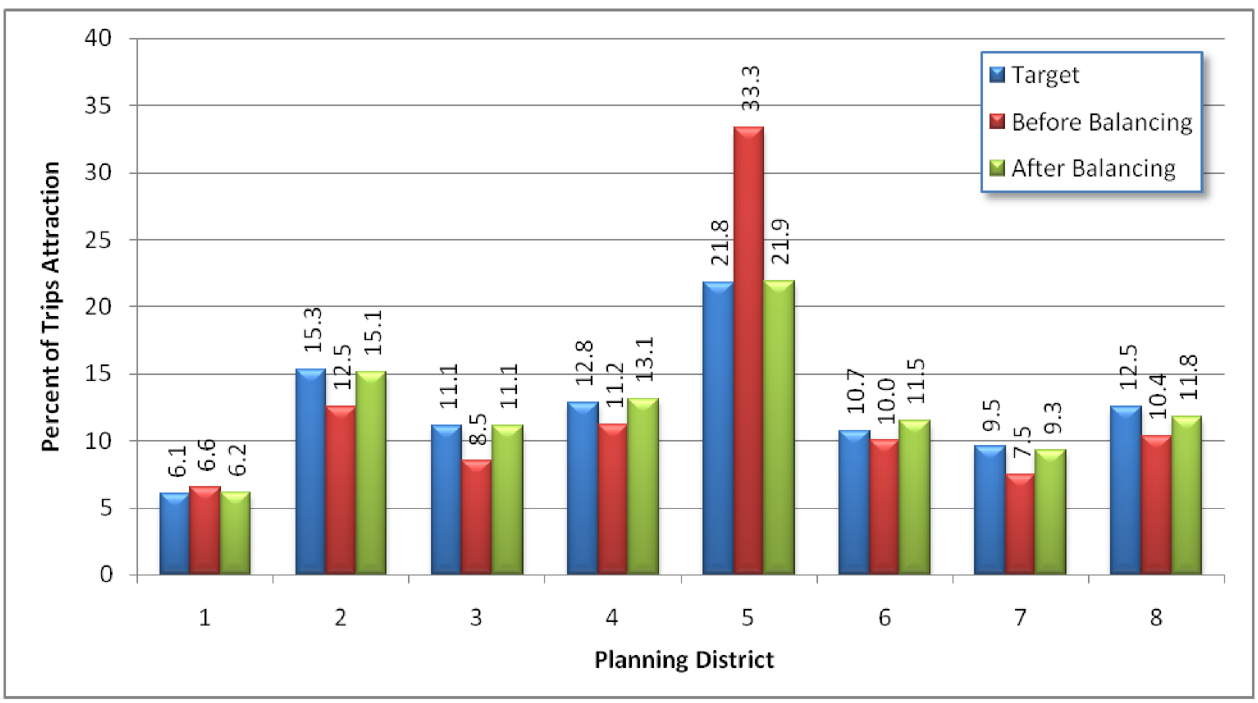


Figure 78: College Trip Attractions by Planning District

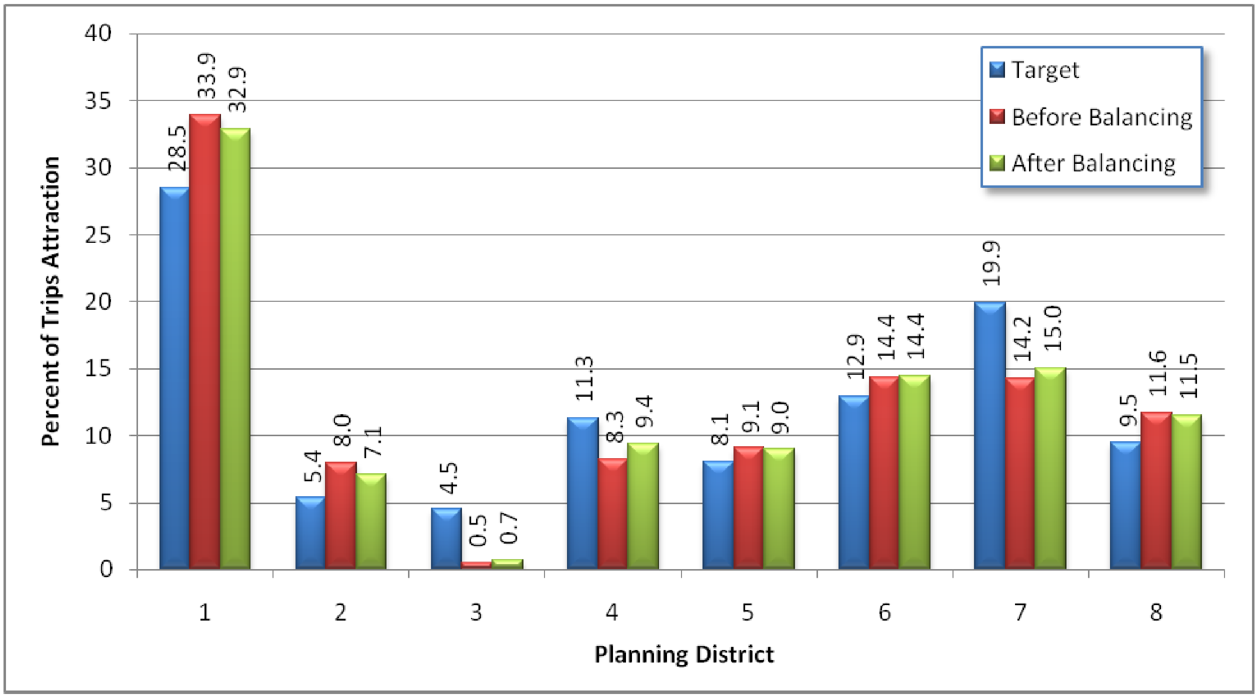
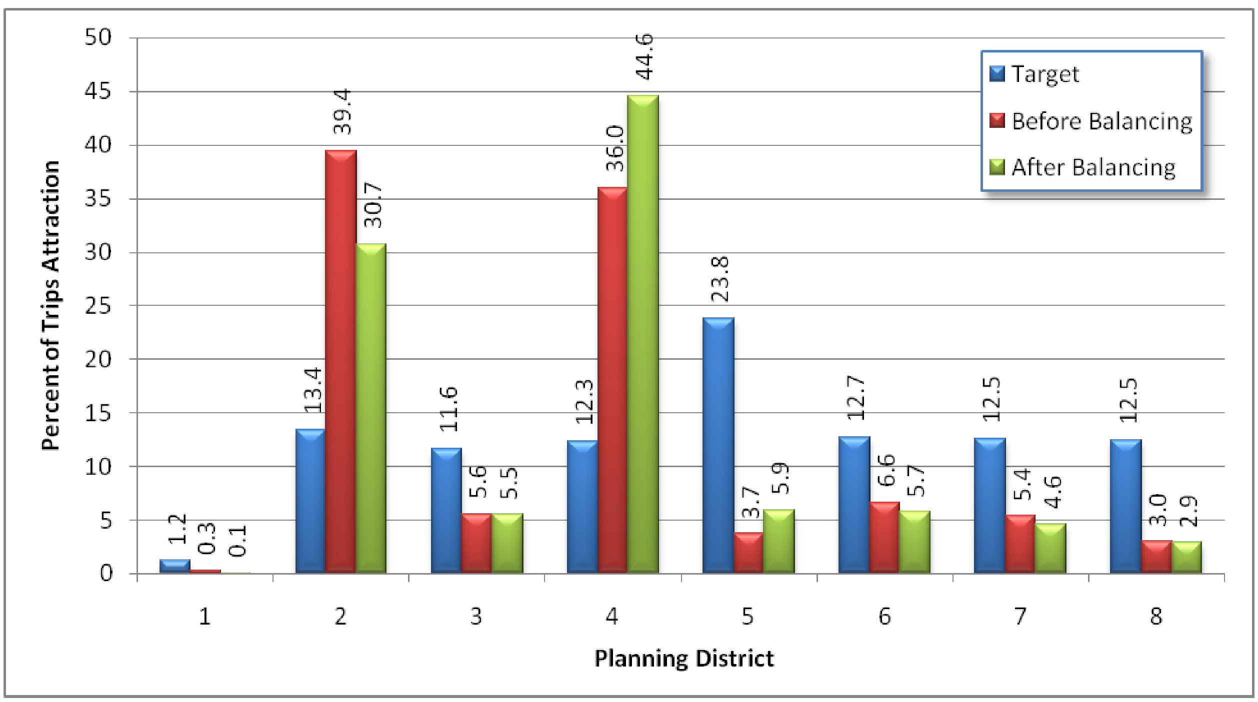


Figure 79: School Trip Attractions by Planning District



Balancing school trips was virtually impossible. The location choice model sent way too many school trips to districts 2 and 4, and way too few trips elsewhere (Figure 79). Balancing improved districts 2 and 6 some, but made the remaining districts worse. However, due to the small market share of school trips in the model, these results were deemed to have little impact on the assignment results.

8.5.3 Trip Length Calibration

In this application ten iterations of coefficient adjustments were needed to calibrate the trip lengths to the observed data. The calibration process started with the latest β_n factors and applied a series of successive adjustments until the average trip lengths matched METRO's internal estimates. There are eight trip purpose coefficients on anchored tours that need calibrating, namely work, school, college, shop, recreation, visit, other, and serve passenger. For simplicity and convenience, the latter five trip purposes were collectively combined as discretionary trips and calibrated using a single coefficient. All non-anchored tours were calibrated as discretionary trips as well. Table 40 presents the final time factors at convergence.

Table 40: Location Choice Coefficients by Tour Type

Activity Purpose	*Tour Type*	*Coefficient*
Work	Anchored	- 0.9
Discretionary	Anchored	- 1.1
School	Anchored	- 0.5
College	Anchored	- 2.1
Discretionary	Non-Anchored	- 0.5

Table 41 and Figure 80 present the average distances by tour leg type. In short, total trips were 3.2 percent shorter than the METRO survey targets, with most tour leg types in close agreement. Similar results are reported for trip types in Figure 81.

Table 41: Comparison of Average Trip Length by Tour Leg Type

Tour Leg Type	*Survey*	*ActGen 4.0*	*Percent Difference*
HW	7.54	7.66	1.6
HD	4.43	4.41	-0.4
WH	7.59	7.71	1.5
WW	5.92	5.99	1.2
WD	4.90	4.87	-0.7
DH	4.41	4.21	-4.6
DW	4.66	5.19	11.3
DD	4.13	4.53	9.7
HSch	2.75	2.15	-21.9
HCol	5.44	4.83	-11.3
Total	**4.91**	**4.75**	**-3.2**

Figure 80: Average Trip Length Comparisons by Tour Leg Type

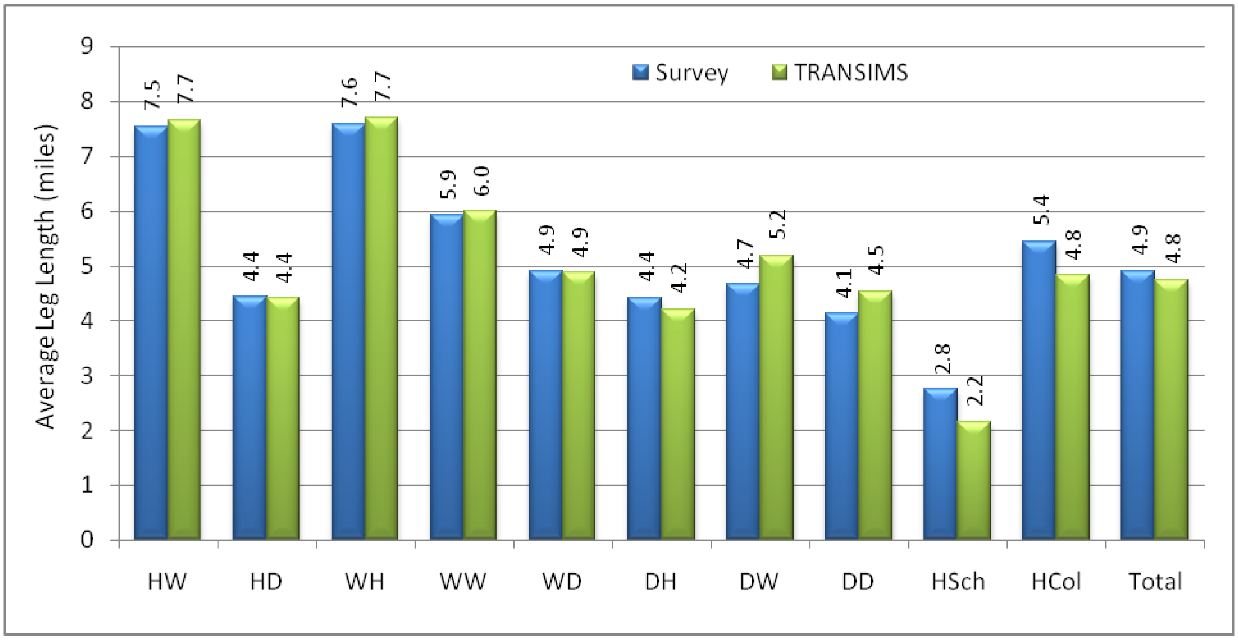
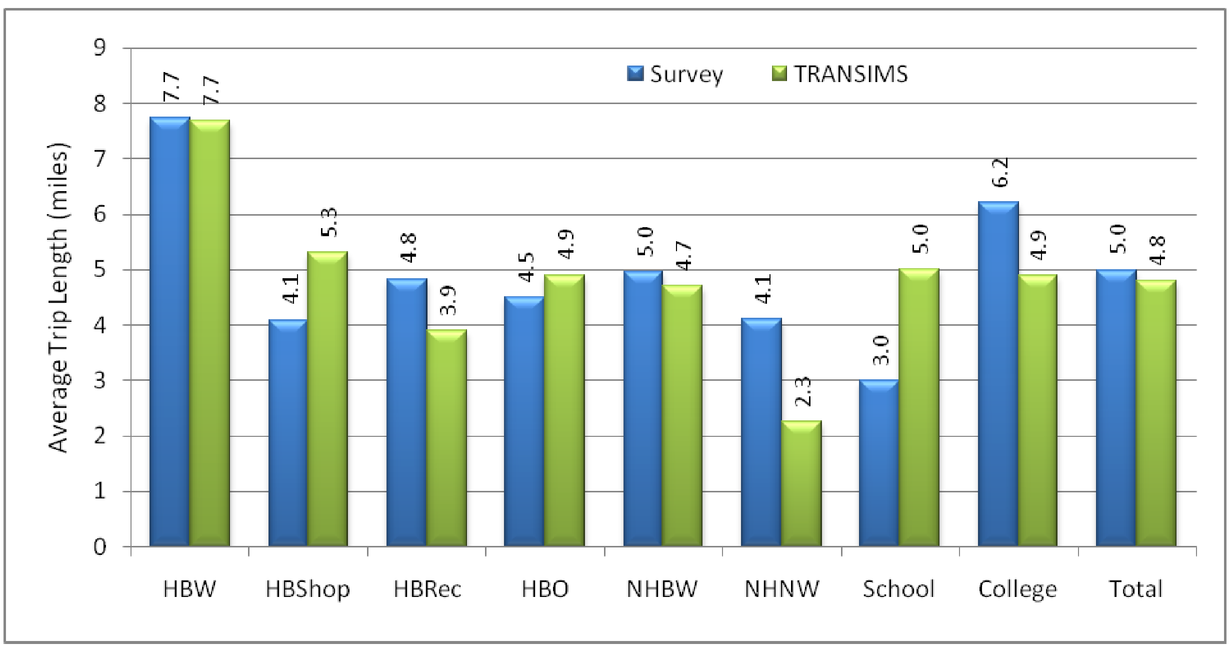


Figure 81: Average Trip Length Comparisons by Trip Type



8.5.4 Trip Length Distributions

As in the previous implementation, the trip length distributions by trip purpose were compared to the Track 1 distributions to evaluate how effectively the location choice model captured the overall shape of the distribution curve. The original comparisons showed that the work trips had significantly longer trip lengths than expected which was one of the primary reasons the simulated volumes were overestimated. As shown in Figure 82, the tail of the work trip distribution for this implementation is much closer to the Track 1 shape. Unfortunately, the short trip lengths are considerably distorted. The activity generator had difficulty assigning trips to work locations less than three miles from home. This may be a consequence of zone size, intra-zonal travel time estimates, minimum time budget constraints, and/or time adjustment factors for nearby locations. Since the other trip purposes underestimate trips of less than three miles as well, some type of algorithm improvement for short trips appears warranted.

Figure 82: Home-Based Work Trip Length Distribution Comparison

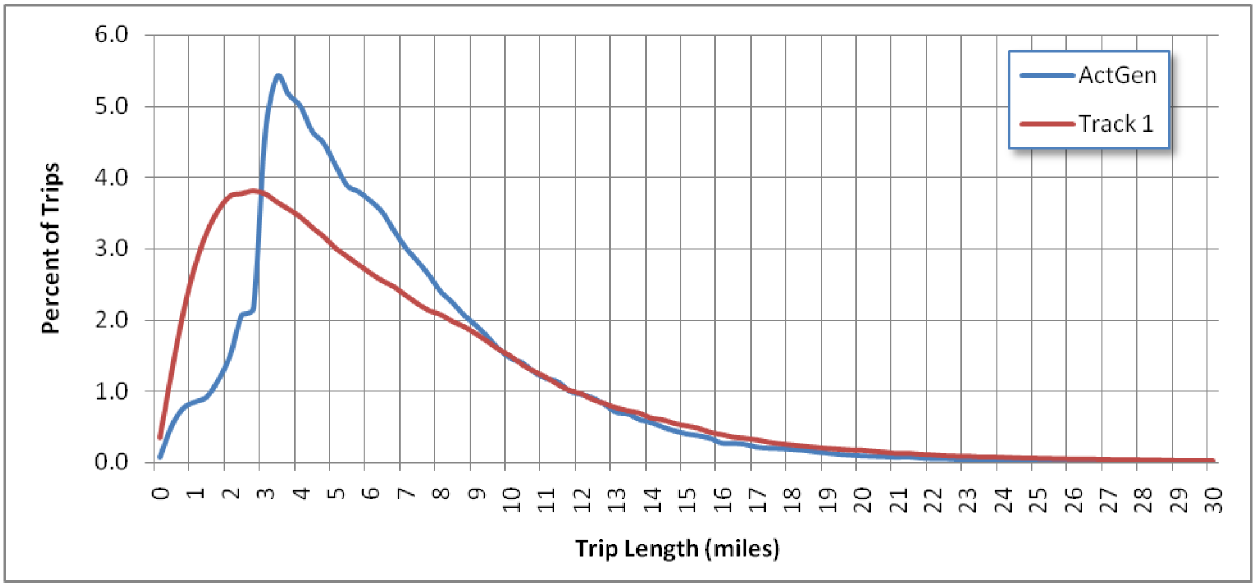
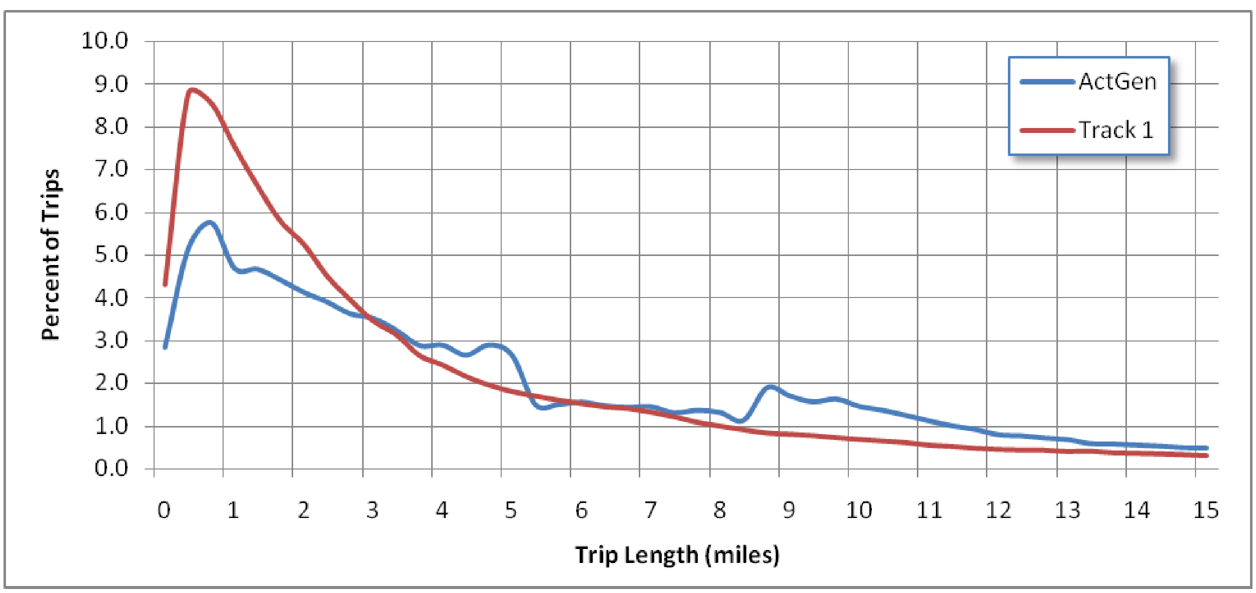


Figure 83: Non-Home Work Trip Length Distribution Comparison



For this implementation, the home-based other (Figure 84) and non-home non-work (Figure 85) trips were calibrated together as discretionary trips. Home-based other trips are by far the largest share of trips in the regional model. The distribution shown in Figure 84 is somewhat flatter than the previous calibration (Figure 63). It includes noticeably fewer trips in the one to two mile range as mentioned previously. Trips in the 9 to 10 mile range are about twice the target distribution. These results may suggest that calibrating a single discretionary model is unadvisable. On the other hand, it may also be a natural consequence of matching synthetic households to survey households that make more trips in order to replicate the overall number of vehicle trips in the region.

Figure 84: Home-Based Other Trip Length Distribution Comparison

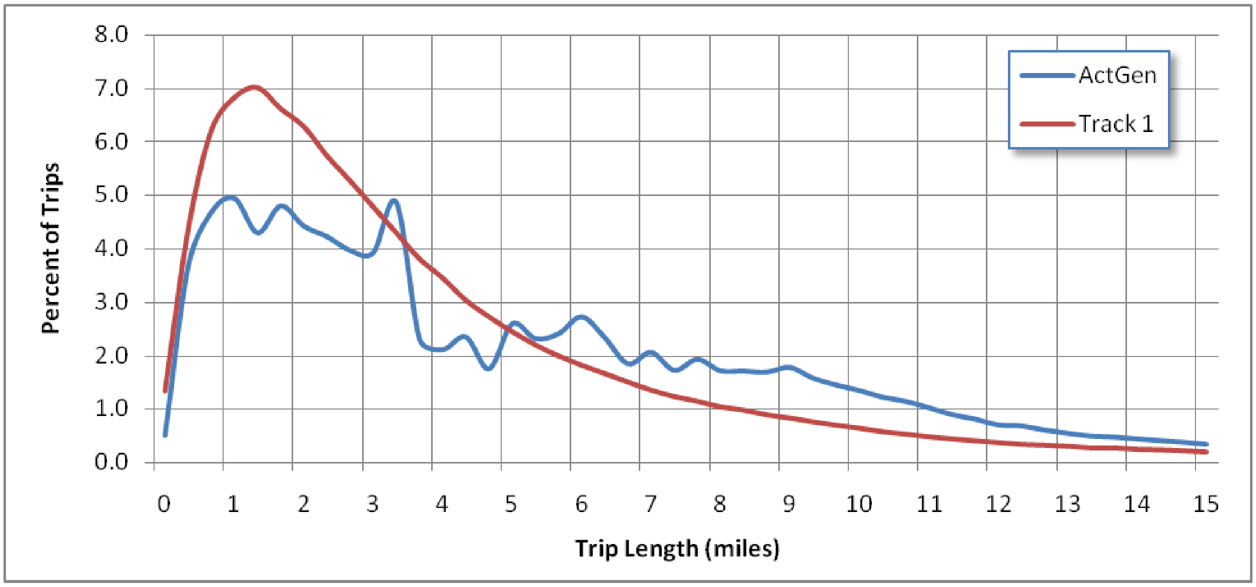
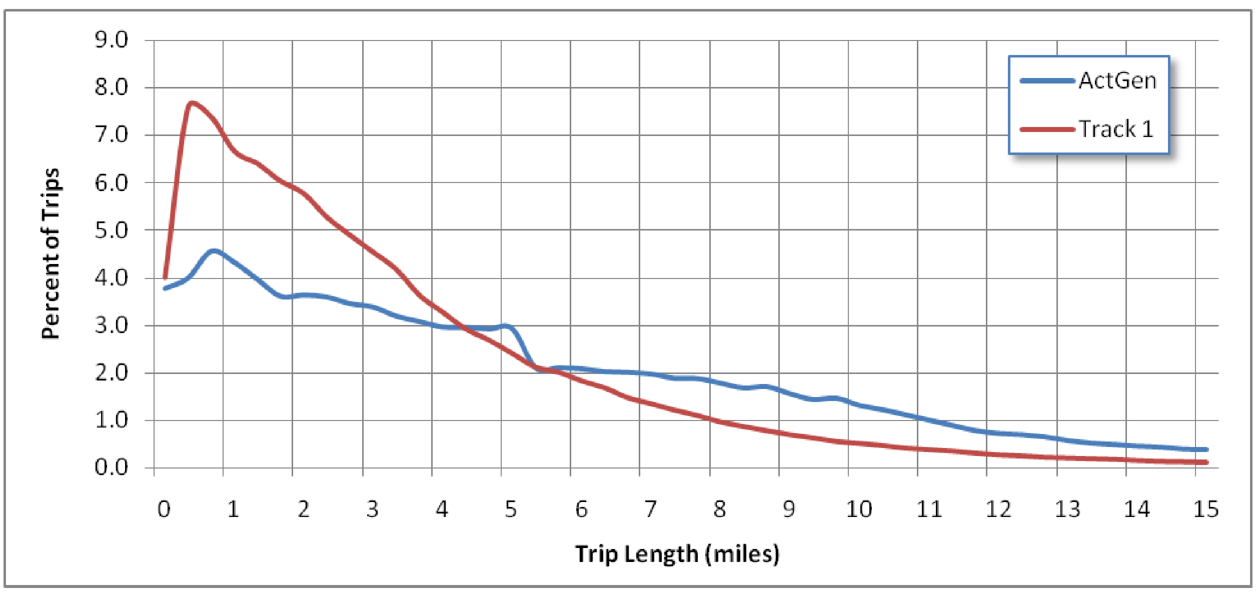


Figure 85: Non-Home Non-Work Trip Length Distribution Comparison



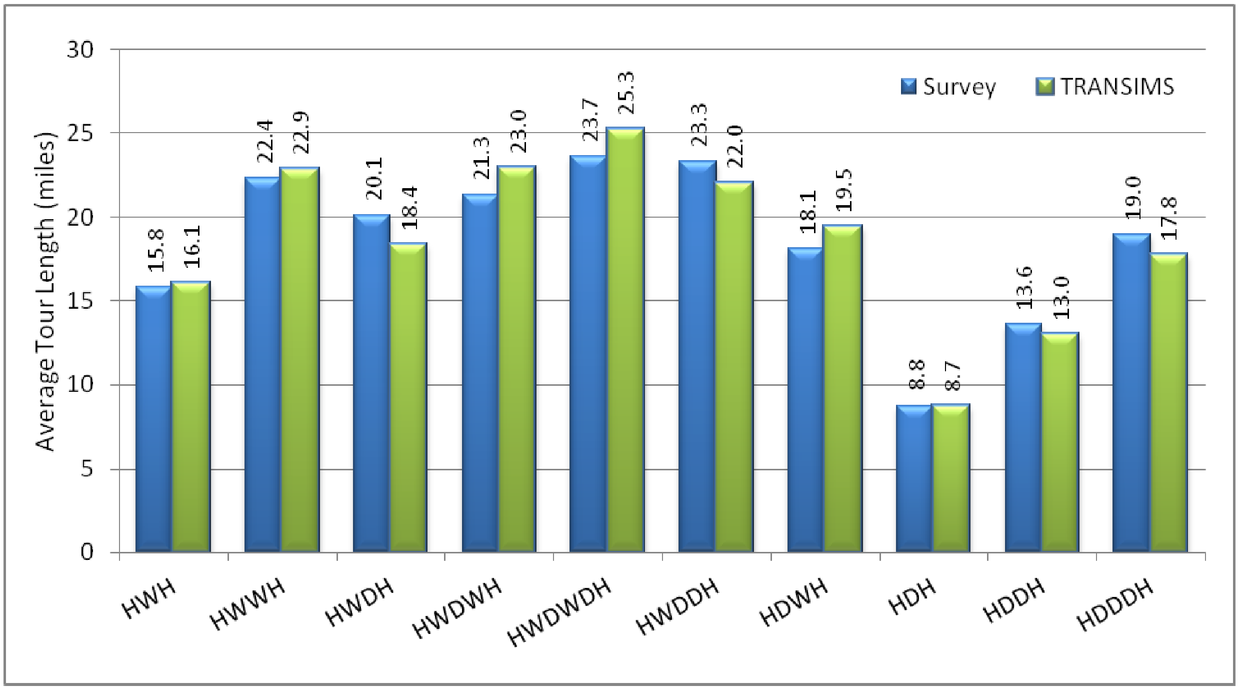
8.5.5 Tour Distance Validation

In addition to trip lengths, tour distances were checked and validated against survey data for several tour types (Table 42 and Figure 86). All of the tour types were within 9 percent of the survey average, with home-work-home, home-work-work-home and home-discretionary-home being in very close agreement. These three types plus home-discretionary-discretionary-home which is only 4.5 percent low represent 82 percent of the types of tours observed in the household activity survey.

Table 42: Tour Trip Length Comparisons

Tour Type	*Survey Frequency*	*Survey Length (miles)*	*ActGen 4.0 Length (miles)*	*Percent Difference*
HWH	3,705	15.8	16.1	1.5%
HWWH	1,221	22.4	22.9	2.4%
HWDH	717	20.1	18.4	-8.6%
HWDWH	422	21.3	23.0	7.7%
HWDWDH	105	23.7	25.3	6.8%
HWDDH	170	23.3	22.0	-5.6%
HDWH	259	18.1	19.5	7.5%
HDH	7,723	8.8	8.7	-0.2%
HDDH	1,936	13.6	13.0	-4.5%
HDDDH	785	19.0	17.8	-6.6%
Other	822	24.9	27.0	8.5%

Figure 86: Tour Length Comparisons by Tour Type and Complexity



8.6 Inter-District Flows

The results of the activity generation process were compared to the inter-district flows produced by the METRO regional model and modeled in Track 1. Table 43 shows the percent difference in market shares for work trips between the activity model and Track 1. While most inter-district pairs track reasonably well, the activity model significantly overestimated the intra-district flows. The same applies to discretionary trips, although the impact on intra-district travel is less apparent (Table 44). Inter-district college trips show a huge influx to district 1 from districts 3 and 7 (Table 45). The main reason is that the original zone attraction weights from the METRO destination choice model were mostly zero for districts 3 and 7. This explains why the intra-district flows for these districts are much less than observed in the Track 1 data. District 8 retained most of its college trips due to river crossing penalties. School trips, on the other hand, are considerably different. As seen in Table 46, the intra-district flows are grossly underestimated. Assigning trips to the nearest school could address this problem, but since school trips constitute only 0.1 percent of total trips, the model was left in place for future research.

Table 43: Percent Difference in Work Trip Distribution (ActGen - METRO)

Origin District	Destination District							
	1	2	3	4	5	6	7	8
1	4.8	2.1	0.7	-1.3	-5.6	-1.4	1.5	-0.7
2	-0.9	9.4	-4.0	-0.7	-1.2	-0.4	-2.3	0.1
3	-1.5	-3.3	16.6	-5.7	-1.8	-1.8	-2.0	-0.5
4	-0.8	-0.9	-8.1	20.8	-3.7	-5.2	-2.0	-0.1
5	2.8	0.7	-0.3	0.5	2.1	-4.7	-2.0	0.8
6	-0.8	-0.7	-1.6	-4.1	-11.8	21.9	-2.2	-0.6
7	5.2	0.0	0.8	-2.4	-6.4	-1.7	5.1	-0.6
8	-2.8	-0.9	-0.5	-1.0	-7.7	-1.9	-2.2	17.1

Table 44: Percent Difference in Discretionary Trip Distribution (ActGen - METRO)

Origin District	Destination District							
	1	2	3	4	5	6	7	8
1	4.6	-0.2	0.0	1.1	-2.4	0.1	-2.8	-0.3
2	-0.7	1.0	-5.2	2.5	4.7	0.3	-2.7	0.3
3	-0.6	-5.3	9.5	-4.9	4.5	-0.2	-3.2	0.0
4	0.0	2.1	-4.1	6.1	-0.7	-3.1	-0.4	0.2
5	-1.4	1.2	0.0	0.4	8.7	-4.6	-3.5	-0.7
6	-0.2	0.8	-0.2	-4.3	-9.4	14.4	-0.2	-1.0
7	-0.5	-2.5	-0.1	0.8	-7.8	-0.4	10.5	0.0
8	-0.2	1.4	0.0	0.7	2.6	-0.4	0.3	-4.4

Table 45: Percent Difference in College Trip Distribution (ActGen - METRO)

Origin District	Destination District							
	1	2	3	4	5	6	7	8
1	-8.0	1.1	0.4	0.8	6.9	1.5	-3.2	0.6
2	4.1	5.4	-0.5	1.6	1.7	0.5	-13.3	0.6
3	18.8	1.7	-24.5	-3.0	1.3	0.1	5.5	0.1
4	7.8	1.0	-3.4	-14.0	3.8	5.7	-1.0	0.2
5	-0.9	1.6	-1.1	-1.9	0.1	4.7	-4.0	1.5
6	0.2	0.4	0.0	-0.4	2.1	-2.3	-0.3	0.2
7	18.0	1.5	-0.8	1.0	2.6	0.3	-23.0	0.4
8	-4.6	1.4	0.0	0.5	-11.1	1.8	0.1	12.0

Table 46: Percent Difference in School Trip Distribution (ActGen - METRO)

Origin District	Destination District							
	1	2	3	4	5	6	7	8
1	-8.3	75.0	0.0	25.0	-75.0	0.0	-16.7	0.0
2	0.6	-68.9	-0.6	52.1	7.6	7.7	1.5	0.0
3	0.9	39.5	-65.3	19.7	1.5	0.3	3.5	0.0
4	0.3	39.7	10.0	-50.8	2.1	-3.3	1.9	0.1
5	0.8	32.6	3.6	49.6	-87.7	2.5	-1.4	0.0
6	0.0	33.0	2.7	13.7	4.3	-61.5	7.6	0.3
7	-1.2	25.2	0.9	56.5	-17.2	0.9	-65.1	0.0
8	0.1	32.4	0.3	23.6	4.0	4.6	11.5	-76.5

8.7 Router-Microsimulator Results

The iterative feedback process outlined in Figure 51 was used to route and simulate the trips generated by the third ActGen implementation. Unlike previous applications, the initial Router build-up (Figure 87) showed that the volumes underestimated traffic counts. Though not the most desirable result, the process at least demonstrated that the model could be further calibrated to replicate the counts.

Figure 87: Initial Router Validation for the Third ActGen Implementation

```
                   Num.  ------Volume------ ---Difference---  --Abs.Error--   Std.     %     R

 Facility Type     Obs.  Estimate  Observed   Volume      %     Avg.     %     Dev.   RMSE   Sq.

       Freeway      51   2464435   2874828  -410393   -14.3   10586   18.8    7609   23.1  0.688
    Expressway      24    562387    409706   152681    37.3    8318   48.7    8563   69.2  0.213
Major Arterial     104    902569   1143236  -240667   -21.1    3867   35.2    2682   42.7  0.539
Minor Arterial      82    289968    507048  -217080   -42.8    2945   47.6    2890   66.5  0.286
     Collector     130    230453    462788  -232335   -50.2    2102   59.0    1609   74.2  0.156
  Bridge/Other      18    177088    286236  -109148   -38.1    6792   42.7    3675   48.3  0.561

         TOTAL     409   4626900   5683842 -1056942   -18.6    4349   31.3    4919   47.2  0.885
```

Analysis of the results revealed that the Router was able to generate travel plans for about 4.3 million vehicles (close to Track 1 numbers). Previous efforts did not generate plans for more than 3.6 million vehicles. Time schedule problems continued to be the primary source of problems. This meant that the travel times used to schedule the activities were not compatible with the travel times calculated by the Router or Microsimulator.

The initial Microsimulator run (Figure 88) was able to start 99 percent of the 4.3 million trips, but could only complete 79 percent. The 618,000 wait time problems indicate that the simulation is highly congested even though the Router volumes are almost 20 percent less than the daily counts. This suggests that the time of day distribution of traffic has peaks that produce cascading queues and gridlock in the Microsimulator.

Figure 88: Initial Microsimulator Performance Statistics

```
Number of Vehicle Trips Processed = 4316272
Number of Vehicle Trips Started   = 4287234 (99.3%)
Number of Vehicle Trips Completed = 3414265 (79.1%)

Total Hours for Completed Vehicle Trips = 1504585.5 hours
Average Travel Time for Completed Trips = 26.44 minutes
Maximum Number of Vehicles on the Network = 264108 at 17:25:21

Number of Travelers with Problems = 895340 (15.9%)

Total Number of Problems = 895829
Number of Wait Time (#9) Problems = 618106 (69.0%)
Number of Departure Time (#14) Problems = 28456 (3.2%)
Number of Arrival Time (#15) Problems = 248288 (27.7%)
Number of Vehicle Spacing (#23) Problems = 979 (0.1%)
```

The TripSum program was applied to the activity file to generate a composite diurnal distribution of the trip start times generated by the activity model. Figure 89 shows the resulting distribution along with the trip start times from the expanded travel survey. It is obvious from this chart that the time reporting bias in the travel survey significantly influenced the time scheduling within the activity generator. The model smoothed out the morning trips relatively well, but failed to make appropriate adjustments to the afternoon and evening activity schedules. Dramatic surges in traffic such as these will easily overwhelm the network and create the levels of congestion predicted by the Microsimulator.

The first attempt to address this problem applied the PlanTrips program to the activity file to adjust the activity schedules based on simulated travel times. This reduced some of the problems, but failed to dampen the major surges in travel schedules. Ultimately this problem needs to be addressed at the activity generation level by including random variance in the time schedules extracted from the survey. At this stage in the progress, the only practical approach was to use PlanTrips to split the activity file into independent trips, route each trip, and then use SmoothPlans to adjust the diurnal distribution. The Router-Microsimulator feedback process was then applied based on the smoothed trip times shown in Figure 90.

Figure 89: Diurnal Distribution of Trip Start Times

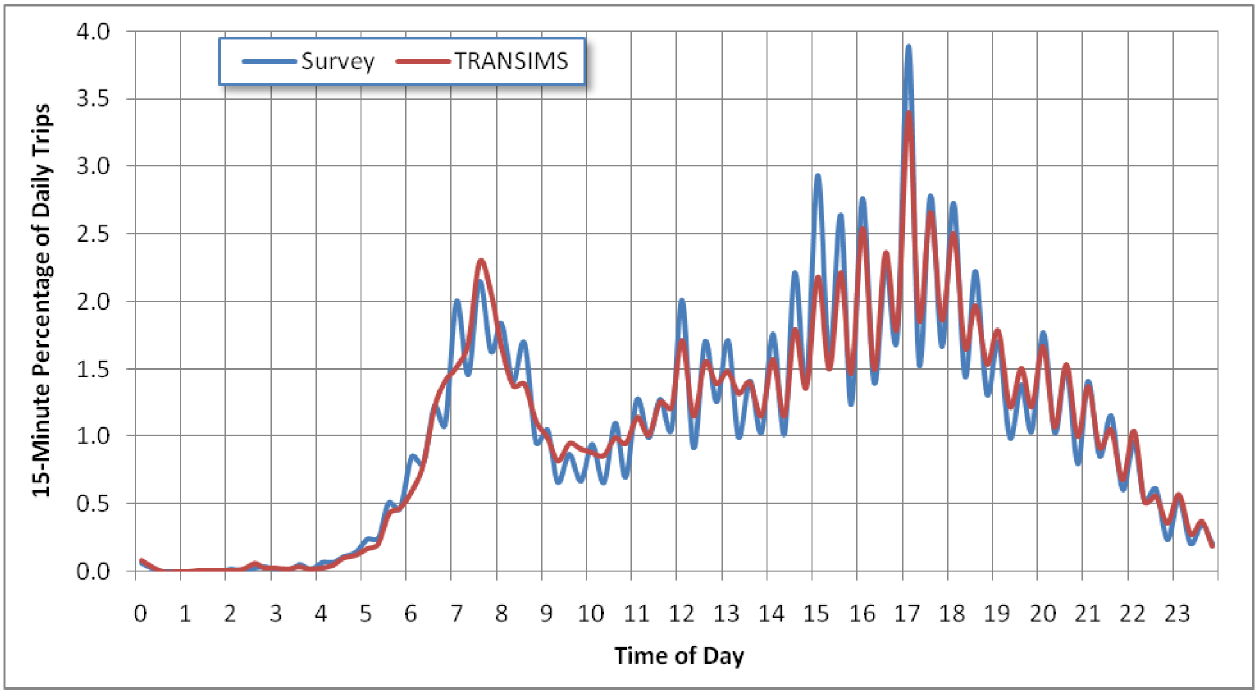
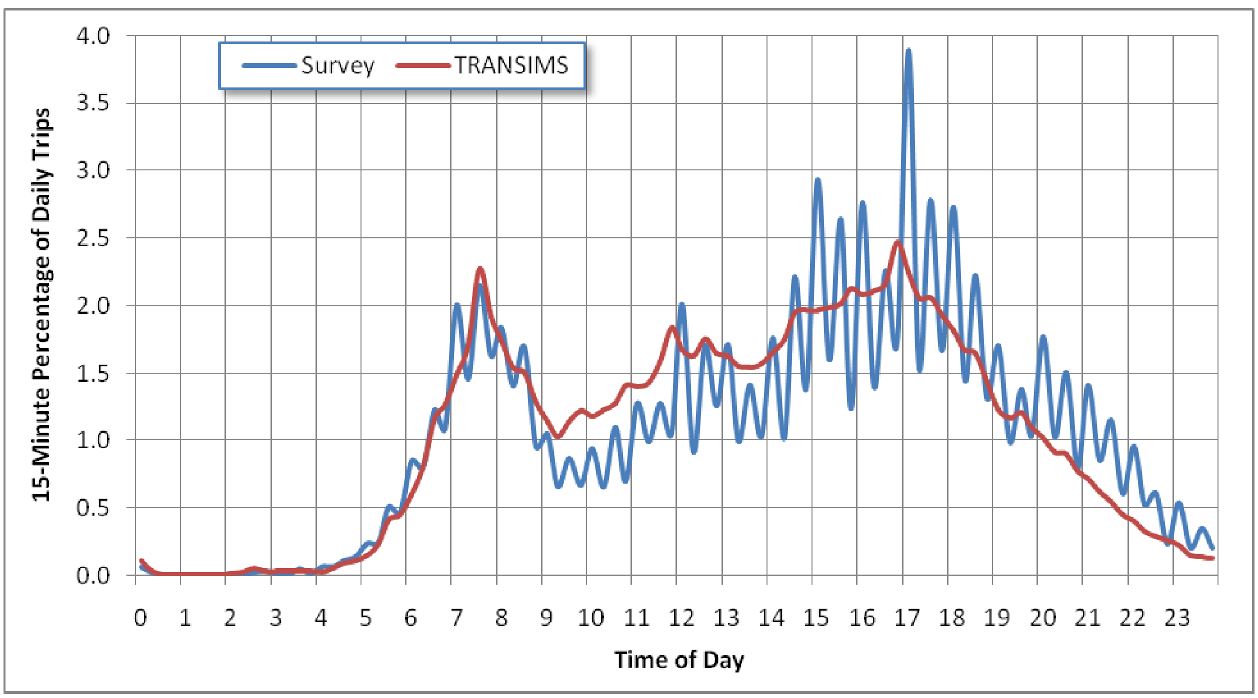


Figure 90: Smoothed Trip Start Time Distribution



After smoothing the trip start times and performing 20 feedback iterations between the Router and Microsimulator, the general Microsimulator performance statistics improved to the values shown in Figure 91. Comparing these results to Figure 88 shows that the number of completed trips increased by

670,000 or almost 20 percent and the number of simulation problems reduced by 75 percent. Notice also that the average travel time per trip dropped from 26.4 minutes to 13.0 minutes or 50 percent. The maximum number of vehicles on the network at any point in time dropped from 264,000 to 127,000 and the time of day when the maximum load occurred moved from 5:25 PM to 4:56 PM. All of these statistics point to a network that is far less congested then it was originally. The 206,000 wait time problems, however, indicate that the simulation is not fully converged and more iterations or network adjustments are needed to fully stabilize the results.

Figure 91: Microsimulator Performance Statistics after 20 Iterations

```
Number of Vehicle Trips Processed = 4316272
Number of Vehicle Trips Started   = 4302152 (99.7%)
Number of Vehicle Trips Completed = 4081242 (94.6%)

Total Hours for Completed Vehicle Trips = 882890.5 hours
Average Travel Time for Completed Trips = 12.98 minutes
Maximum Number of Vehicles on the Network = 126676 at 16:55:54

Number of Travelers with Problems = 230748 (4.1%)

Total Number of Problems = 230824
Number of Wait Time (#9) Problems = 206360 (89.4%)
Number of Departure Time (#14) Problems = 14120 (6.1%)
Number of Arrival Time (#15) Problems = 10180 (4.4%)
Number of Vehicle Spacing (#23) Problems = 164 (0.1%)
```

The encouraging simulation results are more related to the fact that there are 20 percent fewer vehicle trips generated by the activity model than was simulated in Track 1. Figure 92 and Figure 93 compare the simulated volumes after 20 feedback iterations to the daily traffic counts. The facility type results are not appreciably different from the initial Router validation shown in Figure 87. With the exception of the St. John's Bridge, all screenlines show significant negative differences. The only way the validation statistics can improve is by increasing the number of trips or the trip lengths in the activity model.

Figure 92: Facility Type Validation Results after 20 Iterations

Facility Type	Num. Obs.	------Volume------ Estimate	Observed	---Difference--- Volume	%	--Abs.Error-- Avg.	%	Std. Dev.	% RMSE	R Sq.
Freeway	51	1966578	2874828	-908250	-31.6	17809	31.6	10233	36.3	0.555
Expressway	24	401493	409706	-8213	-2.0	2947	17.3	1762	20.0	0.738
Major Arterial	104	1063278	1143236	-79958	-7.0	2662	24.2	2132	31.0	0.540
Minor Arterial	82	451700	507048	-55348	-10.9	2577	41.7	2333	56.1	0.271
Collector	130	369370	462788	-93418	-20.2	1610	45.2	1289	57.8	0.296
Bridge/Other	18	263706	286236	-22530	-7.9	3936	24.7	2185	28.1	0.649
TOTAL	409	4516125	5683842	-1167717	-20.5	4272	30.7	6517	56.0	0.897

Figure 93: Screenline Validation Results after 20 Iterations

Link Group	Num. Obs.	------Volume------ Estimate	Observed	---Difference--- Volume	%	--Abs.Error-- Avg.	%	Std. Dev.	% RMSE	R Sq.
R-1 St. John's Br	2	24675	22610	2065	9.1	2100	18.6	1460	20.7	0.000
R-2, William. River Bridg	16	378477	464966	-86489	-18.6	7330	25.2	6980	34.3	0.870
R-3, Sellwood Bridge	2	22748	31382	-8634	-27.5	4317	27.5	3465	31.6	0.000
R-4 OrCty 205 Br	4	95924	109571	-13647	-12.5	6224	22.7	6221	30.1	0.903
R-5, I-5 Bridge	2	71279	113838	-42559	-37.4	21280	37.4	825	37.4	0.000
R-6 I-5 Hayden	2	80733	128300	-47567	-37.1	23784	37.1	926	37.1	0.000
R-7 I205	2	51082	104074	-52992	-50.9	26496	50.9	1468	51.0	0.000
E-9 W of 122nd	34	281747	348370	-66623	-19.1	2843	27.8	3650	44.7	0.848
E-21 W of Tacoma	20	217942	272032	-54090	-19.9	4686	34.4	8326	68.9	0.895
W-7 Westhills	10	128999	172226	-43227	-25.1	4593	26.7	7167	47.6	0.968
W-09 Tigard Int	6	104065	154962	-50897	-32.8	8944	34.6	8400	45.6	0.930
W-16 NW of Tigard	10	116359	140922	-24563	-17.4	3690	26.2	4525	40.2	0.976
TOTAL	110	1574030	2063253	-489223	-23.7	5682	30.3	7489	50.0	0.909

Chapter 9 Future Research

This project investigated a significant number of strategies for implementing activity generation and integrating activity data with time-dependent network simulations. An important lesson learned was that calibrating a location choice model simply based on average trip lengths by trip or leg type is insufficient for ensuring that the activity patterns are temporally or spatially realistic. The shape of the distribution and the capacity of the destination are critical components of the location choice process.

This study also found that activity patterns with predefined schedules require some level of time budget constraints to make the coordination of activities feasible for complex tours. In addition, the travel time data used to select locations and schedule activities needs much higher fidelity than zone-to-zone skims can provide. The methods developed by this project to include zone range factors in the travel time estimation proved useful in overcoming the limitations of zone skims. The research also suggested that improvements to this methodology may be warranted for trips of three miles or less.

In the end, the project failed to find the exact set of parameters and coefficients that generated the correct number of vehicle trips between the correct locations at the correct time of day. It did demonstrate that the correct values could be bracketed by the TRANSIMS methodology, but was unable to refine the solution within the project schedule. The following are a few ideas about how this process could be improved and streamlined to achieve convergence more expeditiously.

9.1 Potential Process Improvements

As noted earlier, the time reporting bias inherent in all travel surveys was copied to the activity schedules of the synthetic households without sufficient smoothing. This created huge spikes in network demand at specific times of day that the Microsimulator could not accommodate. Smoothing methods were applied to the activity generation results, but the preferred solution would be to integrate a schedule randomization process into the activity generator from the onset. The simplest way of implementing this concept is to apply a single random time shift to all activities of the household. More sophisticated algorithms could introduce randomness for individuals, tours or activities within tours. This would require close coordination of shared travel activities or vehicle usage, but would improve the trip length distributions and the shape of the overall diurnal distribution.

This study was designed with the objective of testing the interface between the activity generator and the Router-Microsimulator procedures. In this regard, the methods and procedures for handling tours could be significantly improved. The fact that the Microsimulator breaks down tours into individual trip legs and rigidly checks the leg constraints leads to situations where differences between the Router time estimates and the Microsimulator travel times generate problems that might easily be resolved if the Microsimulator executed the full tour as a single element. This would enable the Microsimulator to start the next trip earlier than planned if the travel time to the activity proved less than the Router anticipated, or the next trip might start later if the travel time is longer than expected. Of course, there needs to be some limits to how much the Microsimulator distorts the original plan before a problem is registered and corrective actions are taken.

Another concern is that the procedures used to select households for re-routing are inherently trip-based and not tour-based. For example, if a tour has four trips that meet their associated time constraints, but one of the trips has a sub-optimal path, PlanCompare will select this household as a candidate for re-routing, thus leading to building new paths for all four trips. Furthermore, when a household is selected for re-route, the Router builds new paths for all tours by all household members. In many cases, this generates a significant amount of unnecessary path building that could introduce instability and slow down the convergence process. The rationale behind this approach is that there may be interactions between trips, tours, and household members that need to be coordinated during the path building process. This complexity is one of the reasons the original GEN2 model chose to assign a unique household ID to each tour so the tours could be independently modeled.

9.2 Time Schedule Optimization for Tour-Based Models

One of the problems with the current activity routing process is that the time schedules are fixed by the activity generator with relatively generic information able the travel times between the activity locations. If the Router finds a problem with the travel time estimates, the only option for correcting the estimate is to adjust the trip arrival time. If these adjustments conflict with the start time of the next trip, additional problems and inconsistencies are created.

A potential solution to this problem is to provide the Router with additional flexibility in scheduling activities for the whole tour. This would enable the Router to coordinate the scheduling of each trip with actual point-to-point travel times. It is also envisioned that the Router could optimize the trade-offs between travel time and schedule variance during the routing process. A time of day choice model could select the set of travel times by time of day and schedule delays that provides the optimum utility for the whole tour.

The methodology is outlined below.

Let $S = \{1,..., j,..., J\}$ be the set of stops along activity pattern h; A_j = actual arrival time at stop j; P_j = preferred arrival time at stop j; and SD_j = schedule delay associated with stop j (the difference between actual and preferred arrival times.). Two types of schedule delays are defined:

Early schedule delay (ESD_j)

$$ESD_j = P_j - A_j, \qquad A_j < P_j \tag{22}$$

Late schedule delay (LSD_j)

$$LSD_j = A_j - P_j, \qquad A_j > P_j \tag{23}$$

Additionally, one can specify an interval of time, ε_j, during which, delays are tolerated:

$$LSD_j = \begin{cases} 0 & |A_j - P_j| < \varepsilon_j \\ A_j - (P_j + \varepsilon_j) & \text{otherwise} \end{cases} \tag{24}$$

$$LSD_j = \begin{cases} 0 & |A_j - P_j| < \varepsilon_j \\ P_j - (A_j + \varepsilon_j) & \text{otherwise} \end{cases} \tag{25}$$

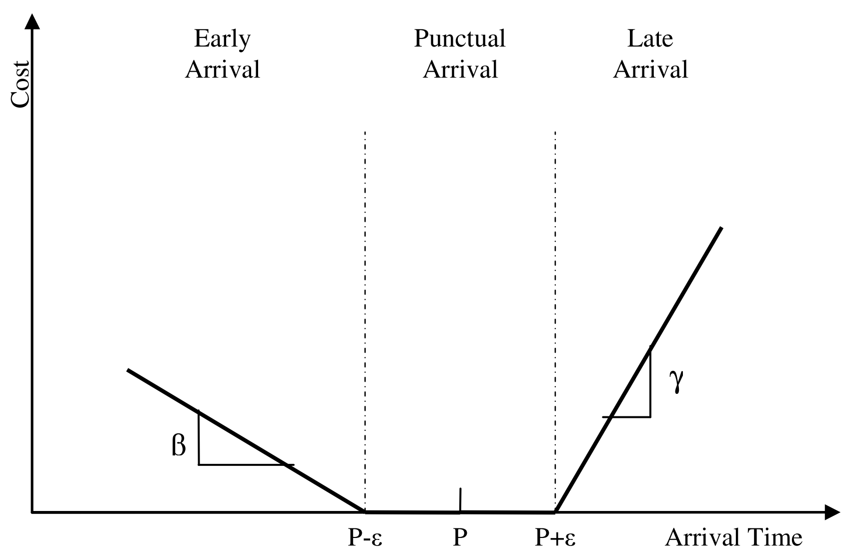
Equation (26) defines the generalized cost associated with activity pattern h and departure time τ.

$$G_h^\tau = \sum_{j=1,\, j\in S}^{J-1} \alpha.T_{j,j+1} + \sum_{j\in S} \left(\beta.ESD_j + \gamma.LSD_j\right) \tag{26}$$

where , $\beta < \alpha < \gamma$, α is the cost of travel time to the user, β is the cost per unit time of early delay, and γ is the cost per unit time of late delay. Figure 94 graphically demonstrates the concept. Prior studies estimated the parameter values as follows:

Value of time distribution: N(0.4, 0.2), [0.01, 3.0] (Lam and Small, 2001);
Value of ESD distribution: N(0.3, 0.15), [0.01, 2.0] (Small (1982);
Value of LSD distribution: N(1.8, 0.6), [0.25, 4.0] (Small (1982).

Figure 94: Time Schedule Optimization Model



The user equilibrium optimality conditions is expressed in terms of the general cost G_h^τ as:

$$f_h^\tau.\left(G_h^\tau - G_h^{\tau^*}\right) = 0 \tag{27}$$

$$\left(G_h^\tau - G_h^{\tau^*}\right) \geq 0 \tag{28}$$

$$f_h^\tau \geq 0 \tag{29}$$

where $G_h^{\tau^*}$ is the least general cost associated with activity pattern h at time τ.