Metropolitan Washington Council of Governments
National Capital Region Transportation Planning Board

Results of FY 2007 Travel Forecasting Research
November 16, 2007

Vanasse Hangen Brustlin, Inc.
8300 Boone Blvd, Suite 700
Vienna, VA 22203
(703) 847-3071

This page intentionally left blank

Table of Contents

Table of Figures

This page intentionally left blank

Introduction

This report combines into a single document the technical memoranda prepared by VHB for TPB during its FY 2007 work program for research in travel demand forecasting. The work tasks for this time period covered herein are:

1. Attend Meetings and Assess TPB Work Program in Models Development and Data Collection
2. Review of the Federal Transit Administration's (FTA) Summit Software
3. Review the use of Dynamic Traffic Assignment (DTA) and traffic simulation models in support of regional planning activities
4. Research the State of the Art in equilibrium assignment
5. Research techniques for peak spreading analysis
6. Research the use of cutlines for model validation
7. Review other MPOs' experience with feeding back the results of a nested logit mode choice model into trip distribution

Rich Roisman (items 1, 4, 6, and 7), Paul Gilliam (items 3 and 5)*, and Dan Goldfarb (item 2)[*] served as the primary authors, with assistance from Tim Brulle, Nitesh Gupta*, Dalia Leven, Kindra Ochoa*, Joe Ojeda, Maggie Qi, Bryon White* and project management from Phil Shapiro and Frank Spielberg.

[*] No longer with VHB.

This page intentionally left blank

Task 1 -- Attend Meetings and Assess TPB Work Program in Models Development and Data Collection

This page intentionally left blank

At TPB's direction, VHB staff have been attending and participating in meetings of the Travel Forecasting Subcommittee and other TPB bodies over the past fiscal year. TPB also requested that VHB provide an assessment of the TPB work program in models development and data collection to provide guidance for future work program activities in upcoming fiscal years.

Overall, VHB believes that TPB's work program in this area is sound, and recognizes that TPB must continue to meet ever-increasing demands within an increasingly constrained budget; however, there is always a need for new initiatives to improve TPB's ability to best serve the travel forecasting and travel monitoring needs of the National Capital Region. The following represent areas where VHB suggests that TPB consider additional work in models development and data collection in the upcoming work program:

1. **Continue development of Peak Spreading Model.** FY Task 5 was a preliminary review of techniques to forecast peak spreading over time. One result of this task was to identify several approaches for peak spreading. Another was to prepare a preliminary inventory of traffic counts that can be used to define existing peaking characteristics, and serve as a basis of modeling future peak spreading. It is recommended that the regional model approach identified in Task 5 be tested initially as this would require much less effort and results could be presented several months after project initiation.

2. **Changes in Travel Behavior due to an Aging Population.** TPB needs to consider now what will happen to trip making characteristics as the baby boomer generation moves into retirement. The baby boomers make up a considerable segment of the regional population and while many of them will leave the area, a majority will probably remain here due to the attractiveness of the region. Travel behavior for the baby boomers will be considerably different over time as there will be more part-time workers and more non-work trips. By looking at the behavior of the retired sub-group now, TPB can get some idea of what will happen when the group gets larger, but continued research is probably needed. As with other aspects of their lives, in travel behavior the boomers will not follow in their parents' footsteps.

3. **Airport Trip Generation.** The major airports in and around the TPB region have undergone improvement and expansion projects in recent years, some of which (such as Dulles International Airport) are not yet complete. As the nature of the airports change, so do their trip generation characteristics and TPB should consider how these changes impact its process. There has been work by others on this issue that TPB may be able to build upon. Some trip generation rates are documented in the Federal Highway Administration (FHWA) publication *Intermodal Ground Access to Airports- A Planning Guide*; however, much of that information is now at least a decade old.

4. **Further Research on External Trips.** In addition to looking at the absolute number of forecast external trips, the purpose and direction of trips should also be considered.

As exurban development grows, the proportion of external productions is likely to grow while the proportion of external attractions is likely to decrease. The relationship between external-external (E-E) and external-internal (E-I) trips may also change. TPB should investigate the regional impact of these potential changes.

5. **Interaction between the greater Baltimore and Washington Regions.** Over time the Baltimore and Washington Metropolitan regions continue to meld into one megalopolis. Today the region stretches from south of Fredericksburg and continues north into Pennsylvania; and east west from the eastern shore of Maryland, Delaware and Virginia to the Shenandoah Valley. As the region continues to become economically stronger this expansion will increase. There have been attempts over the last three decades to develop an integrated modeling approach for the two regions that have had limited success. The availability of the new Household Survey for both regions may provide a unique opportunity to develop a modeling approach that better integrates travel between the two regions.

Task 2 -- Review FTA Summit Software

This page intentionally left blank

The Summit software program developed by FTA is used to prepare information for evaluation of transit projects that are seeking funding under the New Starts program. Summit results are used as part of the evaluation for New Starts applications. Summit measures the difference in user benefits between different transit alternatives. It compares a baseline to a build alternative, which is usually, but not necessarily, a Transportation Systems Management (TSM) type alternative. The software is basically a matrix manipulation type of application. It focuses on the calculations used in the mode choice model at the traffic analysis zone (TAZ) level and calculates the benefits of the build scenario versus the baseline. It then aggregates this information into user-defined districts for the study area and allows for comparison between the build and baseline alternatives. Districts are generally defined as smaller and denser around the proposed transit project, while in outlying areas they include a larger area.

The objectives of this chapter are to:

- Document the structure of Summit.
- Document some applications of Summit.
- Discuss some key factors affecting its successful application.
- Discuss its application in the current TPB forecasting framework.

Summit Structure

Summit currently has a Windows interface, but it was originally written to run in DOS. It can be run through the new Windows interface or via a command prompt. It is a FORTRAN-based program and there are control files the user can update with the necessary input file information and user options.

The input data includes:

- Mode choice model outputs.
- Zone attributes related to accessibility.
- Zone to district equivalence table.
- Control files for execution.

The required mode choice outputs include the logsum of the logit equation used in the model, the in-vehicle time coefficient, and the trip tables. The zone attributes include the type of transit access for each zone. There are three types of access markets for each zone: CW (can walk), MD (must drive), and NT (no transit). Zone-to-zone interchanges are referred to as the access market segmentation. The zone-to-district equivalence table is for aggregating the results. It is similar to the data required for a “data squeeze,” which is an aggregation of the zone data. The control file syntax is listed in the user’s guide. It designates the location and names of the input files and file prefixes for naming and identification. Most input files must be in a binary format. The Summit user’s guide describes the file format and required headers for each file. The zone-to-district equivalence file is a text file, as are the control files.

Summit produces several important outputs. The products of a run are:

- A summary report file.
- User benefits file.
- A trip length file.
- Zone based vector files for geographic information systems (GIS) applications.

The user-benefits file is summed into the user-defined district level. There is also a user-benefits file by zone for productions and attractions. This file is useful for GIS applications. The trip length shows the change in trip length between the baseline and build. It is not overly useful because this data is difficult to interpret in relationship to user benefits. The file details trip length changes while user-benefits are calculated as a measure of time. A spreadsheet program can be used to manipulate and graph the trip length results.

The key Summit output is the table of user benefits. User benefits are a utility expressed in units of time. The development of this measure is based on the concept of consumer surplus. It measures the change in service and the change in price, which is represented as time for this calculation. It is similar to travel time savings but has several key advantages. It is sensitive to changes in both travel times and travel costs, and it recognizes benefits for both existing transit users and new users diverted from other modes. User benefits represent direct transportation benefits as a single unit of measurement and permits them to be totaled. They reflect an overall measure of transportation benefits in the alternatives evaluation.[1]

Calculate the user benefits by converting the denominator of the logit model to equivalent minutes of in-vehicle time based on the locally determined coefficient. That product is then multiplied by the number of trips for that alternative. The user benefits are then equal to the difference of that product from the baseline alternative. The user benefits are expressed in terms of minutes and reflect all i-j interchanges.

$$\text{User Benefits} = \left(\frac{\text{Price}_{ALT}}{C_{IVT}} \times \text{Trips}_{ALT} \right) - \left(\frac{\text{Price}_{BAS}}{C_{IVT}} \times \text{Trips}_{BAS} \right)$$

Figure 2-1: Summit User Benefit Equation

$\text{Price}_{ALT} = \ln(\exp(U_{auto}) + \exp(U_{bus}) + \exp(U_{rail}))$ for the build alternative
$\text{Price}_{BAS} = \ln(\exp(U_{auto}) + \exp(U_{bus}) + \exp(U_{rail}))$ for the base alternative
C_{IVT} = coefficient on in-vehicle time
Trips_{ALT} = Trips for each i-j Interchange for the build alternative
Trips_{BAS} = Trips for each i-j Interchange for the base alternative

The user benefits are reported in terms of totals for access market segmentation as well as at the zonal and district levels for both production and attraction ends.

[1] Summit Users Guide, FTA, April 13, 2006, Washington, D.C.

Summit Applications

Summit compares two transit alternatives and computes the change in user benefits for the build alternative versus the base. It allows for a quantitative analysis of transit alternatives and also for comparison of user benefits across different projects to help determine which projects should get FTA New Starts funding. For each project, Summit compares the user benefits from the build alternative to the user benefits from the baseline alternative. The baseline alternative is usually the TSM alternative, but it does not have to be. For New Starts submittals, the baseline alternative must have service coverage similar to the build alternative. For application submittals, the baseline alternative is approved after the submittal. The key difference between the baseline alternative and the build alternative for many New Starts applications is the baseline does not include a fixed guideway system; the build alternative does have a fixed guideway system.

The Summit results will show positive user benefits when the build alternative shows improved travel time. Currently, Summit only looks at transit-related benefits. It does not consider potential highway benefits resulting from a transit alternative. Transit-related highway benefits are not evaluated because of the instability in the application of the equilibrium highway assignment algorithm. This instability is related to issues with using an equilibrium assignment algorithm to reach convergence between iterations. Because of this instability, it is difficult to measure benefits from highway trips shifting modes to transit.

Summit is a very useful tool for identifying coding and path problems with the transit networks. Summit will produce questionable and illogical results when there are problems in the access coding as well as the mode choice coefficients. The GIS plots of the user benefits by TAZ help to show where there are network or service issues. For example, if there is a new Light Rail Transit (LRT) line and zones adjacent to the LRT show negative user benefits, then it is useful to examine why with the new service and access to that specific zone user benefits decreased. In many cases these types of issues relate to the access coding for those zones. It could be that walk access links under the build alternative were not coded. It might be that the zone was previously drive-only access, and under the LRT scenario walk access links were not coded. The drive access under the build alternative then created a longer path than in the baseline. Summit is very good at highlighting these types of issues within the networks. The challenge is to examine those issues and determine what caused them.

Summit does not work well with step functions or cliffs. Cliffs are artificial barriers represented in the model, but which are not so clear in actuality. For example, in some models walk distance is determined to be a specific distance from the transit access node. For zones on the edge of this walk-shed, a change to a fixed guideway could force trips to go from “walk access to a direct bus” to “must drive to a rail station.” Under the baseline alternative, the bus served the neighborhood with greater access, but slower travel times.

The new transit service is just slightly farther away, but provides faster service. Summit analysis often highlights these types of cliff issues.

Summit is a very useful tool for analyzing networks and mode choice results. It catches errors that may have been overlooked in the past, but it requires familiarity with the inner workings of the mode choice model. There are issues with the mode choice model that can result in illogical user benefits. When there are multiple modes, the results can be very sensitive to slight changes in service. Summit uses paths for all modes and access markets. Small, unforeseen changes in access can create large changes in the user benefits. An example of this is with commuter rail, which in some models has larger bias coefficients. Changing a path from commuter rail to a new build alternative that might include a bus-to-rail transfer to access the new rail can show negative user benefits. The factoring of perceived time can show a shorter actual path being a longer path.

Figure 2-2 is a GIS plot of user benefits for a light rail extension project. This type of plot is one of the more useful Summit outputs. Summit produces files with user benefits by zone that can be input as a field into a GIS file. The files have production-end user benefits and attraction-end user benefits by zone. Summit is not GIS software, but reviewing the results using GIS is very useful and it is required for New Starts submission.

Figure 2-2: GIS Plot of User Benefits

The project in Figure 2-2 involves an extension of an existing LRT line by roughly 3 miles and the addition of four new stations. The GIS plot shows the user benefits at the zonal level for home-based work productions. Green shades show a gain in benefits; red shades show a loss in benefits. Review of the results raises some key questions for this application. For the zones to the southwest of the extension, why was there a loss in user benefits? It would be expected that the new service would provide those zones with increased user benefits. What happened is that with the extension of the light rail, under the build alternative, the project sponsor eliminated the express commuter bus service to the downtown area. Zones that previously had direct service to downtown on the commuter bus line now had to transfer to rail. Other corridor zones demonstrating lower gains in user benefits were caused by similar changes in service, as well as coding errors.

Review of these results raised another question about the gain in benefits on the other side of the downtown. These benefits are not likely to be related to the service extension. These benefits were related to the added service on the light rail continuing from the extended line to another existing line north of downtown. These benefits were questioned in the review because service could be increased on similar existing lines without the proposed extension. The analysis in Figure 2-2 was counting benefits not directly related to the extension being submitted for New Starts. The increase frequency on the other lines did not depend on getting the extension. This is an example of how Summit can be valuable as an analytical tool to identify potential problems with the coding and service being proposed.

These issues highlight the need to ensure user benefits are directly related to the proposed transit investment. The build alternative should not inadvertently reduce service. Figure 2-2 is a good example of this issue. The problem is common in outlying areas where commuter bus service suffers when new rail service is proposed and the commuter bus service is replaced with feeder service to the station. The new path represents not only a slower path, but also a transfer and additional wait time.

Another Summit output is the summary of user benefit calculations, which is a very useful data table. It can be used to evaluate whether there are any issues in the process and where the benefits are by access market segmentation. The user-benefits table summarizes 50 tables produced by Summit. Tables 1 through 10 show the total trips for the baseline alternative. Tables 11 through 20 show the number of total trips for the build alternative. Tables 21 through 30 show the number of transit trips for the baseline, and Tables 31 through 40 show the number of transit trips for the build alternative. Tables 41 through 50 show the user benefits by access market segmentation. Table 2-1 is an example summary table:

Table 2-1: Sample Output from Summit

Summary of User Benefit Calculations

Table	Contents	Conditions	Markets		Total
1	trips	all	BASE	CW-CW	928474 trips
2	trips	all	BASE	CW-MD	0 trips
3	trips	all	BASE	CW-NT	0 trips
4	trips	all	BASE	MD-CW	0 trips
5	trips	all	BASE	MD-MD	1177997 trips
6	trips	all	BASE	MD-NT	0 trips
7	trips	all	BASE	NT-CW	22 trips
8	trips	all	BASE	NT-MD	0 trips
9	trips	all	BASE	NT-NT	2621413 trips
10	trips	all	BASE	TOTAL	4727935 trips
11	trips	all	ALT	CW-CW	928474 trips
12	trips	all	ALT	CW-MD	0 trips
13	trips	all	ALT	CW-NT	0 trips
14	trips	all	ALT	MD-CW	0 trips
15	trips	all	ALT	MD-MD	1177997 trips
16	trips	all	ALT	MD-NT	0 trips
17	trips	all	ALT	NT-CW	22 trips
18	trips	all	ALT	NT-MD	0 trips
19	trips	all	ALT	NT-NT	2621413 trips
20	trips	all	ALT	TOTAL	4727935 trips
21	trips	trn	BASE	CW-CW	260444 trips
22	trips	trn	BASE	CW-MD	0 trips
23	trips	trn	BASE	CW-NT	0 trips
24	trips	trn	BASE	MD-CW	0 trips
25	trips	trn	BASE	MD-MD	61372 trips
26	trips	trn	BASE	MD-NT	0 trips
27	trips	trn	BASE	NT-CW	0 trips
28	trips	trn	BASE	NT-MD	0 trips
29	trips	trn	BASE	NT-NT	0 trips
30	trips	trn	BASE	TOTAL	321816 trips
31	trips	trn	ALT	CW-CW	261477 trips
32	trips	trn	ALT	CW-MD	0 trips
33	trips	trn	ALT	CW-NT	0 trips
34	trips	trn	ALT	MD-CW	0 trips
35	trips	trn	ALT	MD-MD	61439 trips
36	trips	trn	ALT	MD-NT	0 trips
37	trips	trn	ALT	NT-CW	4 trips
38	trips	trn	ALT	NT-MD	0 trips
39	trips	trn	ALT	NT-NT	0 trips
40	trips	trn	ALT	TOTAL	322920 trips
41	userbens	total		CW-CW	90931 minutes
42	userbens	total		CW-MD	0 minutes
43	userbens	total		CW-NT	0 minutes
44	userbens	total		MD-CW	0 minutes
45	userbens	total		MD-MD	2835 minutes
46	userbens	total		MD-NT	0 minutes
47	userbens	total		NT-CW	218 minutes
48	userbens	total		NT-MD	0 minutes
49	userbens	total		NT-NT	0 minutes
50	userbens	total		TOTAL	93983 minutes

It is important to ensure that total number of person trips remains constant. The total number of trips in the build alternative must equal the total trips in the baseline alternative. The build alternative cannot have a different land-use. Land-use assumptions related to transit-oriented development cannot be assumed for the build alternative and excluded in the baseline alternative. A mode shift is expected with the build alternative, because the build alternative should provide better service resulting in a higher mode share over the baseline.

The fifth column of this report shows access market segmentation. The first two letters refer to the access market in the baseline alternative. The second set of letters refers to the access market in the build alternative. Thus in Tables 21 through 30, there can be no trips where the first access market segmentation is NT. For Tables 31 through 40, there can be no trips where the second access market segmentation is NT. If there are trips in these tables, there is a problem with the network or input data. These market segmentations provide information on how the access markets shift between alternatives.

In the previous table, the build alternative moved few trips from one access market segmentation to another. This shows that the baseline alternative provides equal access to the build alternative. Table 37 shows that the build alternative did provide transit access to four new trips. As can be expected, the build alternative did attract most of the new transit trips from the CW-CW access markets.

A review of Tables 41 through 50 shows the total user benefits. In the current software release this is equal to the transit-user benefits because there are no highway-user benefit calculations. Here, the user benefits should correspond to access markets that included trips. The four new trips from the NT-CW access market produced 218 minutes of user benefits, but most of the benefits came from the CW-CW access market segmentation. This report came from an analysis of a transitway for a relatively high-density urban setting. Therefore, it is expected to show large increases in the CW-CW market. If it was a commuter rail alternative, then changes in the MD-MD might show the highest benefits. In reviewing the user-benefit results, it helps to have a good understanding of what type of system is being tested and what the expected access market segmentation benefits might be.

When examining where user benefits occur, it is useful to look at the zonal level GIS plots and review the district-to-district tables. These tables provide, at an aggregate level, the user benefits by purpose for production and attraction ends, and they can help identify potential problems related to network errors or service deficiencies. They are useful in conjunction with the GIS data.

Good Practice - Key Factors Affecting Successful Application

Summit is a useful tool for many reasons, but if a mode choice model does not fit the parameters expected by FTA, then the modeler needs to supply supporting data to FTA showing why the model is designed and calibrated a certain way. FTA will review model assumption or calibration parameters and may accept them when there is adequate supporting data.

There are key factors to emphasize in FTA's reviews of Summit runs:

- Consistency between alternatives is very important. The access and coverage of the build alternative must be matched with the baseline alternative. If this is not so, then questions will be raised when the access market segmentation is reviewed.

- The trip tables must be consistent. The number of trips cannot change between the baseline and build alternative. Trips can shift modes, but the total number of trips can not change.

- Land use has to be held constant. Project sponsors or others may use Summit to test differences in land use, but FTA evaluations are based on consistent land use.

Because the trip tables must be held constant, the final trip table from the trip distribution should reflect the build alternative and any proposed highway improvements that would complement the system (e.g., direct access ramps into stations, etc.). The productions and attractions for zone pairs cannot change, but the mode shares for each interchange can shift. It is beneficial to use the trip table that best reflects the build alternative in the Summit-based analysis.

The in-vehicle time coefficients for all modes must be the same in path building. There cannot be different coefficients for highway modes and rail modes. The weighted time, including wait time, access time, and in-vehicle time, must be the same for all modes. This also includes any transfer penalties. The access-sheds for rail and bus must be the same. Rail cannot have a longer walk access than bus – unless, as stated above, the engineer or planner responsible for the model has data to support it.

The transit coding and networks should be clean. There should be equivalent access points in both the build and baseline alternatives. Summit will quickly identify issues with coding. Therefore, to save cost, debug those issues at the start of the modeling.

Applications in the Current TPB Framework

TPB's current mode choice model (sequential multinomial logit) could be used with Summit, although it could be a challenge to get the results accepted. The current mode choice model does not have different transit modes, and it only develops shares for transit, auto, and high-occupancy vehicles (HOVs). The model does have access-market segmentation, although there is a problem with the different walk-sheds for rail and bus. The sheds would have to be changed to be equal. The mode-specific weighted times for path building would have to be consistent. Bus and rail can not have different weighted times for in-vehicle or out-vehicle components. Currently there is an adjustment to bus in-vehicle time, which is based on a static set of factors that are designed to reflect the effect of congestion. However the relationship of these factors to the highway skims is not clear, and they could be construed as weights for in-vehicle bus time. Also, the current mode choice model would have to produce the output files required for input to Summit. The required file format is outlined in the Summit User's Guide and could be easily added to the mode choice executable.

The current approach for using the MWCOG model for New Starts projects is to apply a post-process mode choice model, as shown in Figure 2-3. The current model applied in this post-process has been used for Summit submissions. This model provides a nested-logit structure for different transit modes, although Summit does not require a nested-logit structure to be executed. Summit can work with a multinomial logit structure. The important element in the post-process mode choice model is the representation of different transit modes in the transit nest and the output of the required input data for Summit.

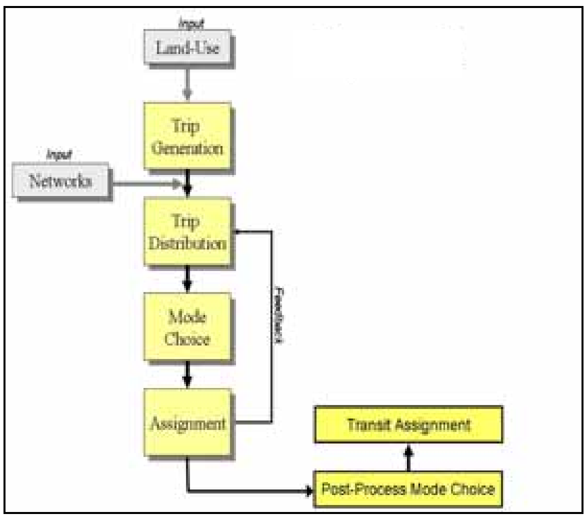

Figure 2-3: TPB Approach for New Starts Forecasts

An issue with the post-process mode choice approach is the lack of any feedback into the model chain. By taking the initial mode choice results and applying another mode choice model after the highway assignment is completed, there is no reflection of mode shifts given the new transit service. There may not be a significant impact on LOV trips, but if the build transit service competes with HOV facilities, then there could be a significant impact on HOV with a shift of HOV trips to transit modes. This mode shift is not addressed with the post-process application of the mode choice model.

A potentially greater issue is the interaction of the transit system with the model chain. By redoing the mode choice as a post-process after the model is completed, problems in the model chain can be overlooked until Summit identifies them. As the mode choice is a post-process, the user trying to fix a problem related to an earlier step in the model chain will not address the real issue. An example of this is when transit times are not given the proper weight for individual interchanges in the trip distribution model. In this case a potential transit trip interchange from a suburb to a downtown CBD would not be identified as a potential transit interchange. This is a trip distribution problem that cannot be addressed by changing mode coefficients and bias constants in the post-process mode choice model in order to achieve reasonable simulated to observed transit boardings.

The easy solution to these issues is to include the post-mode choice model in the model chain, which TPB is currently attempting to accomplish. The post-process mode choice model is currently not developed for the region, but it has been applied for localized areas within the boundaries of the regional model. The post-mode choice model needs to be refined and calibrated for the region before it is applied as part of the regional travel demand model. This can be a very challenging task. Having a post-process mode choice model is essentially rearranging the four-step sequential process, and the benefit of doing this is questionable. It may be a stopgap measure, but the correct action for TPB is to incorporate a better mode choice model in the four-step process. As they review and calibrate the new mode choice model, staff should note the items identified in the "Good Practice" section of this chapter.

Summary

Summit is an evolving tool currently used for all New Starts applications. It has become an important part of the planning process. Summit is useful for highlighting shortcomings in the mode choice model, but the process has resulted in little leeway for using different factors in mode choice models. The only real requirement for the model structure is that the mode choice model be an econometric choice model. Summit will not work on expected default coefficients and unless there is data to support different values, a New Starts applicant must use model coefficients and procedures that are consistent with these requirements. Summit has been applied using traditional four-step models as well as state-of-the-art activity-based models. In San Francisco County Summit was recently applied to the results of an activity based model. The application was viewed as a success.

Task 3 -- Review the Use of Dynamic Traffic Assignment and Traffic Simulation Models in Support of Regional Planning Activities

This page intentionally left blank

TPB staff requested that VHB review the current use of traffic simulation and Dynamic Traffic Assignment (DTA) models among Metropolitan Planning Organizations (MPOs) in the United States. Traffic simulation models can generally be considered to fall into three categories: microscopic, microscopic with DTA capability, and mesoscopic DTA. A mesoscopic traffic model represents traffic as platoons of vehicles traveling through the network. Mesoscopic models are generally associated with DTA, however, there are a number of mesoscopic models without DTA capability that are used for sketch planning purposes where schedule and budget do not allow for detailed microsimulation. By contrast, a microscopic traffic model is the most detailed representation of a vehicle's movement through a network, often on a second-by-second (or smaller) basis. Because of its high level of detail, traffic simulation models are able to produce an animation of individual vehicles moving through the network. This chapter summarizes the two categories of traffic simulation models as well as a number of mesoscopic simulation-based DTA models that have recently been introduced by a number of vendors.

While a number of MPOs have reported the use of a variety of traffic simulation software packages, only a few are currently exploring the use of mesoscopic DTA models, with no MPOs reporting having utilized these tools to date.

Summary of Findings and Recommendations

The use of traffic simulation models and Dynamic Traffic Assignment models are increasing in popularity among MPOs. The results of this research effort reflect this trend, as there are a significant number of traffic simulation models on the market and a growing number of DTA models as well. Table 3-1 summarizes the most popular software and current capabilities; it is important to note that all of the models listed are under constant development which means the current capabilities could and most likely will change in the future.

Table 3-1: Simulation and DTA Model Capabilities

Application							
Traffic Simulation Models	AIMSUN	CORSIM	Cube Dynasim	Paramics	SimTraffic	TransModeler	VISSIM
Pretimed Signals	Y	Y	Y	Y	Y	Y	Y
Actuated Signals	Y	Y	Y	Y	Y	Y	Y
Unsignalized Intersections	Y	Y	Y	Y	Y	Y	Y
Cars, Trucks	Y	Y	Y	Y	Y	Y	Y
Bus Routes, Bus Stops	Y	Y	Y	Y	N	Y	Y
On Street Parking	Y	Y	Y	Y	N	Y	Y
Pedestrians	Y	L	Y	Y	Y	Y	Y
Car Pools	Y	Y	Y	Y	N	Y	Y
Bus and Carpool Lanes	Y	Y	Y	Y	N	Y	Y
Freeway	Y	Y	Y	Y	L	Y	Y
Ramp Metering	Y	Y	Y	Y	N	Y	Y
Roundabouts	Y	L	Y	Y	Y	Y	Y
Right Turn Islands	Y	L	Y	Y	Y	Y	Y
Temporary Events	Y	Y	Y	Y	N	Y	Y
Transit Priority	Y	N	Y	Y	N	Y	Y
Light Rail	Y	N	Y	Y	N	Y	Y
Toll Plazas	Y	L	Y	Y	L	Y	Y
Variable Message Signs	Y	N	Y	Y	N	Y	Y
Dynamic Assignment	Y	N	N	Y	N	Y	Y

DTA Models	Dynamec	Dynasmart-P	Cube Avenue
Queuing Evaluation	Y	Y	Y
HOT Lanes	Y	Y	Y
Variable Message Signs	L	Y	N
Evacuation Plans	Y	Y	Y
Special Event Planning	Y	Y	Y
Ramp Metering	N	Y	Y
Traffic Signals	Y	Y	Y

Y=Yes; L=Limited; N=No

The research illustrates that TPB has a number of traffic simulation and DTA models to select from. There are a number of factors to consider when selecting a software package, including existing staff capability (training implications), potential staff recruitment (the more popular the software, the easier it will be to locate experienced users), potential applications (i.e., does TPB want to evaluate LRT, Intelligent Transportation Systems [ITS], and/or signal timing strategies), and institutional history (i.e., if TPB uses CUBE, one would expect that it would require less effort and training to utilize Dynasim and Avenue than TransModeler or Dynasmart).

Traffic Simulation versus DTA (Mesoscopic)

Traffic simulation models have been used for project planning and traffic operational studies for several decades. Recent advances have added transit priority and a variety of ITS applications to the list of topics that can now be evaluated in detail using these tools.

The challenge with the use of traffic simulation models lies in the amount of labor involved in data collection, network coding, and network calibration, as simulation models require detailed turning movement volume data, traffic signal timing settings, and detailed lane geometries, down to the lengths of turn bays.

While the output of traffic simulation models, particularly ones with three-dimensional (3D) animation and visualization capabilities, are popular among high-level MPO decision-makers and board members, the amount of labor required of technical staff and the associated budget involved makes it difficult to develop these services and maintain them in the long run. These constraints have led to the development of a new breed of models, mesoscopic DTA models that operate at a scale between the typical regional travel demand forecasting models (macroscopic) and traffic simulation models (microscopic).

The mesoscopic DTA models utilize the theory of user equilibrium assignment combined with time dependent origin-destination (OD) matrices to give audiences a richer representation of traffic conditions at the regional and sub-regional level than a regional static planning model while simultaneously reducing the amount of effort required for network coding and calibration when compared to microscopic traffic simulation models. A typical DTA model can utilize the network geometry directly from the regional model, and the user can add additional details such as the number of turn lanes and signal timing data in the areas of greatest interest while using default values for the remaining portions of the network. Since the DTA models are simulation-based, queuing associated with capacity constraints in the network can be readily observed, unlike in static planning models.

Traffic Simulation Models

SimTraffic

SimTraffic is a microscopic traffic simulation model produced by Trafficware. SimTraffic was designed to work with Synchro as its interface, and Synchro serves the dual purpose of performing traffic signal optimization as well, making it perhaps the most widely used simulation package in the U.S. SimTraffic can be used to evaluate arterials, particularly pretimed and actuated traffic signals, unsignalized intersections, pedestrians, roundabouts, and right turn lanes. SimTraffic has the capability of simulating freeways, but the software is limited in that there is no explicit freeway link inherent in the software meaning that the car following and lane changing methodologies that are borrowed from CORSIM are based on the arterial methodologies from the NETSIM component of CORSIM which are not identical to the methodologies used for freeways. This limitation is often overcome to some degree by adjusting the parameters of the Synchro links to reflect speed and merging conditions that would prevail on freeways, but the model often struggles in highly congested conditions partially because of a lack of true freeway car following and lane changing methodologies.

SimTraffic is a relatively easy-to-use tool adding to its popularity among practitioners. The network coding occurs in Synchro and requires less labor than CORSIM and other simulation packages as Synchro nodes are created automatically by overlapping links, eliminating the need to code both links and nodes as in CORSIM. The traffic signal timing coding requires a more experienced traffic engineer as does the calibration process, but Synchro explicitly optimizes both isolated traffic signals as well as signal systems whereas CORSIM requires an experienced traffic engineer who can develop the optimized timing plans manually or through numerous iterations with the model, or by interfacing between CORSIM and SimTraffic. As with the majority of simulation software, Synchro/SimTraffic networks can be coded over aerial photography which allows for quicker base network coding.

Synchro/SimTraffic is also popular among MPOs. It is used by over 20 MPOs including 3 MPOs similar to TPB.

CORSIM

CORSIM is one of the oldest and most popular traffic simulation models in current use. The software was developed by the Federal Highway Administration (FHWA) to support operational analysis and has steadily advanced over the past several decades. Like SimTraffic, CORSIM can model arterials, though signal timing optimization is not a part of the software; signal timing adjustments have to be made manually which adds to the level of effort for arterial analysis when compared to SimTraffic. CORSIM includes car following and lane changing methodologies for both freeway and arterial segments, making it more appropriate for freeway analysis than SimTraffic. CORSIM can be used to evaluate a variety of traffic control strategies, including stop/yield signs, actuated and pretimed traffic signals, and ramp metering. Moreover, CORSIM can simulate bus routes and bus stops including dwell times, on-street parking, pedestrians (using delay factors; not modeled explicitly), carpools, HOV lanes, bus lanes, limited roundabout analysis (all links in a roundabout must be longer than 50 feet which can be difficult to achieve), HOV lane bypass at ramp meters, incidents, weigh stations, toll booths, and airport loading/unloading zones. It should be noted that a number of the applications, such as toll booths and airport loading/unloading require experienced traffic engineers to ensure the model is replicating field conditions in a reasonable manner.

CORSIM historically has been a labor-intensive software package, though much progress has occurred in this regard recently, particularly with the introduction of TRAFED, the graphical editor that allows links and nodes to be coded over network maps or aerial photography. TRAFED also improved on many of the glitches discovered in ITRAF, the previous graphical interface; this dramatically reduces the number of fatal errors in a coded network which require an experienced CORSIM user to debug.

The software requires inputs such as traffic control and signal timing data, roadway and intersection geometries, and demand data; either through explicit turning movement counts, turning movement percentages, or limited capability for OD table input. Depending on the environment simulated, additional inputs may be required including truck percentages, bus routing information, bus dwell times at stops, pedestrian counts, and number of parking maneuvers. Typically, an entry-level engineer or planner can be utilized for basic network coding, such as roadway geometry and intersection control; whereas a number of advanced parameters require a more experienced traffic engineer who has an understanding of traffic signal timing and traffic flow theory. This is particularly important for calibration purposes, which requires the adjustment of factors such as driver aggression and familiarity.

Based on a recent MPO survey, CORSIM is used by over 40 MPOs, including 12 very large MPOs similar to TPB.

Cube Dynasim

Cube Dynasim is a relatively new simulation model developed by Citilabs, the developers of TP+, Cube, and TRANPLAN. Dynasim was developed to provide a relatively seemless transition between Cube regional planning models and detailed traffic simulation. Dynasim can be used to evaluate a variety of scenarios, including arterials, expressways, HOV lanes, toll plazas, ramp metering, taxi stands, transit priority, ITS, evacuation plans, truck terminals and advanced signal systems and technologies.

Cube Dynasim's required inputs include a highway and if applicable, transit network, which can be coded manually using aerial photography or maps, or networks can be extracted directly from Cube or Viper travel demand model networks. Other inputs include signal timing data which can be input manually, or imported from Synchro, demand data in OD format, general traffic control, and bus routing information. The integration with Cube Voyager and TP+ reduces the amount of effort required for coding networks, though a lot of additional coding would still be required including signal timing and detailed intersection geometries that are necessary to calibrate the Dynasim model.

COG/TPB staff has purchased a copy of both Cube Dynasim and Cube Avenue (DTA). Although staff has not had the time to work with the DTA module, staff has done some work simulating traffic around I-395 and South Eads Street in Arlington County using Dynasim. Based on this initial work, staff had the following observations: First, developing a simulation in Dynasim is very labor intensive and time consuming. Although it is true that you can export a sub area network from TP+/Voyager to Dynasim, the exported network then needs a substantial amount of additional coding detail and clean up work. Consequently, it may actually be quicker to simply manually re-code all of the study area links and intersection controls directly in Dynasim, which, in itself is no small task. Second, staff felt that the Dynasim software had a number of bugs in it, which Citilabs worked diligently to correct, but nonetheless, ended up making it difficult to work in a timely manner. Third, acquiring quality traffic counts needed to calibrate the simulation is often difficult or impossible. Fourth, another time consuming feature of developing these simulations is acquiring the AutoCAD files of road geometries from state and local governments. It is important to note, that a number of the limitations pointed out by COG/TPB staff, such as the availability of traffic counts and coding effort would apply to all of the simulation models researched.

Cube Dynasim is currently used by one MPO based on a recent survey; however, this number is expected to grow in the future due to Dynasim's integration within the Cube suite of software which is used for regional planning by a large number of MPOs in the U.S, including TPB. There is currently little experience with the software, so the jury is still out on how well it calibrates, etc.

Traffic Simulation Models with DTA Capability

Paramics

The United Kingdom-based company Quadstone is the manufacturer of Paramics, a suite of microscopic simulation modules in an integrated platform. Paramics is fully scaleable and designed to handle scenarios as wide-ranging as a single intersection to a congested freeway or the modeling of an entire city's traffic system. Paramics can be used to evaluate transit priority, arterials, congested freeways, HOV, ITS, including ramp metering, variable message signs, route control, lane usage, and freeway speed control, parking, incidents, and work zones.

Paramics can translate a variety of common files including Synchro, CORSIM, and Cube/TP+. This helps reduce the labor and learning curve associated with network coding and debugging; however, it is important to note that Paramics uses only OD information to simulate demand. This would potentially add to the calibration effort as typically link speed is one of the primary calibration adjustments. In the case of a Paramics assignment whether static or dynamic, link speed is also a variable in route choice, meaning that link speed adjustments would need to occur in an iterative fashion until the demand pattern matches observed traffic volumes in the field. This could also lead to some instances where the link speed after calibration does not closely match field conditions.

There are currently less than a handful of MPOs using Paramics based on a recent survey. These MPOs are concentrated on the East Coast where Quadstone has its U.S. headquarters.

VISSIM

Over the past decade, VISSIM has become one of the more popular traffic simulation software packages in the U.S., particularly in the context of light rail and bus rapid transit (BRT) evaluation due to its ability to model transit priority. Developed by the German company PTV AG and marketed here through its U.S. subsidiary, PTV America, VISSIM is a microscopic, behavior-based multi-purpose traffic simulation program. VISSIM can be used to evaluate arterials, congested freeways, transit priority, traffic management systems such as alternative route control, traffic flow control, toll roads, access control, HOV and HOT lanes, feasibility analysis of large networks with alternative route choice using dynamic assignment, capacity analysis of toll plazas and border control facilities, traffic calming, parking, parallel vehicle flows (e.g. cars and motorcycles) driving in the same lane as well as overtaking vehicles inside wide lanes, and NEMA and Type 170 signal controller interfaces for real-time evaluation.

VISSIM uses a link-connector system to lay out networks. This allows for greater flexibility with regards to evaluating complicated intersections and roadway/transit networks, but adds significantly to the effort required for network coding. In response to this, VISSIM recently introduced a Synchro interface which allows Synchro files to be translated directly into VISSIM with minimal modifications. VISSIM can also read data in GIS format which can reduce the effort required for network coding; however, since the software gives the user greater control over the network coding and hence, calibration procedures than other software such as Synchro/SimTraffic and CORSIM, VISSIM applications typically require a more experienced traffic engineer to ensure model accuracy and consistency.

Reflecting its increasing popularity in the U.S., VISSIM is used by approximately two dozen MPOs currently, including a dozen similar in size to TPB.

AIMSUN

AIMSUN is a microscopic traffic simulation model that has been compared favorably by practioners to VISSIM in the past. AIMSUN can simulate urban networks, freeways, arterials and any combination thereof. It has been designed and implemented as a tool for traffic analysts to help traffic engineers in the design and assessment of traffic systems. It has proven to be very useful for testing new traffic control systems and management policies including, adaptive traffic control systems such as SCATS, transit priority, Advanced Traffic Management Systems (ATMS) including ramp metering and variable message signs, vehicle guidance systems, and incidents.

AIMSUN also has DTA capabilities. The simulator is able to model the drivers' reasoning for route selection before and during the trip. It includes four different algorithms to model dynamic route choice, a function editor to allow the specification of cost functions, and the option of considering the costs from historical routes and/or considering the driver's memory. A variety of drivers will use different criteria: from always sticking to the same path to changing their path according to advice from a guidance system or traffic conditions.

According to a recent survey, no MPOs reported using AIMSUN.

TransModeler

TransModeler is a traffic simulation model developed by the Boston-area based Caliper Corporation, the manufacturers of TransCAD. TransModeler has the capability to model mixed freeway and arterial networks, HOV lanes, bus lanes, toll facilities, evacuation plans, work zones, traffic signal systems, traffic signal preemption, lane use signs and flexible variable message signs, ramp metering effects on freeway and adjacent urban streets, the impact of real-time traffic information on dynamic driver rerouting, and transit priority.

TransModeler has a unique GIS architecture that integrates traffic simulation models with a GIS that has been extended to store, maintain, and analyze transportation and traffic data. This allows for the storage of information such as traffic counts, lanes, and speeds which becomes a useful database for future studies.

TransModeler is also unique in that it can simulate at the microscopic, mesoscopic, and macroscopic levels, including hybrid simulations in which microsimulation can be intermixed with mesoscopic and macroscopic simulation on any network segments. This allows the network of greatest interest to be simulated at the micro level and others at the mesoscopic and/or macroscopic scale which makes it possible to simulate very large networks with modest computing power.

TransModeler is integrated with TransCAD allowing for integrated travel demand and traffic modeling. Travel demand forecasts can be subjected to more detailed operational analysis with the use of embedded matrix estimation procedures (adjusting the OD table to match existing count data). Conversely, traffic simulation results can be fed back to the travel demand model for improved destination and mode choice.

TransCAD can be used to run a Stochastic User Equilibrium or Dynamic User Equilibrium assignment to generate congested link travel times, flows, and turning movements as input to the TransModeler route choice models.

To help aid the learning curve associated with TransModeler, CORSIM and SimTraffic files can be imported, though it is unclear how much additional effort may be required to further format these files prior to running TransModeler. The required inputs are similar to other traffic simulation software and include detailed lane geometries, traffic control data, demand data including vehicle and truck counts or OD matrices, pedestrian counts, and transit routing information. In this regard, experienced traffic simulation modelers should have a relatively modest learning curve with regards to basic data inputs. The GIS based scenario development implies that the user would also need to develop a basic understanding of GIS, which is unique to TransModeler.

TransModeler is a relatively new entry into the traffic simulation arena. Hence, no MPOs responded as having used the software to date; however, based on conversations with Caliper, several MPOs who use TransCAD for regional travel demand forecasting are exploring the use of TransModeler.

Dynamic Traffic Assignment Models

Dynameq

Dynameq, developed by the Canadian company INRO (the developers of EMME/3), is an equilibrium DTA model for use on large congested networks. Dynameq enables planners to evaluate congested network scenarios with dynamic equilibrium benchmarks, a time varying version of the same well-understood equilibrium assignments used in static analysis for years. Dynameq's equilibrium traffic assignment results represent user optimal network conditions that are immediately useful as an upper-bound on network performance.

Traffic phenomena that trigger congestion are modeled explicitly, including signals, conflicting movements at intersections, lane permissions for turning movements and vehicle classes, and weaving. Each vehicle travels along a particular lane, performs lane changes where appropriate, and crosses signalized and unsignalized intersections. Large networks tend to be more data-intensive. Dynameq is designed with a minimal set of meaningful model parameters to get the model up and running as quickly as possible. The user can focus data collection and network coding effort to the parts of the network that need it most, and use link and intersection default settings, for less critical parts of the network. One can use constant demand extracted from static planning models, or separate the demand matrix into time slices.

The user can draw insight from simulation results using a variety of analysis tools, and communicate results to decision-makers. Decision makers can see the big picture with animated network-scale results to identify congestion patterns and assess the extent of congestion with animated plots of lane-by-lane queues.

The current maximum network size consists of 10,000 links, 5000 intersections, and 1000 transportation analysis zones. Dynameq is used to evaluate lane closures, infrastructure expansion at the sub-regional level, Managed Lanes, HOT Lanes, pre-timed signal control, and incidents.

As Dynameq is manufactured by INRO, the developers of EMME/3, EMME/3 users would require minimal training to use Dynameq. Dynameq, like other DTA models, does not require the level of detailed inputs that traffic simulation models require, which also reduces the amount of labor involved in network coding. Dynameq has not developed an interface as of yet to read in Synchro, CORSIM, and/or Cube/TP+ networks which would make it more cumbersome for experienced users of those software packages to implement Dynameq.

Dynameq is currently not used by any of the U.S. MPOs that responded to a recent survey. This is consistent with the limited use of EMME/3 by those MPOs.

Dynasmart-P

Dynasmart-P uses mesoscopic simulation combined with DTA to model the evolution of traffic flows in a traffic network, which result from the travel decisions of individuals. The model is also capable of representing travel decisions of travelers seeking to fulfill a chain of activities at different locations in a network over a given planning horizon.

Dynasmart-P was developed by the University of Maryland, College Park, in concert with FHWA to address the growing need to evaluate Intelligent Transportation Systems (ITS) in the regional planning context. Dynasmart can provide dynamic traffic assignment methods for traditional planning analyses, assess the impacts of ITS technologies, such as dynamic message signs, ramp meters, and in-vehicle guidance systems; assess the impacts of different traffic operations and control strategies, evaluate regional work zone management, evaluate incident management and special event management strategies, and evaluate congestion-pricing schemes.

However, it is important to note that Dynasmart-P cannot model detailed traffic maneuvers, such as car-following, lane-changing, and weaving operations which would still require microscopic analysis. In addition, there is currently limited transit and inter-modal modeling capabilities, though future versions will be better able to perform this type of modeling.

While Dynasmart-P is a new software package with limited applications to date, it has been used in the following efforts:

- Develop traffic management strategies for major highway reconstruction projects in Zwolle, a city in the Netherlands.

- Evaluate downtown El Paso, TX traffic and environmental impacts of one and two-way traffic flow reconfigurations. This project used a combination of Dynasmart-P and CORSIM.

- Undertake a pilot study to apply DTA as a part of the regional four-step modeling process in the El Paso, TX region.

Dynasmart can translate networks from Cube/TP+, CORSIM, and most GIS formats; this allows large networks to be readily converted into Dynasmart format. Dynasmart can either use default traffic control settings or actual signal timing data which allows for greater flexibility in the amount of labor required as the user can focus their coding efforts on the areas of greatest interest while using default values for the rest. However, it is important to note that using the default traffic control can lead to similar issues as what occurred with TRANSIMS where it was discovered that traffic control and signal timing had a much greater impact on route choice than the researchers expected.

Another area of potential concern with regards to Dynasmart is the lack of transparency in the OD estimation process. To date, all of the efforts involving Dynasmart have required that the OD estimation occur at the University of Maryland. For Dynasmart to become more of a mainstream software package, the OD estimation procedures utilized by the University of Maryland will need to be shared with other users, or other matrix estimation software such as those developed by Citilabs and others will need to be tested on Dynasmart networks.

There are currently a handful of MPOs testing Dynasmart, primarily in the areas of evacuation planning and regional ITS planning.

Cube Avenue

Cube Avenue is a mesoscopic model developed by Citilabs, the developers of Cube, TP+, and Voyager. By explicitly modeling time, Cube Avenue can be used for studies comparing policies for alleviating peak period congestion, such as variably priced toll lanes, as well as evaluating the effectiveness of emergency evacuation plans. Cube Avenue can also be used to quantify impacts of upstream traffic congestion, measure queuing at intersections and merge points in a network, isolate secondary impacts from one intersection to another, ITS strategies such as HOT lanes and ramp metering, emergency evacuation plans and strategies, special event planning, and traffic control, including traffic signals, roundabouts, and stop-controlled intersections.

Cube Avenue works with conventional Cube/TP+ job scripts and networks which minimizes the learning curve for current users of Cube. The networks and associated OD tables can be extracted directly from regional networks, further reducing the amount of labor associated with network coding. Furthermore, as this is a mesoscopic model, the level of detail associated with network representation is less than that of traffic simulation models.

Cube Avenue was recently released by Citilabs. Hence, no MPOs responded as having used this software in a recent survey, though it is likely that a significant number of MPOs will be evaluating and/or utilizing this software in the future based on the large number of MPOs in the U.S. who use the Cube suite of software for regional travel forecasting.

TRANSIMS

TRANSIMS is an agent-based simulation system capable of simulating the second-by-second movements of every person and every vehicle through the transportation network of a large metropolitan area. It was developed by the Los Alamos National Laboratory.

It consists of mutually supporting simulations, models and databases. By employing advanced computational and analytical techniques, it creates an integrated environment for regional transportation system analysis.

TRANSIMS is designed to give transportation planners more accurate, complete information on traffic impacts, energy consumption, traffic congestion, land use planning, traffic safety, intelligent vehicle efficiencies, and emergency evacuation.

TRANSIMS has the capability to analyze traffic over the entire transportation network of a metropolitan area, including local streets and highway ramps, compute precise speed and acceleration information for every single vehicle at any second of the day, and provide second-by-second information allowing for a much more precise analysis of time-of-day effects.

While TRANSIMS is a very powerful modeling platform in theory, when applied to the Portland region it was discovered that the model requires detailed roadway geometry, signal timing, and phasing data to accurately model route choice; acquiring this data for an entire metropolitan area the size of the TPB region would be a labor intensive effort with significant costs associated with it. Moreover, modeling a region the size of Metropolitan Washington to this level of detail would require significant computer processing capabilities and even with this, it would likely take days for the model to run, making it impractical for most TPB applications.

Conclusions

The simulation and/or DTA software selected and implemented by TPB should reflect existing staff and consultant capabilities as well as provide new and/or better solutions to the most pressing modeling questions that TPB faces. For example, peak spreading and managed lanes are two areas where the existing regional model has limitations, and DTA would theoretically provide better answers because of its ability to explicitly model time as well as capture impacts of traffic control and queuing, which all relate to peak spreading. Likewise, Express Toll Lanes (ETL) require the ability to model congestion over time, which would require a DTA model to do this type of analysis at the regional level or a traffic simulation model with DTA capability to evaluate these types of strategies at the corridor level. In addition to these topics, traffic simulation models could be used to evaluate corridors as a part of the federally-mandated congestion management system (CMS) program.

COG/TPB has traditionally used macroscopic traffic assignment methods, such as static user equilibrium traffic assignment, to carry out its regional transportation planning activities. As this chapter points out, the used of traffic microsimulation models require a lot of data and coding effort, which would preclude them as a practical tool for regional planning purposes, unless specific corridors are being evaluated in detail. The more appropriate tool for regional planning purposes, particularly in the context of HOT/Managed Lanes analysis and regional ITS planning would be a mesoscopic DTA model which would explicitly model peak period demand over time and illustrate the queuing affects associated with HOT lanes and other roadway elements. The mesoscopic DTA models do not require the level of detail that microsimulation models require; which reduces the coding effort and allows for a more seamless transition between the regional model networks and the mesoscopic DTA model networks. Given that COG/TPB staff uses Cube for regional planning efforts, Cube Avenue would be the most practical mesoscopic DTA model to use for HOT/Managed lanes analysis. Dynasmart has limited technical support which makes it very difficult for staff to learn and implement the software effectively, and Dynameq has limitations with regards to network size, making it difficult if not impossible to utilize in large metropolitan planning regions such as Washington, D.C.

Moving forward, TPB may want to pursue a pilot study where DTA is used as the fourth step of the modeling process and conduct a screenline and corridor level validation to determine if the DTA assignment is indeed an improvement over the existing static equilibrium process. If the results are positive based on this test study, then it is recommended that TPB test DTA in the context of ETL and HOT lanes as well as peak spreading.

Appendix 3-1 -- Detailed software methodologies

CORSIM Methodologies

The freeway component (FRESIM) in CORSIM uses the Pitts car-following model which is based on the distance headway between vehicles. The objective function of the model is dependent on the lead vehicle length, driver sensitivity of the following vehicle, speed of the following vehicle at time *t*, speed of the lead vehicle at time *t*, and a calibration constant defined by the user. The calibration constant affects vehicle acceleration and deceleration rates which in turn affects the headways that can be maintained between vehicles.

The arterial component (NETSIM) of CORSIM uses car-following logic where the independent or lead vehicle attempts to maintain free-flow speed and the follower avoids collisions with the leader.

The lane changing logic is broken out into three categories:

- Mandatory, which is based on acceptable risks for the driver making the lane change. This logic is used for lane drops, merging, exits, and lane blockages.

- Discretionary, which is based on driver behavior (aggressive vs. non-aggressive).

- Anticipatory, which is applied before on-ramps to allow vehicles to enter the freeway.

Lane changes in general are dependent on gap acceptance models where an acceptable minimum gap between vehicles in the target lane is required to accommodate a lane changer.

Similarly, the gap acceptance models used in NETSIM to simulate stop/yield conditions and permissive turns are based on the time to travel from conflicting point to the opposite stop line (or stop bar). This model is also a function of driver composition and is one of the calibration components in CORSIM, i.e. aggressive driver populations will accept smaller gaps than non-aggressive populations.

SimTraffic Methodologies

In general, SimTraffic uses the same driver and vehicle characteristics as the NETSIM component of CORSIM. With regards to car-following, SimTraffic uses a formula that has vehicles track leaders at a fixed headway. The headway is dependent on speed, driver type, and link characteristics. The acceleration rates used in SimTraffic are identical to NETSIM and the deceleration rates used in SimTraffic are very close to those used in NETSIM.

SimTraffic also has similar lane change logic to NETSIM. The vehicles will complete a lane change when the next lane is clear. To be clear, both the changing vehicle and the vehicle behind must not obtain a deceleration rate above the threshold using the car-following formulas. In SimTraffic a vehicle can be stopped in the middle of a lane change and block 2 lanes. SimTraffic's lane changes tend to be more disruptive than NETSIM because the vehicles require a forward movement to complete the lane change where in NETSIM they do not.

In SimTraffic, gap acceptance is based on the type of turn made and the length of the turning path. The gap times in SimTraffic are more consistent with the amount of time required to complete the turn and are towards the high end of the gap times in NETSIM. In general, SimTraffic will accept fewer gaps but give safer operation.

Cube Dynasim Methodologies

Dynasim uses methodologies developed by Kazi Ahmed at the Massachusetts Institute of Technology (MIT). The car-following model is based on two regimes, the free-flow regime where a driver is assumed to try to attain his/her speed and the car-following regime where the driver is assumed to follow his/her leader. A probabilistic model that is based on a time headway threshold is used to determine the regime the driver belongs to. Heterogeneity across drivers is captured through the headway threshold and reaction time distributions. The parameters of the car-following and free-flow acceleration models along with the headway threshold and reaction time distributions are jointly estimated using the maximum likelihood estimation method.

The lane changing decision process is modeled as a sequence of three steps: decision to consider a lane change, choice of a target lane, and gap acceptance. Since acceptable gaps are hard to find in heavily congested traffic, a forced merging model that captures forced lane changing behavior and courtesy yielding is developed. A discrete choice model framework is used to model the impact of the surrounding traffic environment and lane configuration on drivers' lane changing decision process.

Paramics Methodologies

The lane changing methodology used in Paramics is prioritized into two levels, urgent and non-urgent. Within Paramics, a driver will attempt to execute a lane change maneuver as a response to either a single urgent stimulus or a series of five contiguous and consistent non-urgent stimuli produced by unsuitable transient conditions.

An urgent stimulus is generated if a driver finds itself outside its target range of lanes. Near a hazard, the target range is controlled by the number of lanes available on the exit link appropriate to the driver's choice of route. At all times, the target range is adjusted subject to the behavior patterns associated with the driver and vehicle type; a higher level of aggression causes a driver to move to the outer (higher speed) lanes, a higher level of awareness causes a vehicle to adopt the target lane for an impending turn sooner.

An urgent stimulus is also generated if the vehicle caught in a stationary line of traffic (as a result of an incident for example).

A non-urgent stimulus can be generated for a number of conditions, which are themselves prioritized as follows:

- Move in or out because of constraints imposed by a fixed physical object such as a ramp joining, or a climbing lane.

- Move in or out as suggested by free-flow lane-changing model. This can be defined by the user, or the standard free-flow model can be used.

- Move in or out on an urban road in such a way as to spread the total demand over the available road space. In the absence of other stimuli, this prevents false congestion from building up.

Note that these conditions describe what is necessary for a Drive Vehicle Unit (DVU) to receive a stimulus to attempt to change lanes in either direction. The actual lane-changing maneuver will not occur unless a suitable gap exists. The gap acceptance function can be defined by the user, or the default settings can be used.

With regards to car-following, each DVU in the simulation has a target headway. The mean value for target headway is one second by default; however this can be adjusted by the user to match field conditions as necessary. The target headway for each DVU varies around the mean target headway parameter, depending upon the value of certain parameters assigned to the DVU.

In terms of driver behavior, a high aggression value will cause a DVU to accept a smaller headway. Similarly, a high awareness value will affect the use of a longer headway when approaching a lane drop in order to allow DVUs in other lanes to merge more easily.

If not constrained by an approaching junction, a DVU will vary its speed in order to attain its target headway with the DVU in front.

The reaction time of the driver is simulated by basing the calculation of the necessary acceleration/deceleration on the speed at which the DVU in front was traveling at some point in the past.

A default mean reaction time of one second is used, and this is modeled by giving each DVU a memory, so that it carries out with it not only its current speed and position, but a record of its speed and position for a specified number of timesteps in the past. This is referred to as “speed memory” within Paramics. Reducing the driver reaction time is an important factor when considering the throughput of vehicles along a link.

A DVU changes its speed according to its perception of the speed of the DVU in front. These changes are normally smooth, following linear functions, but may be abrupt following the detection of one of two binary signals. These signals are visible brake lights and perceptible acceleration of the DVU immediately ahead. There are therefore three modes of following within the Paramics model, referred to as braking, cruising, and acceleration modes.

For all modes of following, the concept of target point is used. This point is based on a position at an initial distance behind the leading DVU; the target point is then adjusted to improve the car-following behavior.

In addition to the use of an adjusted target point, a bunching acceleration is also used to bring DVUs together rapidly.

In cruising mode, there are five discrete areas, A, B, C, D and E in the headway/velocity-difference phase space. Each of these regions has a separate expression for acceleration. Of these five, three correspond to conditions where the DVU ahead is cruising:

- In Region A, the following DVU has overshot the target point (the headway is less than the target value) and an attempt is made to achieve the target speed as quickly as possible, i.e. as fast as the physical constraints of the DVU allow.

- In Region B, the leading DVU is pulling away from the following DVU.

- In Region C, the DVUs are at a constant separation or coming together.

When the DVU ahead is perceived to be braking (its deceleration is greater than a certain threshold), its perceived speed is decreased by an amount dependent on its maximum deceleration rate. This action models a driver's expectation that if the DVU ahead is braking, its speed in the next time step will be considerably less than at the current time step. The method of application of speed difference and current separation to acceleration ensures that a DVU will over-compensate if the DVU ahead is braking, and that this over-compensation will increase as the distance between the DVUs decreases. This application combined with the time-lag introduced by modeling reaction time results in the shock-wave characteristics as seen typically in highway traffic flow.

However, because the speed of the DVU ahead is predicted, and may have a resultant value of zero, a threshold is used to test whether the following DVU is close enough to be in danger of collision. If not, the acceleration is set to a positive value.

If the DVU ahead is perceived to be accelerating at a high rate, and is more than the following DVUs safe stopping distance away, acceleration is set to the maximum value.

DTA Methodology

The driving force of the Paramics simulation model is an OD matrix applied to a zone map combined with a time-varying profile. This means that the demand on the network between each OD pair can vary in time and can also vary relative to other OD pairs. This leads to a congestion pattern that is also time-variant. To model the route choice decisions that drivers would make based on their knowledge of a time-varying congestion pattern, the user can enable cost table recalculation on a regular basis, perhaps every five minutes of simulation time. The cost recalculation option, when selected, uses mean simulated travel times for links, rather than estimated free-flow travel times. This revised travel time cost can then be injected back into the weighted and factored link cost calculation used previously to create a new routing tree.

Only the route tree for familiar drivers is recalculated at each stage: unfamiliar drivers will still follow the sign-posted routes on the links marked as being major. The ratio of unfamiliar to familiar drivers will determine the damping factor in the feedback control loop: a higher ratio will result in a reduced likelihood of instability.

The justification behind this method is that familiar drivers will have developed experience over time of the true costs of each of the possible routes in the network, and cost feedback and dynamic route recalculation aims to model this phenomenon. It is possible within Paramics to run the model with cost feedback enabled for a period of time, and then save the link costs to file. These link costs can be used as background, or base costs that can be loaded into subsequent runs of the simulation. However, it should be pointed out that cost feedback within Paramics leads to equilibrium only if the time profile of the demand applied to the network is completely flat. For a realistic simulation, it is almost always necessary to model the peaks and troughs of demand, and unless these variations happen at exactly the same time for every OD pair, there will never be a state of equilibrium within the network.

VISSIM Methodologies

VISSIM uses a psycho-physical car-following model for longitudinal vehicle movement and a rule-based algorithm for lateral movements.

The basic idea is the assumption that a driver can be in one of four driving modes:

- Free Driving: No influence of preceding vehicles observable. In this mode, the driver seeks to reach and maintain a certain speed, his individually desired speed. In reality, the speed in free driving cannot be kept constant, but oscillates around the desired speed due to imperfect throttle control.

- Approaching: The process of adapting the driver's own speed to the lower speed of a preceding vehicle; while approaching, a driver applies a deceleration so that the speed difference of the two vehicles is zero in the moment he reaches his desired safety distance.

- Following: The driver follows the preceding car without any conscious acceleration or deceleration. He keeps the safety distance more or less constant, but again due to imperfect throttle control and imperfect estimation the speed difference oscillates around zero.

- Braking: The application of medium to high deceleration rates if the distance falls below the desired safety distance; this can happen if the preceding car changes speed abruptly, or if a third car changes lanes in front of the observed driver.

For each driving mode, the acceleration is described as a result of speed difference, distance, and the individual characteristics of driver and vehicle. The driver switches from one mode to another as soon as he reaches a certain point that can be expressed as a combination of speed difference and distance. For example, a small speed difference can only be realized in small distances, whereas large speed differences force approaching drivers to react much earlier. The ability to perceive speed differences and to estimate distances vary among the driver population, as well as the desired speeds and safety distances. Because of the combination of psychological aspects and physiological restrictions of the driver's perception, the model is called a psycho-physical car-following model.

There are basically two kinds of lane changes in VISSIM, a necessary lane change and a free lane change. In case of a necessary lane change, the driving behavior parameters contain the maximum acceptable deceleration for the vehicle and the trailing vehicle on the new lane, depending on the distance to the emergency stop position of the next connector of the route.

In case of a free lane change, VISSIM checks for the desired safety distance of the trailing vehicle on the new lane. This safety distance depends on its speed and the speed of the vehicle that wants to change to that lane.

In both cases, when a driver tries to change lanes, the first step is to find a suitable gap (time headway) in the destination flow.

DTA Methodology

The DTA procedure in VISSIM is based on the idea of iterated simulation. That means a modeled network is simulated not only once, but repetitively and the drivers choose their routes through the network based on the travel cost they have experienced during the preceding simulations. To model the "learning process", several tasks have to be addressed:

<u>Routes from origins to destinations must be found.</u> VISSIM assumes that not everybody uses the best route but that less attractive routes are used as well, although by a minor portion of the drivers. That means not only the best routes must be known for each OD

pair, but a set of routes must be known for each OD pair. Ideally, one would have the set of the *k* best routes but there are no efficient methods to compute this set of routes directly-at least not in a way that makes sense for traffic assignment. The solution adopted in VISSIM is to compute the best paths in each repetition of the simulation and thus find more than one route because traffic conditions change during the iteration. During the iterated simulations, VISSIM builds a growing archive of routes from which the drivers choose.

<u>The routes must have some kind of assessment on which the drivers base their choice.</u> In VISSIM for all routes the generalized costs are computed, i.e. a combination of distance, travel time and "other" costs (e.g. tolls). Distance and costs are defined directly in the network model, but travel time is a result of the simulation. Therefore VISSIM measures travel times on all edges in the network during one simulation so that the route choice decision model in the next simulation can use these values.

The choice on one route out of a set of possible routes is a special case of the more general problem of discrete choice modeling. Given a set of routes and their generalized costs, the percentage of the drivers that choose each route is computed. VISIM uses the logit formulation for this model.

The iteration of the simulation runs is continued until a stable situation is reached. Stable here means that the volumes and travel times on the edges of the network do not change significantly from one iteration to the next. A convergence criteria, either default or user defined determines what a "significant" change is between iterations, similar to what is utilized in static equilibrium assignment in regional models.

AIMSUN Methodologies

The AIMSUN car-following model is based on the P. G. A. Gipps model (Gipps, 1981) which developed as an empirical model consisting of two components, acceleration and deceleration, defined as functions of variables that can be measured. The first represents the intention of a vehicle to achieve a certain desired speed, while the second reproduces the limitations imposed by the preceding vehicle when trying to drive the desired speed.

The AIMSUN car-following model evolved from the Gipps model by making the desired speed a local parameter where the desired speed of vehicle *n* is for the current section of the roadway. Additionally, AIMSUN considers the influence of adjacent lanes so that speeds on adjacent lanes are within reasonable ranges.

The influence of the section grade in the vehicle movement is modeled by means of an increase or reduction of the acceleration and braking capability.

Lane change is modeled as a decision process analyzing the necessity of the lane change, the desirability of the lane change, and the feasibility conditions for the lane change that are also local, depending on the location of the vehicle on the road network.

In order to achieve a more accurate representation of the driver's behavior in the lane changing decision process, three different zones inside a section are considered, each one corresponding to a different lane changing motivation:

Zone 1 is the farthest from the next turning point. The lane changing decisions are governed by the traffic conditions of the lanes involved; the feasibility of the next desired turning movement is not yet taken into account. To measure the improvement that the driver will get on changing lanes several parameters are considered: the desired speed of the driver, speed and distance of the current preceding vehicle, and speed and distance of the future preceding vehicle.

Zone 2 is the intermediate zone. Mainly it is the desired turning lane that affects the lane changing decision. Vehicles who are not driving on a valid lane (i.e. a lane where the desired turning movement can be done) tend to get closer to the correct side of the road where the turn is allowed. In this zone vehicles look for a gap and may try to accept it without affecting the behavior of vehicles in the adjacent lanes.

Zone 3 is the nearest to the next turning point. Vehicles are forced to reach their desired turning lanes, reducing the speed if necessary and even coming to a complete stop in order to make the lane change possible. Also, vehicles in the adjacent lane can modify their behavior in order to allow a gap big enough for the lane-changing vehicle.

Lane changing zones are defined by two parameters: distance to Zone 1 and distance to Zone 2. These parameters are defined in time (seconds) and they are converted into distance whenever it is required for each vehicle at each section using the vehicle desired speed at a section. This means that these distances are then local parameters; their value depending on the current traffic conditions on the section.

The gap-acceptance model used to model give way behavior determines whether a lower priority vehicle approaching a junction can or cannot cross depending on the circumstances of higher priority vehicles (position and speed). This model takes into account the distance of vehicles from the hypothetical collision point, their speeds and their acceleration rates. It then determines the time needed by the vehicles to clear the junction and produces a decision to cross or not which is also a function of the level of risk for each driver. Several vehicle parameters may influence the behavior of the gap-acceptance model, acceleration rate, desired speed, speed acceptance, and maximum give-way time.

DTA Methodology

AIMSUN also has DTA capabilities: both en-route and user equilibrium. The user equilibrium is the same concept as in static planning applications, whereas the en-route assignment uses a combination of link costs, historical paths, and a logit model that assigns a probability to each alternative route between each OD pair depending on the difference of the perceived utilities which are a function of both the link costs and historical path selection.

TransModeler Methodologies

The car-following model in TransModeler is quite complex and dependent on the acceleration rate of the subject vehicle, speed of the subject vehicle, speed of leading vehicle, distance between the subject and leading vehicles, model parameters, and vehicle-specific error term for the car-following regime. Like the other simulation models, headway is an important calibration variable in TransModeler. In TransModeler, the headway is used to determine the boundary between the car-following regime, the emergency regime where the vehicle will apply an appropriate deceleration rate to avoid collision, and the free-flow regime, where the subject's speed is not constrained, or in any way influenced, by the speed or relative position of the vehicle in front.

TransModeler models lane changing behavior in three steps: selection of eligible lanes, lane changing decision-making process, and target lane selection. These steps determine the feasibility, desirability, and safety of a lane change. The selection of eligible lanes will result in a feasible or rational choice set of alternative lanes including as many as three choices: the current lane, and the lanes on the right and left, if they exist. A lane may be excluded from the choice set if the lane use rules in that lane are not compatible with the vehicle's type or if the lane properties restrict lane changes in that direction.

If there is more than one alternative, the selection of the target lane depends on the lane changing regime. The three lane changing regimes are Discretionary Lane Change (DLC), Mandatory Lane Change (MLC), and Forced Lane Change (FLC).

All lane changes are classified as either mandatory or discretionary and a different model and set of parameters is associated with each. Mandatory lane changes are those that are required, for example, to reach an exit ramp or to enter a left turn lane to remain on one's path. A vehicle might make a mandatory lane change to move around an incident or comply with a lane use message. A discretionary lane change is one made in order to achieve a perceived improvement in driving conditions, such as a gain in speed.

A forced lane change is a special case of a mandatory lane change where either an extended period of time has passed where an acceptable gap has not been found or the location before which a lane change must be executed is very near, or both.

Once the MLC, DLC, or FLC model has been applied and both the lane change and the target lane have been decided, the gap acceptance model is applied each time step in the model until an acceptable gap is found and the lane change is completed.

With regards to gap acceptance, when crossing an opposing or conflicting stream, for example making a permitted left turn, vehicles compare their anticipated time to pass through the conflict point with their perception of the time it will take the conflicting vehicles to arrive at that point. If the difference between these times is greater than a minimum acceptable crossing headway, the vehicle will proceed into the intersection. Minimum crossing headway thresholds are likely to be different than those for merging. In the model parameters, the minimum acceptable headways for crossing and merging are defined by a distribution, with headway thresholds varying by segment of the driving population.

DTA Methodology

TransModeler also has DTA capabilities using OD tables from the regional model, either Cube or TransCAD, or other sources. Travel times by time period and network segment can be input from external data or developed by running traffic assignments and traffic simulations. Vehicle paths can also be input from external files including those generated by TransCAD and/or created or edited by analysts. When unexpected delays occur due to incidents, etc. some drivers will change their routes during their trip, which reflects an en-route DTA.

Dynameq DTA Methodology

Dynameq uses a dynamic user equilibrium assignment algorithm, where the equilibrium conditions vary over time based on the temporal profile of demand and congestion in the network. The equilibrium approach to DTA is to allocate vehicles over the best paths on the network for each OD pair so that vehicles leaving the origin at roughly the same time have approximately the same travel times. Dynameq accomplishes this with an iterative method, where each iteration consists of one execution of a traffic simulation and one execution of a path-choice model. The traffic simulator receives time-dependent flow rates from the path-choice model, and simulates the resulting traffic patterns on the network. The simulator then provides time-dependent travel time information back to the path-choice model, which consequently modifies the path choices for the next iteration. The process continues cyclically until converging to an equilibrium, as defined within a threshold defined by the user.

Dynasmart-P DTA Methodology

There currently is limited published information on the methodologies used in Dynasmart. The limited data on the software revealed that Dynasmart uses a dynamic equilibrium assignment, either user optimal or system optimal. Dynasmart also utilizes the First-In, First-Out (FIFO) constraint that other DTA programs utilize as well.

Cube Avenue DTA Methodology

Cube Avenue uses mesoscopic techniques including FIFO constraints where each downstream link maintains a FIFO queue of packets that want to enter but are blocked. Whenever an event on an upstream link says that a packet should move to the next link, the downstream link is queried to determine if it can accept any packets, if not, the packet is removed from the event queue and put at the back of the downstream links blocked queue. Whenever, a packet successfully moves out of a downstream link, the link checks whether it can accept the front packet(s) from the blocked queue. This allows for the representation of queuing in the network. The user can specify whether to examine traffic as individual vehicles or as platoons of multiple vehicles. The user can also specify time increments in terms of minutes or hours and intersection characteristics.

Using these inputs, Cube Avenue computes the lowest-cost path for each vehicle unit, based on its departure time, and computes interactions among vehicle units as they travel through the network. Cube Avenue estimates travel speeds based on vehicle density on road segments during each time increment.

As Cube Avenue is a part of the Citilabs suite, most urban areas utilizing Cube can use the regional transportation model to implement Cube Avenue. The inputs include the roadway network in Cube Voyager format, peak period trip tables, vehicle storage area (generally specified as [distance*lanes]/[average vehicle length]), roadway distance, capacity, and lanes, and traffic signal locations and characteristics.

Cube Avenue uses dynamic equilibrium assignment and loads and tracks the movement of vehicle packets throughout the highway network. The packets can be any size, from individual vehicles up to platoons of 20 or more vehicles.

The outputs can be specified for the time period specified by the user. The outputs include:

- Total traffic volume on a road link
- Total traffic in queue
- Link operating speed and travel time
- Link occupancy/utilization
- Intersection LOS and operating conditions

References

U.S. Department of Transportation/FHWA CORSIM Theoretical Underpinnings presentation, 2007

SimTraffic Help Menu, 2005

Kazi Iftekhar Ahmed, February 1999. Modeling Drivers' Acceleration and Lane Changing Behavior.
http://web.mit.edu/its/papers/DRIVIN.PDF

Email from Wade White, Citilabs, November 2007

Quadstone Paramics V5.2 Modeller Reference Manual Quadstone Limited, 2006

VISSIM 4.20 User Manual, 2005

J. Barcelo, J. Casas, D. Garcia, and J. Perarnau. Methodological Notes on Combining Macro, Meso, and Micro Models for Transportation Analysis, 2005
http://www.aimsun.com/papers/Sedona-Modeling_Workshop_2005.pdf

TransModeler Traffic Simulation Software User's Guide, June 2006

Michael Mahut, Michael Florian, Daniel Florian, Shane Velan, Nicolas Tremblay. Equilibrium Dynamic Traffic Assignment for Large, Congested Networks, 2005.
http://www.inro.ca/en/products/dynameq/whitepaper_access.php?id=1389&acc=1FhwExVU66

http://www.fhwa.dot.gov/crt/lifecycle/dynasmart.cfm, 2007

Task 4 -- Research the State of the Art in Equilibrium Assignment

This page intentionally left blank

TPB Staff asked VHB to research the State of the Art in equilibrium traffic assignment as a follow-up to the FY 2006 research on the MPO State of the Practice on traffic assignment as a whole. The FY 2006 research found that while a majority of MPOs nationally use equilibrium assignment, there are several outstanding problems with the equilibrium method, including a failure to reach closure within a reasonable number of iterations and instability in the assigned results. These problems with the widely-used Frank-Wolfe (F-W) equilibrium algorithm have been documented by other researchers, and further research has yielded some new algorithms that potentially overcome the issues with the F-W method. These new algorithms are just now starting to appear in major travel demand forecasting software packages and being applied by MPOs and other agencies. This chapter summarizes our findings on the status of the State of the Art in equilibrium assignment and provides guidance for TPB going forward with model development on the specific issue of traffic assignment.

Background: Overview and Current TPB Traffic Assignment Methods

The final element of a traditional four-step travel demand forecasting model is traffic assignment. This step allocates travel demand (vehicle-trips, developed in the previous three steps of the model chain) to a transportation (usually highway and transit) network between origin-destination (OD) pairs according to a specified method. The qualities of a good assignment method include reasonable accuracy, fast and precise convergence, short computing time, and stable results. Among MPOs, the most-widely utilized assignment method is equilibrium assignment, which simultaneously solves for link flow and cost.[2] Equilibrium assignment is predicated on two underlying assumptions: 1) travelers have perfect information on conditions on all possible routes, and 2) travelers always make a rational route choice to minimize their travel time / cost. When the network reaches equilibrium, all trips are assigned to those paths with the minimum impedance (e.g., travel time or travel cost) between each OD pair, and no traveler can improve his or her travel time by switching to an alternate path. Like most MPO models, the current TPB production travel demand model utilizes equilibrium assignment, as will the next production model. Specifically, the TPB model uses the Frank-Wolfe (F-W) algorithm for equilibrium assignment.

Link-based Frank-Wolfe Algorithm

The F-W algorithm, also known as the convex-combination algorithm, is a classic algorithm in operations research and the most widely-applied equilibrium assignment algorithm in travel demand forecasting.[3] The F-W method views the traffic assignment problem as a minimization problem using linear programming. At each step the objective function is linearized and a solution is calculated to reduce the objective. In general, the F-W algorithm performs well during the first several iterations, but it slows down significantly when close to the minimum point (that is, approaching equilibrium) and never reaches its objective function's minimum.[4] Therefore, the algorithm may be best

[2] See MWCOG (2006) and Spielberg and Shapiro (2006).
[3] See Boyce, *et al* (2004) and Slavin, *et al* (2006).
[4] See Dial (2006).

used to find an approximate solution rather than a true equilibrium.[5] Besides the commonly used performance measures – gap, relative gap and average excess cost, the stabilization of link flows from iteration to iteration gives the forecaster some assurance that an adequate approximation has been achieved.

In recent years the F-W method has been widely used for determining the equilibrium flows in transportation networks. Theoretically, true user equilibrium can only be achieved in an artificially small or virtually uncongested network; for a highly congested transportation network, equilibrium can only be closely estimated. Most travel demand forecasting software packages use the F-W method. Compared with other equilibrium assignment methods, it is easy for software developers to code and requires the least computer memory since at each iteration it deals with only a single path between each origin-destination pair.[6]

Several issues with the F-W algorithm have been reported in previous research, such as slow convergence, long computational time, and unstable assignments, in which a relatively small change to the travel network or other conditions produces unexplainable results across the whole network.[7] According to Wolfe, the unsatisfactory performance occurs because the search direction tends to become orthogonal to the steepest descent direction as the optimum solution is approached.[8] In addition, the F-W algorithm has no mechanism to avoid the introduction of cyclic flows. A cyclic network normally contains a cycle, a path from a node to itself, which may be one reason for the slow convergence of the F-W method.[9] The highlighting of all these issues has increased interest in alternatives to the F-W algorithm.

Literature Review / Overview of Emerging Equilibrium Assignment Algorithms

VHB conducted a literature review to obtain information on emerging equilibrium assignment algorithms. There are two major algorithms emerging (or in some cases, reemerging) as potential improvements over the link-based F-W algorithm: path-based and origin-based.

Path-based Algorithm

Different from a link-based solution, a path-based algorithm (also called a route-based algorithm) for equilibrium assignment provides a complete picture of the travel pattern and offers modelers the capability to keep track of the distribution of the O-D flows among the different routes as well as the corresponding turning details.

---

[5] The level of approximation is inversely related to the number of assignment iterations; that is, more iterations bring the solution closer to a true equilibrium, and therefore directly related to the level of computational power used to run the forecasting model.
[6] Dial, *ibid.*, and Jayakrishnan, *et al* (1994).
[7] See MWCOG (2006).
[8] See Wolfe (1970).
[9] See Janson and Zozava-Gorostiza (1987)

Path-enumeration algorithms were first proposed in the late 1960s.[10] At that time they were infeasible because of the computing power required to store all utilized paths from all origins to all destinations. As computing power became greater, cheaper, and more available over the last decade, path-based algorithms were re-examined as a solution for equilibrium assignment. The path-based algorithm currently uses a gradient projection method or other algorithms to reach convergence faster and more efficiently.[11] After an initialization with an all-or-nothing assignment, the path-based algorithm searches for other paths with shorter travel time between each origin-destination pair and shifts some traffic from previously identified paths to new shorter paths. In the course of the path search, the paths with zero flow are dropped. Path-based methods are still generally considered more computationally-intensive when compared to link-based and origin-based methods. PTV's VISUM modeling platform includes the option of a path-based algorithm as part of its equilibrium assignment module, as does Caliper's TransCAD software.

Origin-Based Algorithm

Origin-based algorithms (OBAs) attempt to retain the advantages of path-based algorithms; e.g., providing immediate route flow interpretation, while further reducing computational requirements. While the solution variables of both the F-W and path-based algorithm are link flow and path flow, the OBA defines the solution variables in an intermediate way between links and paths. The main variables for the OBA are origin-based approach proportions, which allow efficient storage of a complete description of the route flows. The OBA has three distinct advantages:

- The capability to deliver detailed solutions.
- Substantially lower computation time.
- Lower memory requirements compared with path-based algorithms.

Bar-Gera[12] presented and implemented the first OBA, for which the underlying concept is acyclic flows. An acyclic network does not include any cycles. A cycle could be a travel path around a city block or traversing opposite directions on the same roadway segment. The algorithm consists of two main steps: update the restricting acyclic subnetwork and shift flow within the subnetwork. The main solution variables are approach proportions and are updated when the flow shifts from high-cost alternatives to low-cost alternatives. Then the route proportions are determined as the product of approach proportions of all links along the route. Finally, the route flows are calculated using origin-destination flow and route proportion. An acyclic restricting subnetwork is maintained for every origin so that only the links that are included in this subnetwork are assigned approach proportions and unused links are removed. Therefore, only routes that are limited to the links in the subnetwork can be used. The computation efficiency of the OBA results from the following features:

[10] See Dafermos and Sparrow (1969).
[11] See Bertsekas and Gafni (1983).
[12] 1999.

- Boundary search procedure with well-estimated search direction
- Restricted acyclic origin-based subnetwork.
- Origin-specific topological ordering of the nodes.

Boyce[13] applied the Bar-Gera OBA to solve the user-equilibrium traffic-assignment problem in a practical large-scale roadway network in New Jersey. Different alternatives were tested to evaluate the addition of two proposed ramps. Compared to the F-W method, the OBA achieved highly converged solutions with significantly fewer iterations. In an OBA, the routes serving the OD pairs are efficiently identified and optimized because all the destinations for each origin are grouped together. In general, origin-based assignment methods require more computational resources than the F-W method but less than path-based methods. Other experimental results for medium and large model networks have demonstrated that the OBA can efficiently find a highly accurate solution for equilibrium assignment, but that the process still takes an excessive amount of computing time. There is a free, open-source OBA available for download from the Internet.[14]

Dial introduced an updated path-based user-equilibrium traffic assignment algorithm which eliminated the need for path storage.[15] However, Caliper's later research work classified Dial's algorithm as an origin-based equilibrium method because it decomposes the UE problem into a sequence of single-origin problems on acyclic sub-networks or "bushes."[16] Using these simpler sub-networks, it efficiently locates and shifts flow from costly paths to the cheaper paths until the costs of all used paths are within a user-specified range of the cheapest path. Dial's algorithm has several primary benefits:

- It avoids oscillation when approaching equilibrium and achieves a precision unreachable by the F-W algorithm regardless of the network's size and congestion level.
- It avoids explicit path storage and enumeration by restricting attention to a relatively few path segments in sequential acyclic sub-networks; this in turn improves computational efficiency.
- It uses the relative gap for measuring solution quality.
- It provides a "warm start" feature to compute a new equilibrium much faster using the solution obtained for a similar, previous problem.

Dial tested and reported the new algorithm's performances on two networks, and in both cases it significantly outperformed both the F-W algorithm and the Bar-Gera OBA. The

[13] 2004.
[14] See http://www.openchannelfoundation.org/projects/Origin-Based_Assignment. The site includes executable code provided by Bar-Gera for the OBA as well as a Chicago transportation network and trip tables.
[15] 1999 and 2006.
[16] For purposes of organization we have followed Caliper's characterization and included Dial's algorithm with the origin-based methods.

new algorithm routinely achieves the precision F-W was never able to approach, and it reached a relative gap below 10^{-3} in less time than the OBA.

Building on their previous research, Caliper Corporation has implemented an origin-user equilibrium (OUE) method modified from Dial's "Algorithm B" which demonstrated superior performance in reaching tight equilibrium within much lower computational times.[17] The OUE establishes an order-dependent acyclic sub-network from each origin to all destinations and calculates shortest path more efficiently. During each iteration, the algorithm examines and updates the origin-based link flow to improve travel time. Caliper has done an empirical comparison of alternative traffic assignment methods which demonstrates the OUE method is a promising and feasible solution based on its high level of convergence, fast computing time, and modest memory requirements.[18] Caliper's OUE has been incorporated into TransCAD 5.0, which is currently in beta release and will have a final release this summer.

Status of New Equilibrium Assignment Techniques and Computational Advances among Major Forecasting Software Vendors

The impact of new assignment algorithms on computer hardware requirements and subsequently model run-time remains a chief concern among travel forecasters. Currently it takes approximately 18 hours to run the TPB travel demand model (v2.1D#50) on a computer with a 2.99 GHz dual-core processor with roughly 1GB of memory running Windows XP Professional. An increase in run-time is anticipated with the incorporation of the model improvements planned for the Version 2.3 model.[19] Machines with multi-core processors and/or multiple processors have become more widely available and more affordable in recent years, and travel demand software vendors have been working to take advantage of the increased computing power. While all traffic assignment methods can benefit from multithreading and/or distributed processing when more than one processor is available, there is particular benefit for advanced equilibrium algorithms due to the complexity and volume of calculations required for large networks. VHB contacted the major travel demand forecasting software vendors: Citilabs, Caliper, PTV, and INRO, to discuss their current implementation or plans for implementing both advanced computing processes and advanced traffic assignment methods. In general, the findings are as follows:

- Citilabs will include origin-based assignment in Cube Voyager 5.0, to be released in May 2008; their current release focuses on improving model run-time under F-W assignment by using distributed processing.

[17] See Slavin (2006).
[18] Ibid.
[19] Nearly all of the increase in run-time is due to the implementation of the nested logit mode choice model with speed feedback within the v2.3 model; on a machine with a 3.73 GHz Xeon processor and 2GB of memory, run time was reduced to 12 hours compared with the statistics on v2.1D #50 above, but increased to 21 hours when using the nested logit mode choice model.

- INRO is working on incorporating advanced traffic assignment algorithms and advanced computational methods to its software, but did not say when these features would be available.
- PTV currently includes advanced traffic assignment algorithms in VISUM; advanced computational methods are under development.
- Caliper includes both advanced traffic assignment algorithms and advanced computational methods in TransCAD.

Citilabs

Citilabs' new Cube Cluster reduces run time by distributing modeling tasks across multiple processors. There are two methods to distribute model processes: intrastep distributed processing and multistep distributed processing. The former splits zone-based tasks from a single step into groups based on the availability of processors while the latter distributes the independent steps to available processors. Users may define the cluster range using model scripts. Cube Cluster will run on multiple computers which share Windows files or a computer with multiple processors. However, the hardware setup for Cube Cluster is not automatic. Cube Cluster does not impose scaling restrictions on the number of processors or machines in the cluster, although there are obviously practical limits due to physical space, cost, and other issues. A Cube Cluster license costs $1,500, plus $1,500 per node machine for licenses of Cube Voyager.

PTV

PTV's VISUM modeling software provides a path-based multiclass assignment implemented in the equilibrium procedure to distribute demand over the network. It keeps track of all utilized routes and equilibrates flow over different paths. Equilibrium is reached by multiple iterations based on an all-or-nothing assignment or an incremental assignment as a starting solution. The outer iteration step searches for the new routes in the system (those routes with lower impedances) while the inner iteration step balances the network by shifting vehicles among the competing routes. VISUM's path-based algorithm provides two advantages: 1) it stores the paths for later analysis, and this feature in turn allows for 2) path reloading, where a previous assignment is used as the starting point for a new assignment (so-called "warm starts"). Furthermore, VISUM takes advantage of path compression techniques to improve computing efficiency in both assignment processing and path storage.

The latest version of VISUM also includes a new continuous implicit path formulation for the user-equilibrium assignment problem developed at the University of Rome.[20] This method works with time-varying demand and time-varying supply and is reported to be an excellent choice for mesoscopic modeling with reasonable computational requirements and model run-times. So far PTV has released multithreading functionality for VISSIM microsimulation software, but not for the planning software VISUM. The VISUM development team has started to work on this capability, and it is expected that

[20] See Gentile, *et al* (2005).

VISUM will multithread all highway assignment and demand modeling procedures over the next two or three years. VISUM's path-based equilibrium offers level of convergences (10^{-7}) that exceed those used in practice and has path reloading.

Caliper

Caliper has successfully reduced computational time in TransCAD by multi-threading the F-W algorithm with multiple processor or multiple-core machines. In TransCAD, some of the key procedures in travel forecasting are <u>automatically</u> multi-threaded if used in a multiple-processor environment; for example, traffic assignment, which runs almost twice as fast on a dual-processor computer than on a single-processor machine. TransCAD also supports distributed processing or clustering, similar to Citilabs' Cube Cluster. Distributed processing is not automatic due to the complexity and setup and implementation varies with different models. In terms of advanced assignment algorithms, the OBA in TransCAD 5.0 offers exceptionally tight convergence down to as little as 10^{-15}, much better than current standard practice.[21] Caliper's OBA also includes the "warm starts" feature, similar to that found in VISUM.

INRO

INRO's current major release of EMME/3 does not include support for advanced traffic assignment algorithms nor advanced computing processes.[22] Both of these features are under development, but it is not known when they will be fully implemented in the software. Given EMME's history and foundation built on early adoption of the F-W method, it is reasonable to assume that INRO's next major release will include some implementation of both features in order to keep pace with other software vendors and meet the demands of their user base.

Table 4-1 shows a comparison of the major software packages and their features.

[21] Boyce (2004) recommends using convergence of at least 10^{-4} to reach stability of link-flow difference for a large-scale network. Anecdotal evidence suggests that most MPOs converge between 10^{-2} and 10^{-4} and / or fix their number of iterations (like TPB) after repeated tests end up in this range. In the recent TRB survey on the State of the Practice, only 32% of respondents indicating that they used equilibrium assignment also indicated their model's closure tolerance; indicating that many MPOs may use the default settings of their modeling software or don't know certain characteristics of their traffic assignment. Of those responding with closure tolerance, 96% indicate a convergence at a gap between 10^{-1} and 10^{-3} Yet even with the capabilities offered by TransCAD, Caliper recommends using convergence at a relative gap of 10^{-2} for most applications.
[22] Most of INRO's work developing EMME/3 focused on improving the user interface rather than adding model chain features.

Table 4-1: Comparison of Major Forecasting Software Advance Assignment and Computing Features

	Cube/TP+	VISUM	TransCAD	EMME/3
Equilibrium Assignment	Link-based Frank-Wolfe Method	Frank-Wolfe Method, Path-based Multiclass Method	Frank-Wolfe Method, Origin-based Method	Frank-Wolfe Method
Stopping Criteria	GAP, RELATIVE GAP, AAD, RAAD, PDIFF, PDIFFVALUE, RMSE*	Relative Gap	Relative Gap	Relative Gap
Computational Capabilities	Cube Clusters	Warm Start	Clusters Multithreading Warm Start	N / A

* *GAP* – Relative difference in system cost between two iterations; RELATIVEGAP – An alternative GAP measure; *AAD* – Average absolute difference in volumes between two iterations; *RAAD* – Relative average absolute difference in volumes between two iterations; *PDIFF* – Fractional portion of links whose change in volume between two iterations is less than the value of PDIFFVALUE; *PDIFFVALUE* – The value to be used with PDIFF; *RMSE* – Root mean squared error of the difference in volumes between two iterations.

To date, both PTV and Caliper (in VISUM and TransCAD, respectively) have implemented assignment methods touted in the literature as converging more rapidly than the conventional link-based Frank-Wolfe method. TransCAD reports excellent runtimes with the origin-based assignment.[23] The performance of this new method in terms of convergence, runtime and network stability will be best reported by the planning practitioners who are actually using it.

The "warm starts" feature is particularly attractive to planning practitioners since it allows them to perform a series of model runs with feedback and analyze multiple scenarios when making slight changes to the land use or network facilities.

MPO Contacts / Use of New Features

VHB contacted several MPOs to discuss either their use of advanced algorithms for equilibrium assignment, or advanced computing options to improve model performance (run-time), or both (if applicable). Because these features are so new and not yet in wide use, additional contacts were made with state departments of transporation (SDOTs) and one county planning agency to capture the full scope of agencies known to be employing the new assignment techniques. The list of agencies using the new techniques was obtained from the software vendors, and VDOT was contacted due to their interaction with TPB's forecasting work.

[23] Bar-Gera reported much slower run-times using his own code.

Agency	Modeling Platform(s)
Virginia Department of Transportation (VDOT)	Cube
Ohio Department of Transporation (ODOT)	Cube Voyager (Cluster), VISUM
Metropolitan Council of the Twin Cities (Minneapolis / St. Paul MPO)	Cube Voyager (Cluster)
METRO (Portland, Oregon MPO)	VISUM
East-West Gateway Coordinating Council (St. Louis, MO MPO)	Cube Cluster
The Maryland-National Capital Park and Planning Commission, Prince George's Planning Department	TransCAD
Capital District Transportation Committee (Albany, NY MPO)	VISUM

Virginia Department of Transportation (VDOT)

VDOT's modeling group in Richmond currently has no plans to move to Cube Cluster, as 1) they are satisfied with the performance of their individual workstation licenses, and 2) VDOT's "one computer per user" computing policy effectively prohibits use of any cluster or distributed processing feature by anyone except IT staff. VDOT report that their on-call consultant (Corradino) has experienced significant performance enhancement running models using Cube Cluster. Finally, VDOT noted that the latest version of Cube Voyager has the ability to create path databases that store assignment information in a way similar to VISUM for easier select link and other analysis.

Ohio Department of Transporation (ODOT)

ODOT uses Cube Cluster to run a variety of models ranging from small (200 zones) and medium-sized (800 zones) MPO models (written in Voyager and Application Manager) to the Columbus model (1900 zones) and the Ohio statewide model (5000 zones). These last two are both activity-based models written primarily in Java, so Cube (and the distributed processing is used only for the network skims and assignment). Equilibrium closure gap used by ODOT range from 10^{-3} to 10^{-5}. ODOT's run-time gains are so significant that they will not run their large models <u>except</u> under Cube Cluster. ODOT has a significant investment in hardware – approximately $80,000 for a cluster of 9 machines, each with 2 dual-core processors.[24] ODOT's modeling staff reports minimal effort required to begin using Cube Cluster if the users already are familiar with Cube, and offered to share sample scripts as well as discuss specific issues with TPB staff.

[24] ODOT's system is a more robust version of the clusters at the Columbus transit agency (COTA) and MPO (MORPC), where the tour-based models were implemented prior to being used at ODOT and their statewide model developed. More details on those smaller systems can be found in the TPB FY 2006 report on activity-based models.

VISUM assignment was used on a few projects in the area; however, the ODOT staff VHB contacted could not provide more detailed information on the performance of the software.

Metropolitan Council of the Twin Cities (Minneapolis / St. Paul MPO)

The Metro Council tested Cube Cluster by creating a cluster of their three existing modeling workstations – those machines had single-core processors with clock speeds ranging from 3.0 to 3.2 GHz and memory ranging from 0.5 GB to just under 1 GB. A typical model run on the agency's most powerful modeling workstation had a run-time of 35 to 40 hours. Testing with the initial cluster yielded run-times that ranged from approximately 35% to 61% of pre-cluster run-time, depending on the processing methodology and number of machines used in the cluster (the best performance used the multi-step methodology with the maximum available number of machines). The Metro Council was pleased with the test results and is planning to invest approximately $30,000 for a full implementation, with most of that expenditure going to purchase two quad-core workstations for an improved cluster. Cube Cluster will be used for all MPO modeling activities. The Metro Council also reported a relatively easy learning curve for Cube Cluster, noting that most of the startup time was for converting their model execution scripts into a format that best utilized the Cluster features (including some legacy FORTRAN programs). They plan to address these issues more directly in the next version of the model by having it fully implemented in Cube Voyager.

METRO (Portland, Oregon MPO)

METRO uses VISUM assignments for major corridor studies, some of them multimodal and at least one currently with tolls under consideration. In addition, many cities and counties within the Portland area also use VISUM for impact studies and system management plans, so there is compatibility between agencies. Prior to implementation, METRO compared the results of the VISUM assignment to those assigned using INRO's EMME/2 forecasting software. The results were sufficiently comparable that METRO moved ahead with their use of VISUM. VISUM's network structure allowed METRO to more precisely define intersection capacities, which has improved their analytical capabilities.

METRO noted that VISUM's path storage is a significant benefit when performing select link analysis; since all the paths are stored as part of the assignment, there is no need to run a new assignment for purposes of the analysis, and there are resulting time-savings for staff. Assignment run-time is about four hours for a regional network with 2013 zones and 25,000 one-way links, with 3-4 vehicle classes and a high degree of convergence. METRO's workstations use a 2.8GHz processor and 4GB of memory. Assignment results have been stable. METRO will continue using VISUM for regional forecasting work, including future analysis of the regional long range plan. They are also planning to move toward regional dynamic traffic assignment (DTA) in the longer term, which can be performed using VISUM. Finally, METRO noted that the next version of

VISUM will use real numbers instead of integers for assignment, and this switch will lead to a faster and tighter convergence.

East-West Gateway Coordinating Council (St. Louis, MO MPO)

East-West Gateway uses Cube Cluster for all model runs on a three-machine cluster where each machine has the following specifications: single-core 3.6GHz processor and 2GB memory. They report significant time savings due to Cube Cluster and a minimal learning curve.[25] East-West Gateway staff indicated a willingness to answer further questions.

The Maryland-National Capital Park and Planning Commission, Prince George's Planning Department

TPB staff are already familiar with the Prince George's TranForM model, which is essentially the v2.1D model with a disaggregate zone structure for Prince George's County, a conflated, true-shape regional network, and a few model structure changes, all currently implemented in TransCAD 4.8 and soon to be in production using TransCAD 5.0.[26] However, by running in TransCAD, the Prince George's model takes advantage of the advanced assignment algorithms and advanced computational methods that are native to the software platform. The Prince George's model runs in about 2-3 hours, running 100 iterations with a relative gap of 10^{-2} and two feedback loops. The modeling hardware was recently upgraded to a quad-core workstation. After moving their production model to TransCAD 5.0, Prince George's will be able to use the Caliper multi-threaded UE, path-based, or OUE algorithm for its assignments.

Southern California Association of Governments (SCAG)

SCAG is currently using TransCAD 4.8 for its regional transportation model and is upgrading the model to TransCAD 5.0. There are 4149 internal zones (4191 total zones) and 65,000 links in the SCAG network. Congestion varies widely among the Los Angeles subregions. Peak period average freeway speed is about 30 mph. Due to the size of the model, the OUE feature is not used but will be tested for version 5.0.

Prior to moving to the TransCAD platform, SCAG used TRANPLAN for their year 2000 model validation, which was the basis for the 2004 Regional Transportation Plan (RTP). They utilized five feedback loops with flow smoothing between loops. For each loop, the assignments were done with a maximum of 30 iterations. For their year 2003 model validation, which is the basis for the 2008 RTP, SCAG is using standard user equilibrium assignment in TransCAD. The model is setup for up to 10 feedback loops and a maximum of 40 iterations with a relative gap of 10^{-2}. With a five loop application, it takes about 24 hours to complete their model on a quad-core PC.

[25] East-West Gateway did not specify current model run-time; however, during testing in a ten-processor cluster, model run-time decreased to eight hours from 48 hours.
[26] See Slavin, *et al* (2006) for more details.

Capital District Transportation Committee ([CDTC], Albany, NY MPO)

CTDC uses VISUM for typical MPO modeling applications, including corridor studies, scenario testing for the New York State Department of Transportation (NYSDOT), and testing of projects for the regional long-range plan. Processing time for a typical assignment is about one hour for 12 user equilibrium iterations with use of a feedback loop back to trip generation and 10-20 minutes without feedback. CTDC's model network contains 1,000 zones and 10,000 links. Traffic assignments use VISUM's path-based algorithm. CDTC staff report satisfaction with the assignment results, stability, running time, and convergence.

The CTDC model was recently used for a series of different tests to improve computational time and network convergence when applying feedback.[27] When applied, the most successful methodology converged to 10^{-7} after between 15 and 20 feedback loops – a computational time of between 1.5 and 2 hours. A relative gap of 10^{-6} was reached after only six feedback loops with six user equilibrium iterations per loop. Tests using the most successful methodology with VISUM's "warm start" feature did not show any significant improvement in performance due to applying a previous solution rather than computing an initial solution for travel cost. This result may be in part due to the relatively small and less congested network in Albany (compared to TPB).

Comparison of Alternative Traffic Assignment Methods

Caliper's recent research summarized an empirical comparison of alternative user equilibrium traffic assignment methods on large-scale regional transportation networks.[28] The methods under the comparison were as follows:

- Caliper TransCAD UE using F-W[29]
- Caliper Path-Based
- Caliper Bar-Gera OBA
- Caliper OUE

The origin-based and path-based algorithms were coded and tested based on the existing literature. Modifications were made in the initial stage of implementation to improve the convergence performance. The tests revealed that the Bar-Gera origin-based method converged tightly but only after very long computational times while the path-based method did not converge well on medium to large size networks until modifications were made to the gradient search. The memory requirement and computing times were still issues for the path-based method on larger networks. The origin user equilibrium (OUE)

[27] See Boyce, *et al* (2007).
[28] Slavin, *et al* (2006).
[29] Caliper uses a proprietary implementation of F-W that reportedly runs faster than comparable algorithms in other modeling programs; the key procedures of this algorithm are now multithreaded to create TransCADs "standard" assignment algorithm.

method reached a tight equilibrium in significantly less computing time than F-W. Furthermore, the warm start feature of the OUE method requires much less time to reach a new equilibrium solution for a similar problem where the user previously obtained a good solution and saved those results. This confers significant benefits to practitioners when performing scenario analyses. The research concluded that OUE makes it feasible to calculate traffic assignments with gaps of 0.0001 or lower with reasonable computation times for virtually all large models in the U.S.

Caliper's most recent research, presented at the recent TRB Planning Applications conference in Daytona Beach, builds on the above work by comparing only the multithreaded UE F-W and the OUE side-by-side using the Prince George's TransForM model and performing multi-class assignments for different time periods and using feedback.[30] This work again shows significant benefits for both advanced assignment algorithms and advanced computing techniques.

Conclusion / Recommendations

It is important to understand that even with the same term "relative gap" or "origin-based method", the calculation equations and the implementation procedures could be totally different in different software packages, and the proprietary nature of software development makes it difficult to make true "apples to apples" comparisons between platforms. It is the practitioner's responsibility to ensure an adequate approximation of the equilibrium solution is achieved within a reasonable computation time in their model networks. Experience and professional judgment are needed to evaluate whether advanced application procedures for travel forecasting projects actually produce meaningful results. Both the Caliper and PTV research stress the importance of repeating their tests with other models and/or other platforms. Finally, other issues besides the assignment algorithm and computational efficiency may affect convergence and assignment run-time – these include model design, zone structure and size, delay functions, network capacities, and others.[31]

Given TPB's commitment to the Citilabs modeling platform, the next logical step is to pursue whatever run-time and convergence gains can be achieved under the TP+/Cube environment. In order to accomplish this, TPB must convert the v2.2 model to a form that can be used under Cube Cluster. At the time of this report, TPB has purchased Cube Cluster and has been working with it. Staff has gotten a demo model to run, but have not yet gotten their regional travel model running under Cube Cluster. They must also identify either existing machines for cluster creation, or purchase new hardware. If new hardware is to be acquired, specifications must be created, and an appropriate level of investment for software and hardware upgrades to support future model applications determined. At this time of this report, TPB has purchased a modeling server (it has two dual-core Xeon processors) to use for both Cube Cluster and general model runs. Any distributed processing work will also be conducted on this machine.

[30] See Slavin, *et al* (2007).
[31] These thoughts were echoed by Dick Walker of Portland METRO in his response to questions.

TPB modeling staff should maintain contact with representatives from St. Louis, Minneapolis / St. Paul, and Columbus as a resource throughout this process, and should follow-up immediately with Citilabs to discuss any issues encountered while moving the production model into Cube Cluster. The level of benefits reported by Cube Cluster users should improve the TPB model performance sufficiently while Citilabs implements alternatives to F-W in future versions of Cube. TPB should also consider using the model to test the efficacy of Citilabs' future implementation of any advanced assignment algorithms.

References

Bertsekas, D. P. and E. M. Gafni. Projected Newton Methods and Optimization of Multicommondity Flows. IEEE Transactions on Automatic Control, Vol. AC-28, pp. 1090-1096. 1983.

Boyce, David, O'Neill, Christopher, and Scherr, Wolfgang. New Computational Results on Solving the Sequential Procedure with Feedback. Paper presented at TRB 2007 Planning Applications Conference, Daytona Beach, FL, May 2007.

Boyce, David; Ralevic-Dekic, Biljana; and Bar-Gera, Hillel, "Convergence of Traffic Assignments: How Much is Enough?" *Journal of Transportation Engineer*, ASCE, January/February 2002.

Cube Cluster case studies from Citilabs website – http://www.citilabs.com, downloaded 2007.

Dafermos, S. C. and F. T. Sparrow. The Traffic Assignment Problem for a General Network. Journal of Research of the National Bureau of Standards, Vol. 73B, pp. 91-118. 1969.

Dial, Robert B. Algorithm B: Accurate Traffic Equilibrium (and How to Bobtail Frank-Wolfe, Volpe National Transportation Systems Center, Cambridge, MA, July 25, 1999.

Dial, Robert B., "A Path-Based User-Equilibrium Traffic Assignment Algorithm that Obviates Path Storage and Enumeration," *Transportation Research Part* B, March 13, 2006.

Email from Dick Walker, Portland METRO

Email from Guoxiong Huang, Southern California Association of Governments

Email from Lubna Shaoib, East-West Gateway Coordinating Council

Email from Mark Filipi, Metropolitan Council for the Twin Cities

Email from Jeremy Raw, VDOT

Email from Rebekah Anderson, ODOT

Email from Howard Slavin, Caliper Corporation

Email from Michael Clarke, Citilabs

Email from Michael Florian, INRO

Email from Fang Yuan, PTV America

Email from Chris O'Neill, CDTC

G. Bellei, G. Gentile and N. Papola. A within-day dynamic traffic assignment model for urban road networks, Transportation Research Part B: Methodological Volume 39, Issue 1, January 2005, Pages 1-29.

Janson, B., and Zozaya-Gorostiza, C. (1987). The Problem of Cyclic Flows in Traffic Assignment. Transportation Research, 21B, pp. 299-310.

Jayakrishnan, R., W. K. Tsai, J. N. Prashker, and S. Rajadhyaksha. A Faster Path-Based Algorithm for Traffic Assignment. In Transportation Research Record 1443, TRB, National Research Council, Washington, pp. 75-83. 1994.

Metropolitan Washington Council of Governments, National Capital Region Transportation Planning Board. Results of FY 2006 Travel Forecasting Research. Produced under contract by Vanasse Hangen Brustlin, Inc., 2006.

Slavin, Howard; Brandon, Jonathan and Andres Rabinowicz. "An Empirical Test of Alternative Traffic Assignment Methods," TRB 2006 Annual Meeting, Washington, D.C. January 2006.

Slavin, Howard; Brandon, Jonathan, Rabinowicz, Andres, and Srinivasan Sundaram. "Achieving Planning Model Convergence," TRB 2007 Planning Applications Conference, Daytona Beach, FL, May 2007.

Spielberg, F., and P. Shapiro._Metropolitan Planning Organization Travel Forecasting State of the Practice. Presentation to Transportation Research Board Annual Meeting, Washington, DC, 2007.

P. Wolfe. Convergence Theory in Nonlinear Programming. In Integer and Nonlinear Programming, edited by J. Abadie, pp. 1-36, North-Holland, Amsterdam. 1970.

Task 5 -- Research Techniques for Peak Spreading Analysis

This page intentionally left blank

Traffic congestion in large metropolitan areas has become so acute that many commuters are adjusting their departure and/or arrival times for work and other destinations to avoid the worst of what is now called the "peak period". The adjustments in departure times combined with travel times that can last beyond the peak hour have led to the phenomena of peak spreading, where the peak hour demand on a particular roadway exceeds the peak hour capacity and causes demand to shift to the "shoulders" of the peak hour, or the hours adjacent to the peak hour. This situation is so pronounced in the TPB region, that most of the major freeways in the areas have peak periods that last from roughly 6 AM to 10 AM in the morning and 3 PM to 7 PM in the evening where stop and go traffic is common throughout.

TPB requested that VHB review and summarize the state of the practice and the state of the art with regards to modeling peak spreading at the MPO level. VHB began this effort by reviewing the recent MPO survey and following up with staff at large MPOs with characteristics similar to the TPB to gain further insight and documentation into their peak spreading modeling efforts. The results of this research are summarized into state of the practice (most typical) and state of the art (new or unique approaches).

In addition to synthesizing the results of the research, VHB also conceptualized additional approaches to modeling peak spreading at the regional level that may benefit TPB in the future.

Current TPB Practice

TPB currently uses time-of-day factors to divide the daily trip tables into three time periods, AM peak, PM peak, and off peak time periods. These trip tables are then assigned to the regional network using congested skims for the peak periods and free flow skims for the off peak assignment. TPB uses a series of volume-delay functions for assignment with the primary outputs being link volume and speed. TPB then utilizes a post processing procedure where the final assignments are divided into hourly increments and if the hourly volume exceeds the capacity of a particular link, the excess volume is shifted to adjacent hours and the link speeds updated. The post-processing procedure is typically only run for air quality analyses.

State of the Practice

Most metropolitan planning agencies use time-of-day factors which are applied to the daily trip tables output from the mode choice model. The factors are typically derived from household survey data and validated to some degree with traffic counts. There are a number of limitations to this approach including:

- Regional time-of-day factors do not capture the temporal variations in demand throughout the region. For example, in the Washington region, I-270 in Montgomery County would have different peaking characteristics than US 50 in Prince George's County.

- The time-of-day factors are applied for the entire peak period which does not capture the variation of demand within the peak period. A number of large MPOs in addition to TPB use this method, including SCAG (Southern California Association of Governments, BMC (Baltimore Metropolitan Council), and SEMCOG (Southeastern Michigan Council of Governments).

- The time-of-day factors do not "see" congestion. For example, the same factors are applied throughout the Washington region despite the large variation in congestion patterns. The factors are not adjusted based on congestion because there is currently no feedback from assignment to post mode choice where these factors are applied.

As volume delay functions are used to calculate speeds associated with link v/c ratios, the impacts of traffic control and roadway constraints at specific points in the network are not explicitly considered. This constraint, combined with high v/c ratios that prevail in oversaturated networks such as the TPB network, often result in unrealistic speeds which necessitates further post processing to link volumes and speed during the air quality conformity analysis.

The research revealed several variations that MPOs use to mitigate these limitations to some degree. The North Central Texas Council of Governments (NCTCOG, the Dallas/Fort Worth MPO) uses a modified volume-delay function in the form:

Total Travel Time=
(travel time at the uncongested free flow speed) + link congestion delay.

The congestion delay consists of taking the minimum of two values: a minutes-per-mile parameter C or the v/c calculated delay using the curve shown in Figure 5-1:

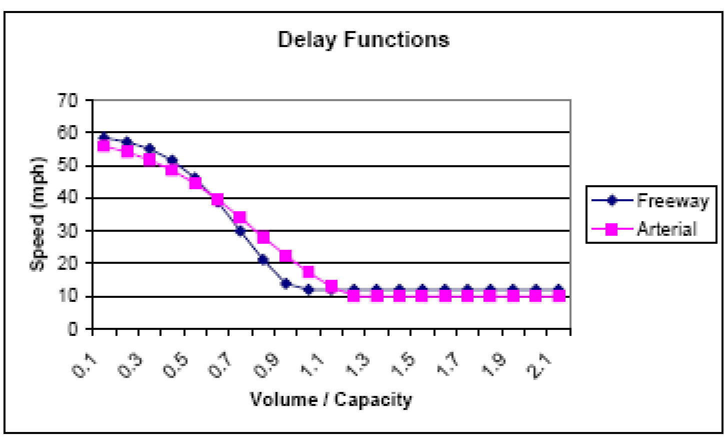

Figure 5-1: NCTCOG Volume-Delay Function (Source NCTCOG)

This process effectively caps the minimum speed on any link in the network to prevent unrealistically low speeds from being fed back into trip distribution and mode choice; however it is acknowledged that floor speeds violate a property of user equilibrium formulation which could lead to problematic convergence. Furthermore, these types of cliff functions or caps are problematic when it comes to New Starts applications with FTA, which is a particularly important consideration for TPB.

Another variation of the time-of-day factoring approach is used by Metro (Portland MPO), which utilizes additional time periods for assignment to better capture the peak demand within the overall peak period. In addition to the 3 hour AM peak period time-of-day factors, Metro also calculates time of day factors for the 2 hour AM peak as well. Likewise, time-of-day factors are applied for both the 4 hour PM peak period and the 2 hour PM peak. This gives a better approximation of demand in the "peak within the peak period), though is still subject to the same limitations with time-of-day factors in general.

To overcome the limitations in the regional assignment (i.e. unrealistically low speeds, over-assigned links), most MPOs post-process assignment results during conformity analysis. A typical post processing approach would look at each link in the network and divide the time period volume into each hour of the day. An analysis is completed testing if any one hour of volume exceeds the hourly capacity for the link. If this is the case, then move the excess volume to the preceding and following hour. Moreover, to overcome unrealistically low speeds output from assignment, separate speed calculations such as the one developed by Richard Dowling and Alexander Skabardonis[32] is used where:

Average Link Speed = Average Queue Speed * (Average Queue Length/Length) + uncongested speed * (1-Average Queue Length/Length))

WHERE:
Uncongested Speed = 1.24 * Survey Speed (fc,h)/ (1+ (V/C) ^ 11
Fc= functional class
H= hour of day
Avg. Queue Speed= Capacity/lane * 25 feet/vehicle
Avg. Queue Length= Average Queue * 25 feet/vehicle
Average Queue= (Q1 + Q2)/2
Q1= Queue at start of time slice
Q2= Q1 + (1 hour traffic/lane – 1 hour capacity/lane)

State of the Art

While time-of-day factoring procedures are considered as state of the practice, there are a few innovative, state of the art approaches to modeling peak spreading. The Puget Sound Regional Council (the Seattle-Tacoma MPO) has two mechanisms that account for peak spreading within the modeling process. In the AM and PM peak (3 hour) assignments, the delay functions incorporate a factor ranging from 0.455 (at v/c=0.0) to 0.333 (at v/c=1.0) to allocate the 3-hour volume to the worst hour for calculating delay. This accounts for flattening of the peak hour within the peak period on a link-by-link basis.

[32] 1992.

The model also includes a time-of-day model which calculates the shares of trips in each time period by direction for auto trips within each homed based purpose which accounts for spreading outside the 3-hour peak periods, on a TAZ-to-TAZ basis. The schematic of the time-of-day model is shown in Figure 5-2:

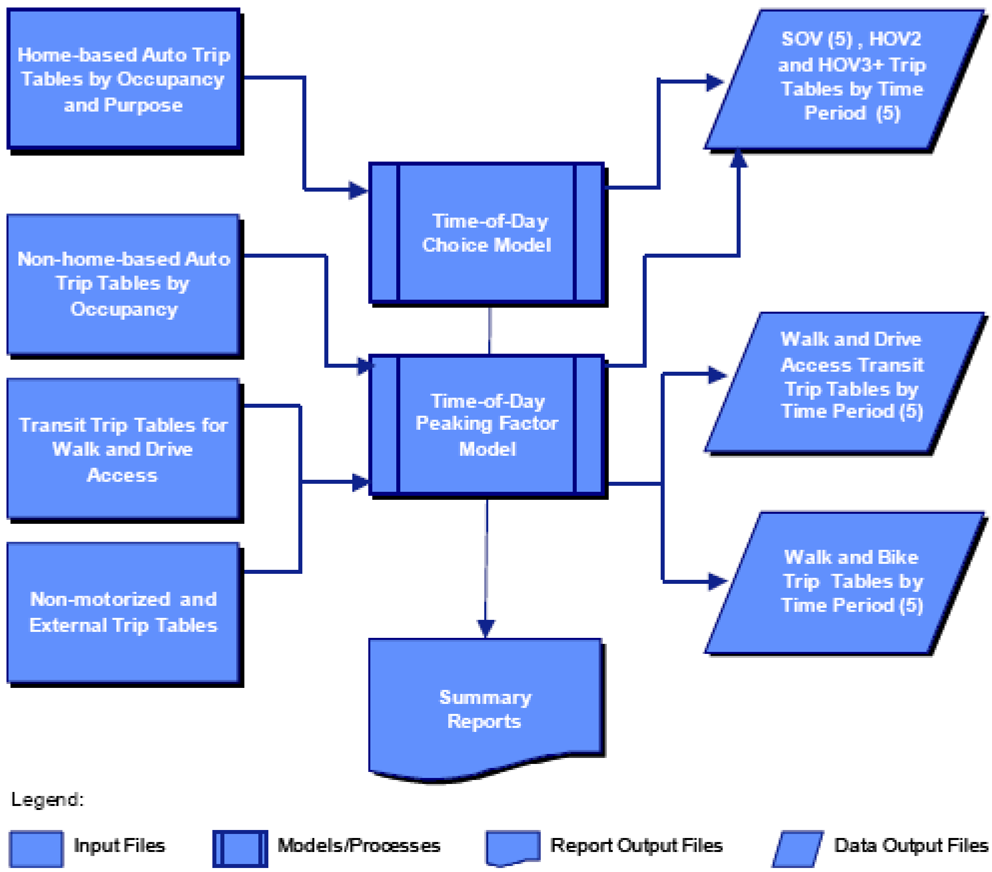

Figure 5-2: PSRC Time-of-Day Model (Source: PSRC)

The Metropolitan Transportation Commission (MTC) is the metropolitan planning organization for the nine-county San Francisco Bay area; MTC has also developed a logit model to address peak spreading. The model is a binomial logit choice model with the choices of AM peak (two-hour) period departure and non-AM peak period departure. The choice is estimated using data from the 1990 Bay Area household travel survey, using data variables such as free-flow and AM peak period congested travel time, trip distance, household income, and dummy variables for bridge crossers, carpooling and retail employment. Highway assignments were calibrated and validated against 1990 daily and peak period traffic volumes and peak period speeds.

This peak-spreading model has a tendency to divert trips from the peak period to the shoulders of the peak period due to increased congestion levels. The result is that the peak period traffic volumes are sometimes lower than the peak shoulder period traffic volumes, yielding too fast speeds in the peak period and too slow speeds in the shoulder period. This was called the "snow plow" effect by MTC, with traffic piling up on the shoulders to allow traffic to flow during the peak period. The quick fix to this problem was to prepare a four-hour AM period traffic assignment based on peaking factors derived from the household travel surveys. The slower of the two-hour and four-hour AM peak period assignments are used to feed back to all mode choice models for purposes of forecast equilibration.

Many theoreticians believe that activity-based modeling is the answer to most of the time-of-day questions the profession faces, including peak spreading, with the idea being that if a model is estimated based on daily activities, some with time constraints others without; then we will be able to better model household's responses to future congestion levels. For example, if a person in a household has to be at work at 9 AM and also currently engages in the activity of purchasing coffee at 8:30 AM on the way to work, the person may well eliminate the non-mandatory coffee purchasing activity in the future just to make it to the required activity (work) by 9 AM.

The Mid-Ohio Regional Planning Commission (MORPC, the MPO for Columbus, Ohio) developed one of the first regional activity-based travel forecasting models in the United States. The model is a disaggregate tour-based model applied with the microsimulation of each individual household, person, or tour, mostly using Monte Carlo realization of each possibility estimated by the models, with the use of a random number series to determine which possibility is chosen for that record.

The model consists of nine separate models that are linked and applied sequentially. In order, these nine models are: Population Synthesis, Auto Ownership, Daily Activity Pattern (mandatory tour generation), Joint Tour Generation, Individual Non-Mandatory Tour Generation, Tour Destination Choice, Time of Day Choice, Tour Mode Choice, and finally, Stops and Trip Mode Choice.

The Tour Destination Choice, Time of Day Choice and Tour Mode Choice models are all logit based and applied together. The "LogSum" composite impedance measure from the mode choice model is available to the other choice models, making them sensitive to changes in travel times due to congestion. The Time of Day (TOD) model is based on the "time windows" concept, accounting for the use of a person's time budget over the day (16 hours available per person). These models are applied at the tour level, yielding the primary destination, time of day, and mode choice for the entire tour, and consider both the out-bound and in-bound portions of the tour.

The TOD model is a hybrid discrete choice departure time and duration model. The model has a temporal resolution of one hour for the modeled period between 5 AM and 11 PM. All tour departures before 5 AM were shifted to the 5 AM hour, and all tour arrivals after 11 PM were shifted to 11 PM. The TOD model is applied sequentially among tours, with mandatory (work, university, and school) tours being scheduled first. The model determines the departure time of each tour and the duration of the activity associated with the tour. Therefore, the 190 departure and arrival time combinations can be applied with relatively few variables. As a result of this time-windows constrained formulation, the timing of the departure and arrival times on both legs of the tour is determined by both the duration of the activities and by the travel times to and from them.

Additional Approaches

There are other potential approaches to modeling peak spreading, including feeding back congested network conditions to pre-assignment where time-of-day factors are applied. This approach would use the congested travel times and v/c ratios output from assignment to adjust the time-of-day factors until all v/c ratios are below an accepted threshold. While this approach would model peak spreading at the regional level, it would still not capture the variations in time-of-day throughout the region.

VHB staff researched and developed another procedure that could be utilized to evaluate peak spreading in the TPB region. This approach would begin with the base year validated travel model and hourly traffic count data at the regional screenlines (ADT count data is collected at the screenline level currently for validation purposes). Hourly counts are also available at many of these locations, and additional counts could be conducted as necessary to augment available count information.

The process would begin by estimating OD tables for the 2, 3, 4, and 5 hour peak periods (the 1 hour peak period is rare in the TPB region). The hourly screenline count data would be used with Cube Matrix Estimation Software to estimate the OD tables for the above referenced time periods. These OD tables would then be divided by the daily regional OD table which would lead to 4 peak period “k” factor tables or more specifically OD tables that reflect the percentage of peak period travel to daily based on existing count information in the region. The “k” factor OD tables would then be applied to the forecast year regional daily trip table, resulting in 2, 3, 4, and 5 hour forecast peak period trip tables which would then be assigned to the regional network using the congested skims.

The resulting assigned networks would represent the 2, 3, 4, and 5 hour peak periods for the forecast year. To examine the duration of a peak period on a particular roadway segment for example, the first step would be to conduct the 2 hour assignment and plotting the resulting v/c ratios (using the hourly capacity x # hours in time period). If the v/c ratio is greater than a pre-defined threshold (1.1 for example), then a 3 hour assignment would be conducted and the roadway segment re-evaluated; this process would be repeated until the v/c ratio for the peak period assignment is equal to or less than the threshold value, so for example, if the 3 hour assignment results in a v/c of 1.34, and the 4 hour assignment results in a v/c of 0.99, then the assumption would be that there would be approximately 4 hours of congested conditions on this particular roadway in the future. The procedure is illustrated in Figure 5-3.

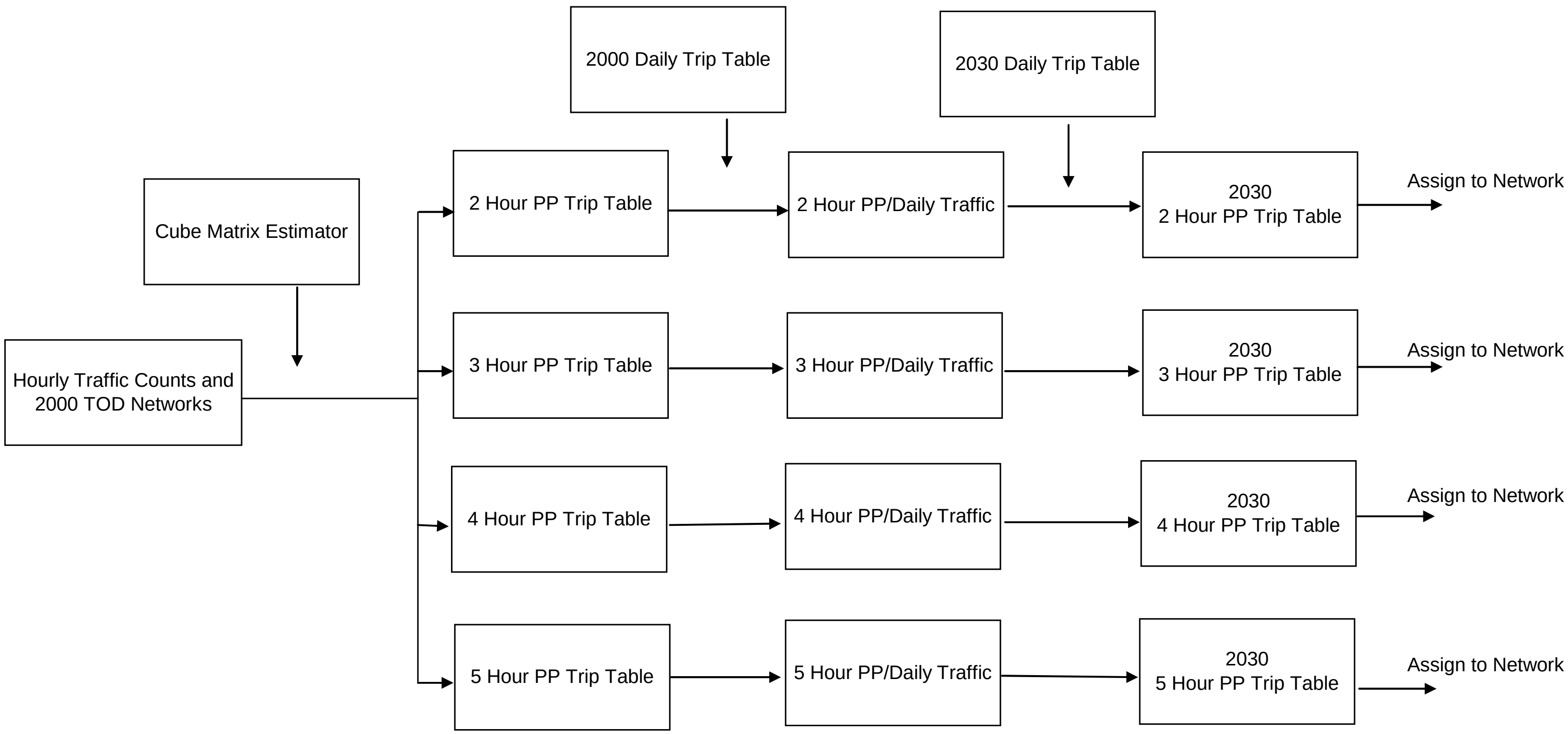

Figure 5-3: Proposed TPB Peak Spreading Approach

The benefit of this approach is that the time-of-day factors are based on actual count data which includes behavior not revealed in household and transit surveys, and the time-of-day factor is effectively disaggregated to the interchange level which would capture the variances in peaking patterns in a region the size of TPB.

VHB completed an initial evaluation of traffic count data to determine data availability and evaluate the peaking characteristics along the I-270 corridor. This data is presented in graphical format in Appendix 5-1. The results of the initial analysis show that locations along I-270 already experience peak periods lasting four or five hours. Appendix 5-1 also contains regional maps showing the availability of observed traffic data for expanding the geographic coverage of the peak spreading analysis beyond the I-270 corridor.

Finally, with the introduction of mesoscopic, simulation based, Dynamic Traffic Assignment models, it is computationally feasible to estimate and assign an OD table on a more detailed network which would include traffic control and a more detailed network representation. This approach would estimate time dependent OD tables (15 minute intervals recommended), in a similar fashion as the previous approach; likewise the forecast peak period OD table would be calculated using k factors. The forecast peak period OD table would then be dynamically assigned to the detailed network which would more accurately reflect the roadway constraint aspect of peak spreading at point locations such as major intersections or interchanges. Additional benefits of this approach is that it could be introduced into the 4-step modeling process (the El Paso, TX MPO is testing DTA in this context currently) and allow TPB to evaluate ITS and ATMS strategies at the regional and corridor level which cannot be done explicitly with the current regional model. However, it should be noted that like any new tool, there are still a number of unknowns with DTA models and OD estimation, so TPB would benefit by testing these strategies incrementally before investing significant resources.

Next Steps

TPB could begin testing one or more of the existing approaches based on staffing and budget availability. VHB has developed additional approaches that could also be considered, one using the existing regional model network, another using DTA and a more detailed network. As developing a regional network for a DTA model would require a significant investment on TPB's part, it is recommended that the regional model approach be tested initially as this would require much less effort and results could be presented several months after project initiation. In the longer run, TPB may want to evaluate using a DTA model for the 4th Step in the modeling process and/or developing a state of the art logit model for evaluating the effects of peak spreading.

References

Rebekah S. Anderson, P.E., Robert M. Donnelly, Modeling of Peak Hour Spreading with an Activity-Based Micro-simulation Model.

Chuck Purvis, Peak Spreading Models Promises and Limitations Presentation at the 7th TRB Conference Application of Transportation Planning Methods Boston, MA.

Cambridge Systematics, PSRC Travel Model Documentation.

Dowling, R., and A. Skabardonis. Improving the Average Travel Speeds Estimated by Planning Models. In *Transportation Research Record 1366*, TRB, National Research Council, Washington D.C., 1992. Pg 68 - 74

SEMCOG, Ozone and Carbon Monoxide Conformity Analysis for the Proposed Amendment of the SEMCOG 2030 Regional Transportation Plan.

Email from Charles Baber, Baltimore Metropolitan Council.

Email from Arash Mirzaei and Ken Cervenka, North Central Texas Council of Governments.

Email from Guy Rousseau, Atlanta Regional Commission.

Email from Larry Blain, Puget Sound Regional Council.

Email from Deng Bang Lee, Southern California Association of Governments.

Portland Metro, Metro Travel Forecasting 2005 Trip-Based Demand Model Methodology Report.

North Central Texas Council of Governments, Travel Model Documentation.

Appendix 5-1: Initial Peak Spreading Procedure for TPB Region

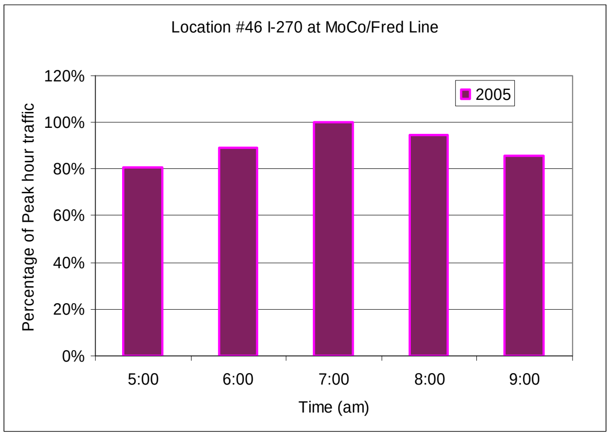

Figure 5-4: Screenline 25 AM Peak Hour Analysis – I-270

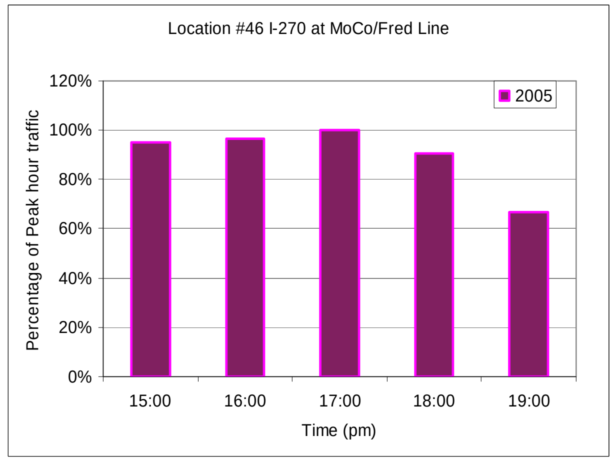

Figure 5-5: Screenline 25 PM Peak Hour Analysis – I-270

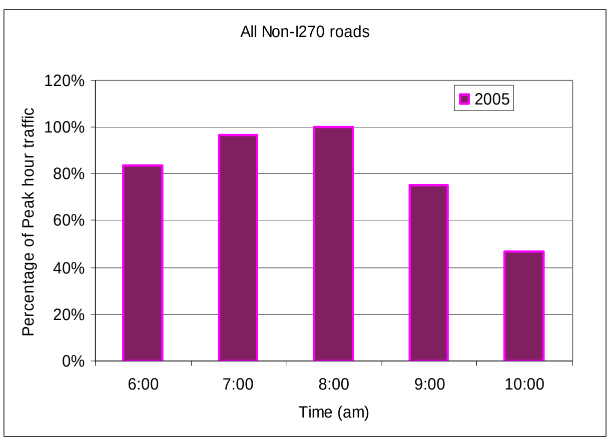

Figure 5-6: Screenline 25 AM Peak Hour Analysis Non-Freeways

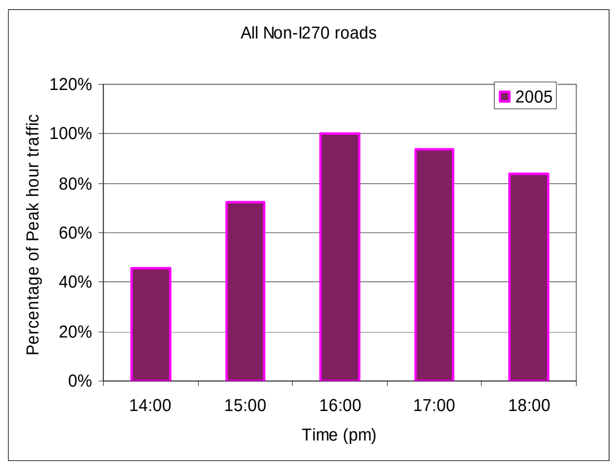

Figure 5-7: Screenline 25 PM Peak Hour Analysis Non-Freeways

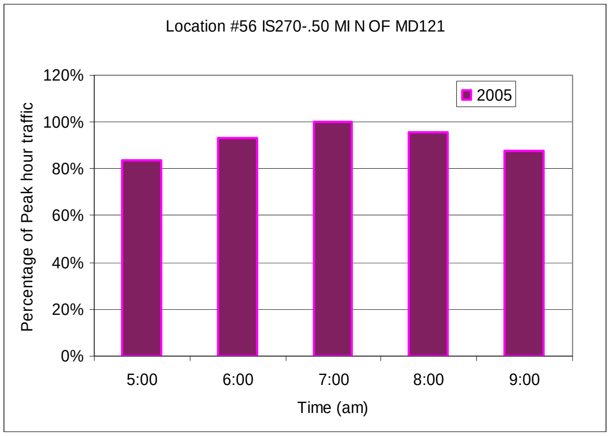

Figure 5-8: Screenline 23 AM Peak Hour Analysis – I-270

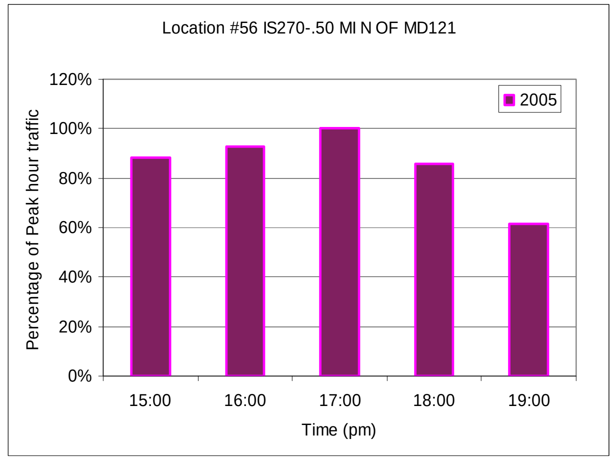

Figure 5-9: Screenline 23 PM Peak Hour Analysis – I-270

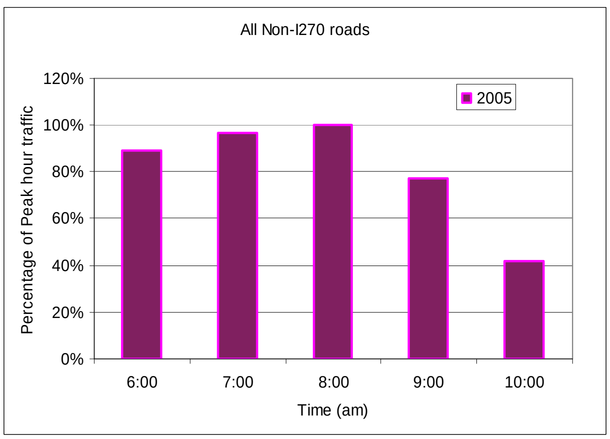

Figure 5-10: Screenline 23 AM Peak Hour Analysis Non-Freeways

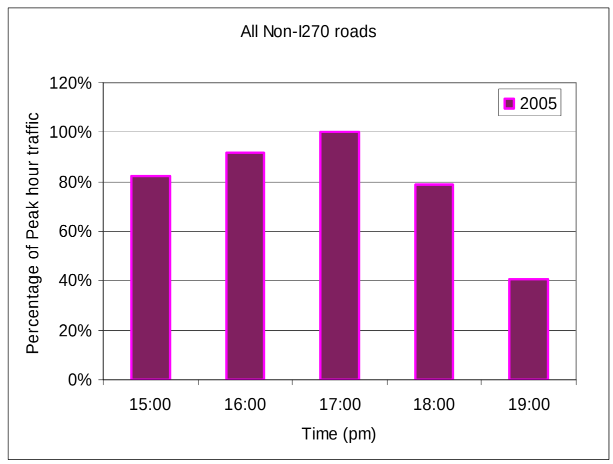

Figure 5-11: Screenline 23 PM Peak Hour Analysis Non-Freeways

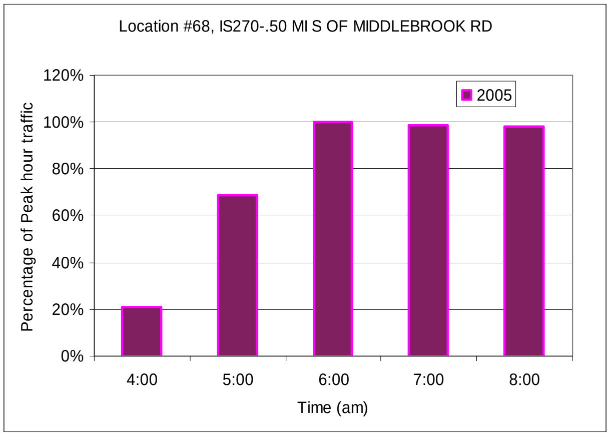

Figure 5-12: Screenline 22 AM Peak Hour Analysis – I-270

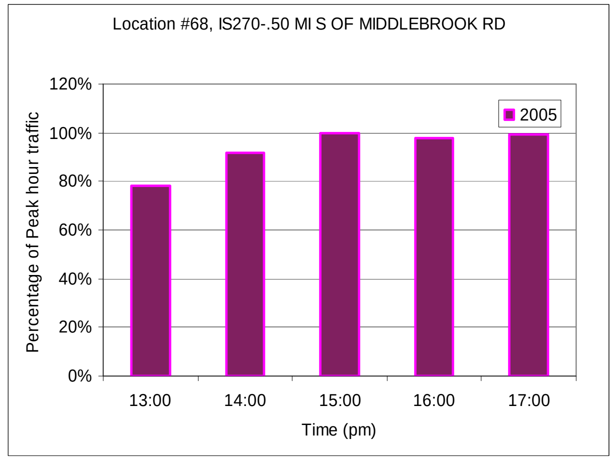

Figure 5-13: Screenline 22 PM Peak Hour Analysis – I-270

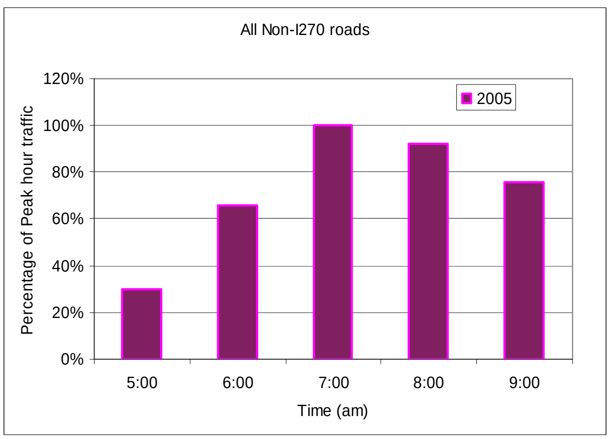

Figure 5-14: Screenline 22 AM Peak Hour Analysis Non-Freeways

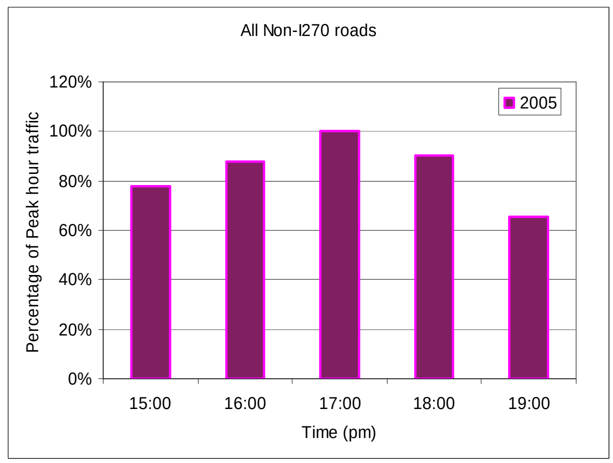

Figure 5-15: Screenline 22 AM Peak Hour Analysis Non-Freeways

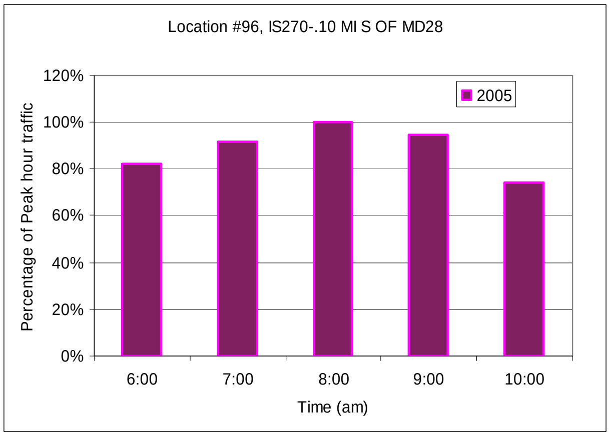

Figure 5-16: Screenline 8 AM Peak Hour Analysis – I-270

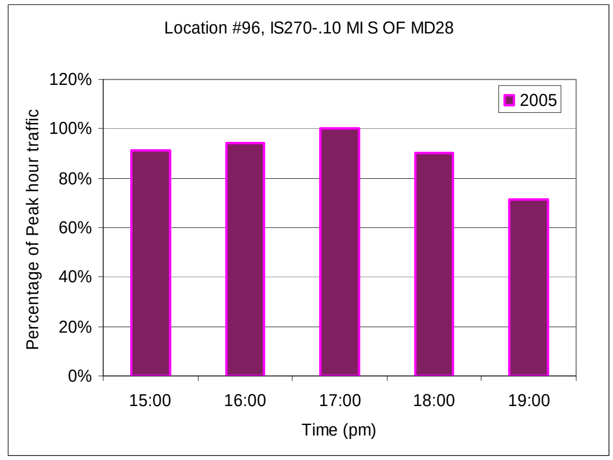

Figure 5-17: Screenline 8 PM Peak Hour Analysis – I-270

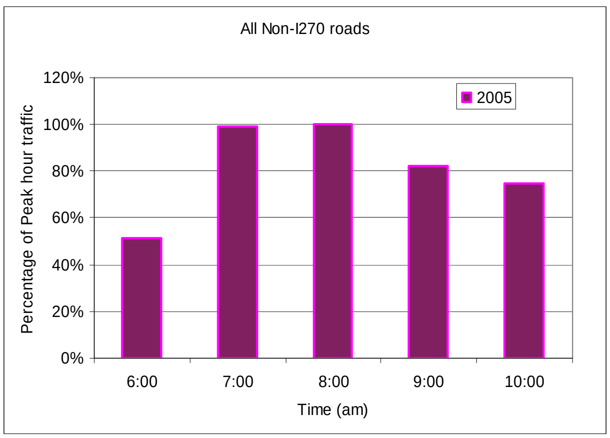

Figure 5-18: Screenline 8 AM Peak Hour Analysis Non-Freeways

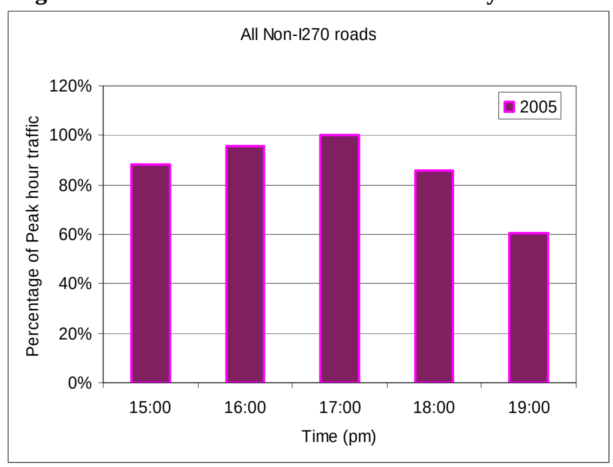

Figure 5-19: Screenline 8 PM Peak Hour Analysis Non Freeways

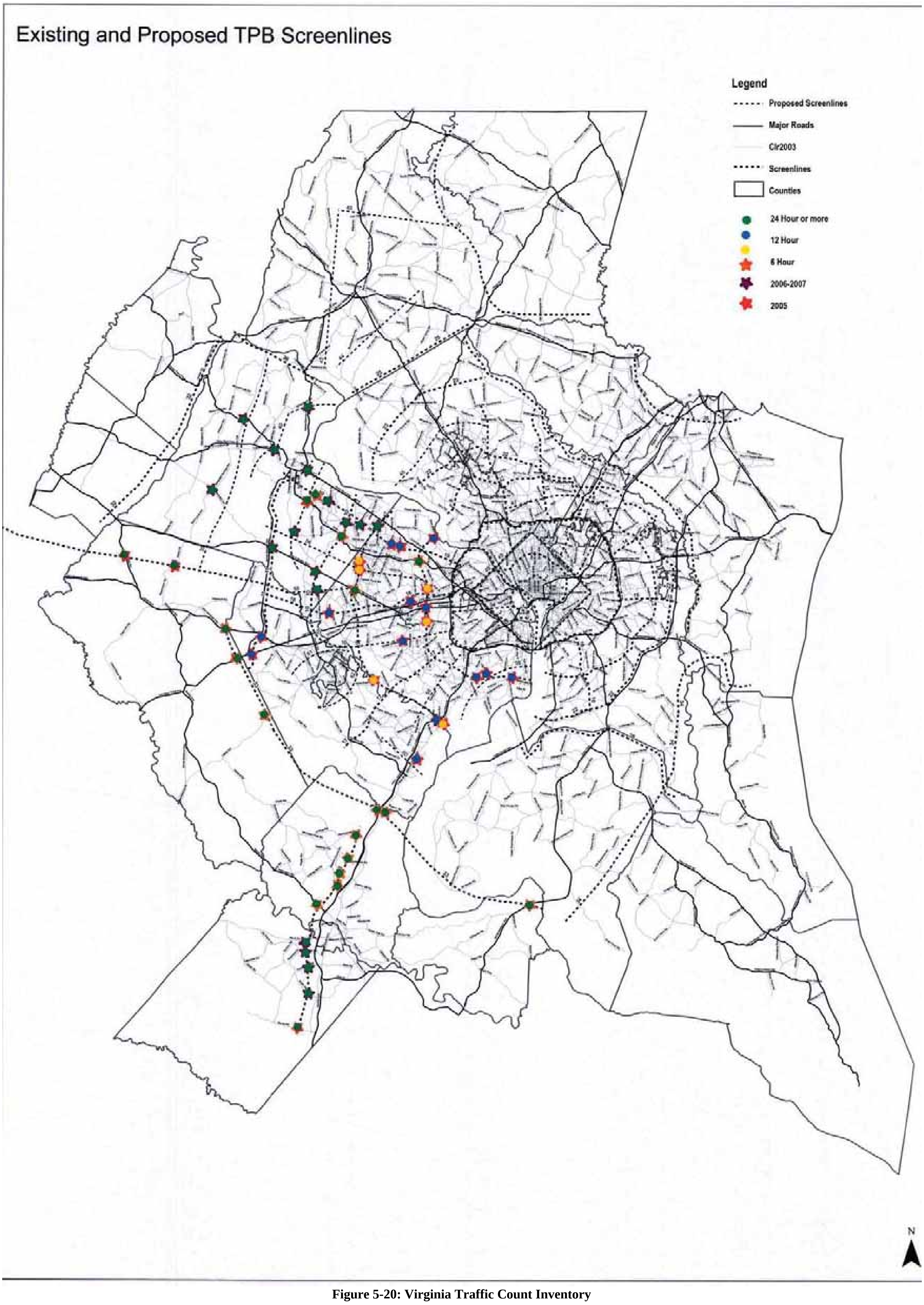

Figure 5-20: Virginia Traffic Count Inventory

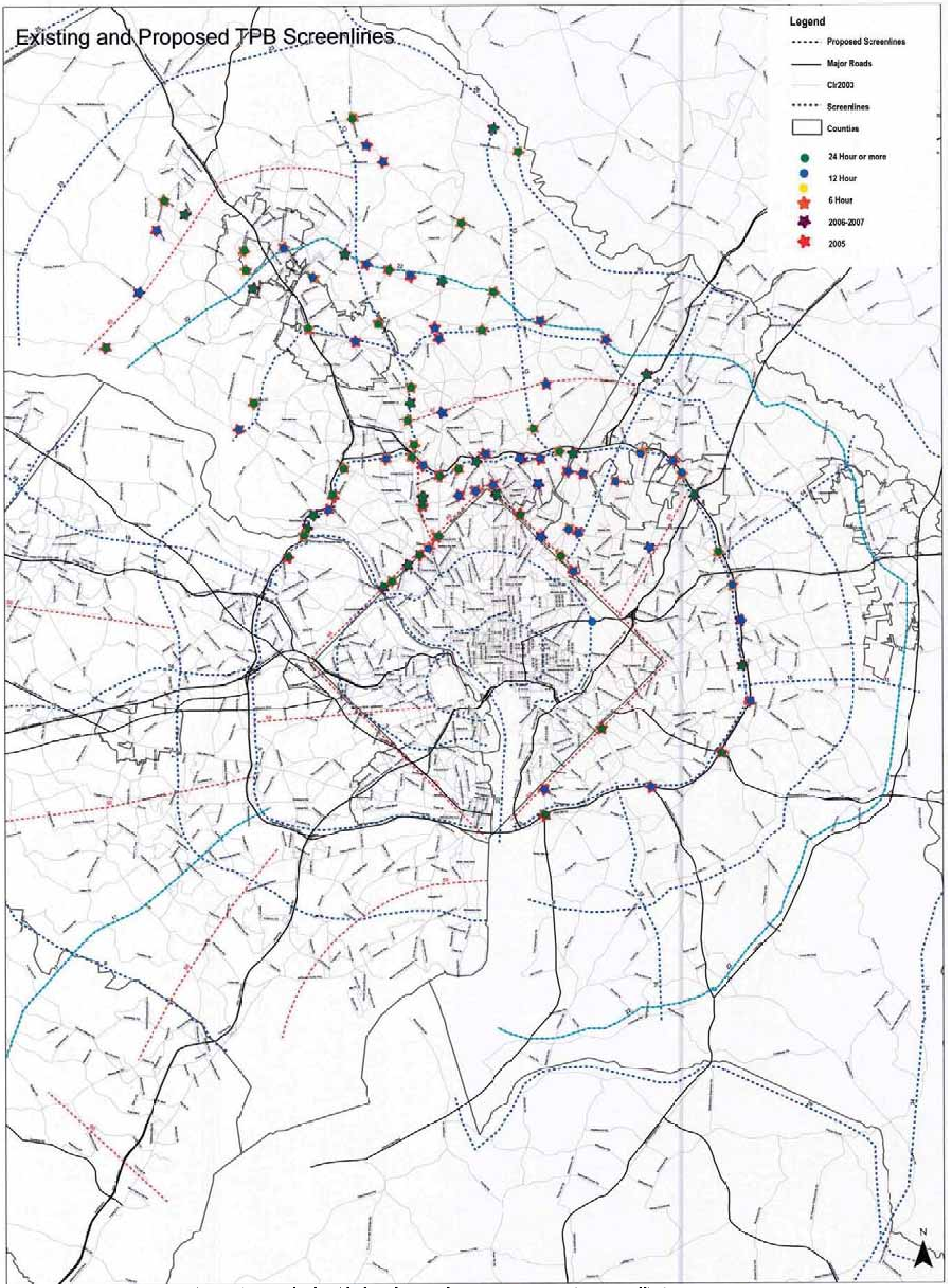

Figure 5-21: Maryland Inside the Beltway and Lower Montgomery County Traffic Count Inventory

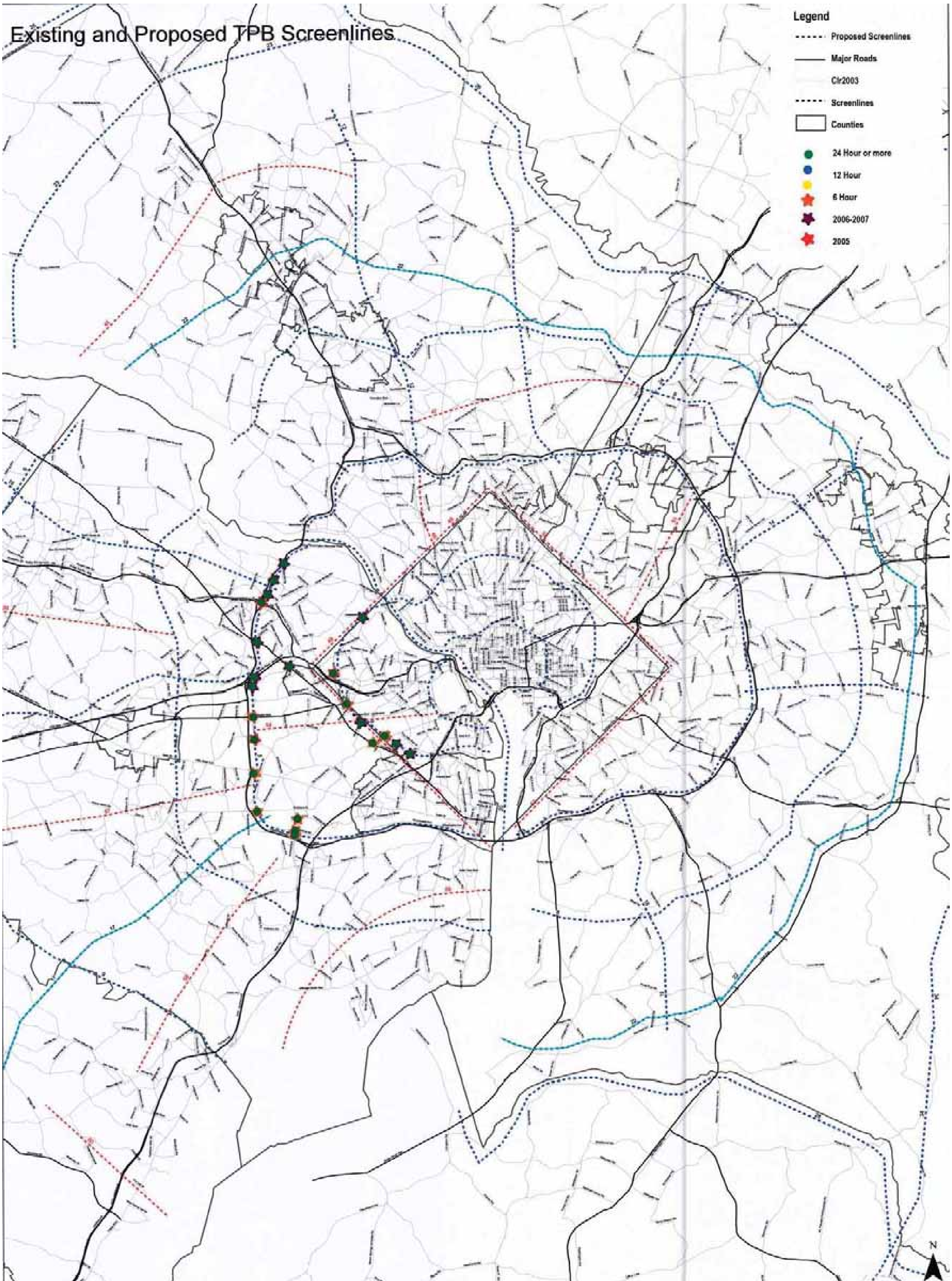

Figure 5-22: Virginia Inside the Beltway Traffic Count Inventory

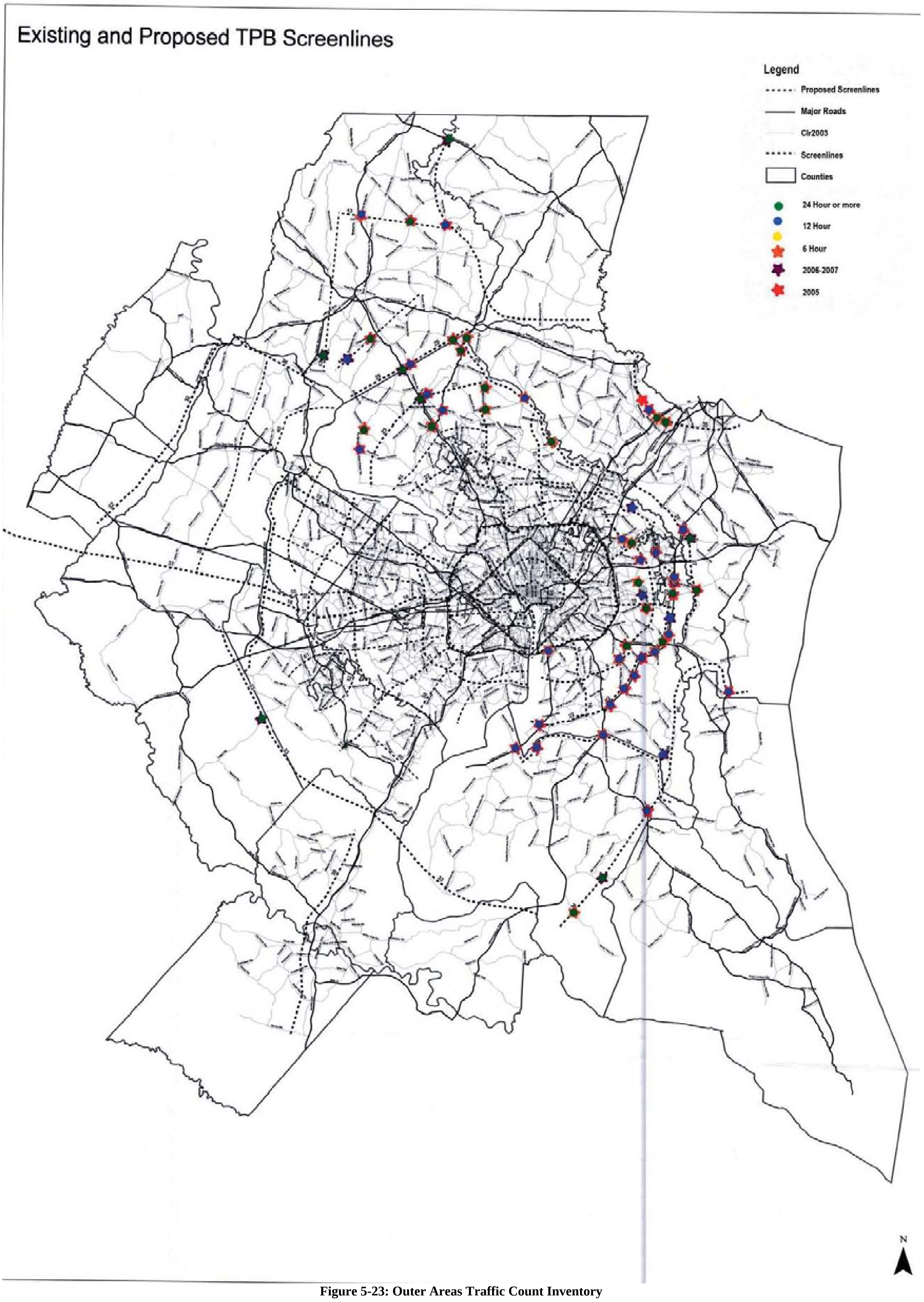

Figure 5-23: Outer Areas Traffic Count Inventory

This page intentionally left blank

Task 6 -- Research the Use of Cutlines for Model Validation

This page intentionally left blank

TPB Staff requested that VHB review the use of cutlines for model validation. TPB currently uses screenlines to compare estimated (model-forecasted) and observed traffic volumes at regional locations as part of validation; however, travel forecasting staff has expressed interest in applying cutlines for future validation tests to compare estimated and observed data at a more detailed level, particularly for forecasting work to support project planning studies. This chapter summarizes our review.

Background: Cordons, Screenlines, Cutlines, and the Current TPB System

"Screenlines" is a generic term used to refer to three hierarchical types of imaginary lines placed across a series of parallel facilities or a series of facilities serving the same travel market. In order from broadest to most tightly focused, these lines are:

Cordons – the lines of a cordon form a closed polygon in order to compare estimated and observed traffic flows into and out of the enclosed study area. Examples with the TPB modeled area include the Metro Core Cordon, the cordon around the Capital Beltway (I-495 / I-95), and the external cordon surrounding the modeled area which is used to compare estimated and observed volumes for external-internal (E-I) and internal-external (I-E) trips as well as modeled area through (external-external / E-E) trips.

Screenlines – actual screenlines capture cross-regional travel flows. The best example of this within the TPB modeled area is the Potomac River screenline (see Figure 6-1).

Cutlines – cutlines capture travel flows through a major corridor. Many of the lines for capturing flows within the TPB modeled area fit this definition even though they are both collectively and individually referred to as screenlines. Screenlines are still an appropriate term within the TPB model for what amount to very long cutlines, although according to FHWA guidelines, cutlines "…should be used to intercept travel along only one axis" (see Figure 6-2).[33]

The "line" of demarcation and definition between a cutline and a screenline can be somewhat blurry. Validation of a large regional model such as the TPB model requires comparison of estimated and observed traffic volumes at a *regional* level; i.e., county-to-county flows or flows to and/or along parallel facilities within a broad, *regional* travel corridor such as "outer" jurisdictions to "inner" jurisdictions. These long screenlines are in fact agglomerations of potential, shorter cutlines that could be focused on smaller corridors such as I-95 and its major parallel facilities (US 1 in Virginia; US 1, US 29, and MD 295 in Maryland).

[33] See Barton-Aschman and Cambridge Systematics (1997). It can be argued that travel patterns in the TPB region have changed radically enough in the intervening ten years since these guidelines were published that meeting this criterion is both difficult and lacking value to the ability to model travel markets, particularly with the percentage of very long and multiaxial trips, such as circumferential travel.

Figure 6-1: Example of Screenline Locations[34]

Figure 6-2: Example of Cutline Location[35]

Comparison of modeled versus counted traffic across cordons or screenlines provides an indication of how well a travel demand model performs in replicating major trip patterns and movements throughout the network. The screenline or cordon will usually correspond with a recognized visible boundary feature (a river or major transportation facility) or a well-delineated political boundary (a county or city border). Screenlines typically encompass all facilities that serve the same definable travel corridor to allow for the fact that the model may not perfectly represent competition between parallel facilities. The definition heavily depends on the delineation of the travel corridor. Historically, cutlines have been reserved for use in project planning studies where the study area is a smaller subset of the regional modeled area and the model must be revalidated so that the estimated volumes adequately match the observed data on the network within the study area. Ideally the cutlines are selected at a very early stage of the study to ensure the availability of reliable observed data for use in revalidation and so that adequate time is available for adjustments to model parameters, if necessary. Most recently, cutlines for project planning work have been used successfully with the TPB model for studies such as the Intercounty Connector (ICC), the Base Realignment and Closure (BRAC) Environmental Impact Statement (EIS) for Fort Belvoir, and the various I-270 studies.

Figure 6-3 shows the existing TPB screenline system. There are 38 regional screenlines in the TPB modeled region in addition to the external cordon. The last model validation compared estimated and observed volumes along the screenlines as well as checking county-to-county flows. Sufficient growth and subsequent changes in regional travel patterns have occurred in the years since the last model validation that some consideration needs to be given to moving screenlines or adding new screenlines. In addition, the screenlines should be easily subdivided for use as cutlines when project planning studies are undertaken or for possible use in regional validation and sensitivity testing.[36]

Literature Review

The primary guiding document on the treatment of screenlines in travel forecasting is National Cooperative Highway Research Program Report Number 255: Highway Traffic Data for

[34] Source: Ibid.
[35] Source: Ibid.
[36] Sensitivity testing, sometimes referred to as dynamic validation, describes the process by which the model's response to specific, targeted changes in land use or network inputs is tested and documented.

Urbanized Area Project Planning and Design (hereafter NCHRP 255). The guidelines contained within NCHRP 255 are so widely used in travel demand forecasting activities around the United States that it is not an understatement to call the report "the bible" on the subject; every other guidance document found in the literature uses the criteria from NCHRP 255 as its starting point and do not radically depart from them.[37] Furthermore, none of the MPOs contacted for this memo follow procedures significantly different than those found in NCHRP 255 or its child documents. Those procedures and guidelines are summarized below:

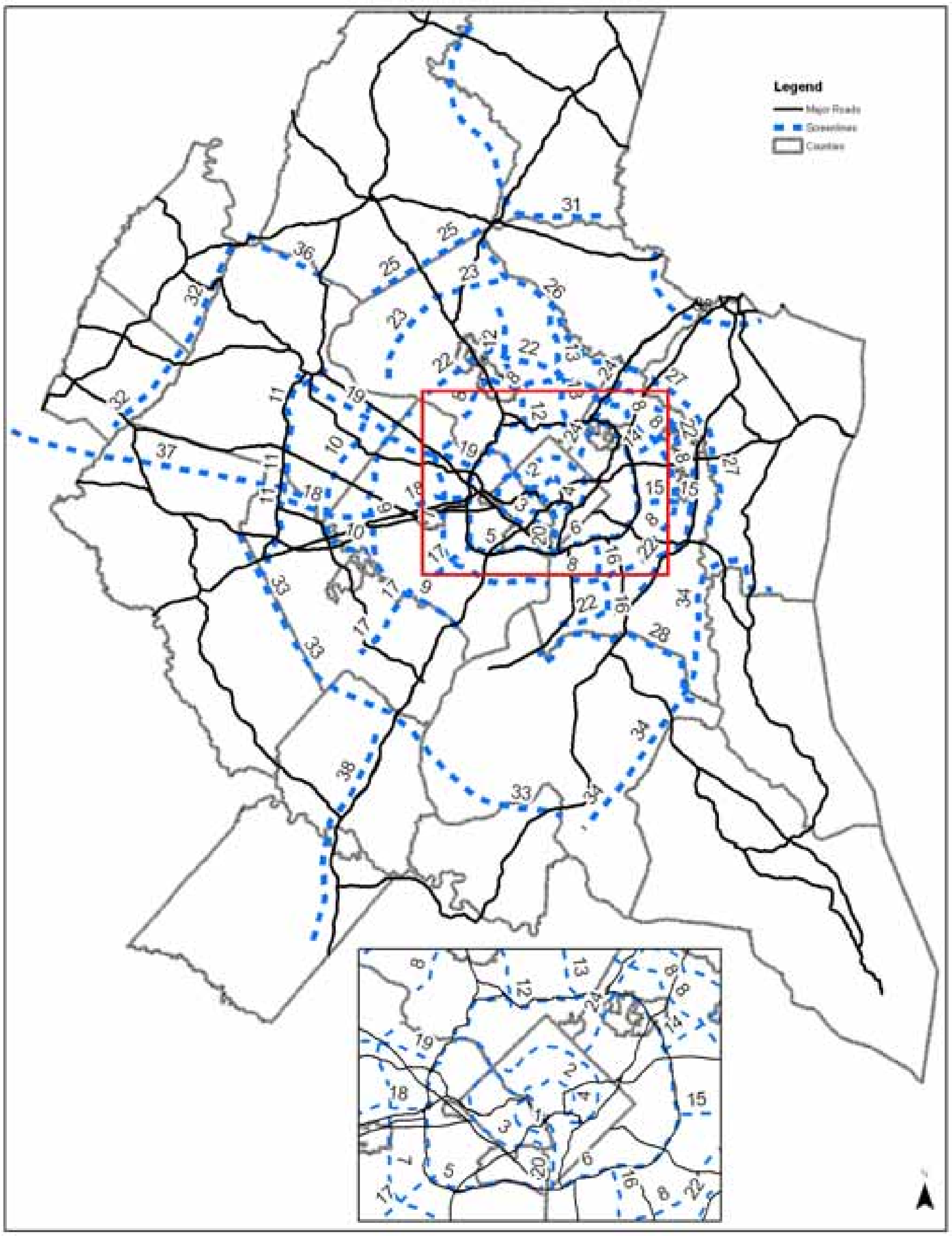

Figure 6-3: Existing TPB Screenline System (Source: TPB GIS)

[37] This includes the relevant sections of Barton-Aschman and Cambridge Systematics (1997) and state-level model reasonableness checking and validation documents.

NCHRP 255 Overview

NCHRP 255 details the procedures that MPOs should use to validate traffic counts and traffic assignment volumes modeled during travel demand forecasting. The methodology includes the use of screenlines as count validation points and includes procedures for adjusting the modeled volumes on each link to more closely replicate count data.

Selecting Screenlines

Creating meaningful and useful screenlines requires familiarity with the transportation network. NCHRP 255 recommends that several rules be followed when selecting locations for screenlines and for determining which links should be included in the analysis.

- A single screenline should capture traffic on all roadways that are alternatives in a corridor. Non-parallel facilities in the same area should not be included in the screenline.
- Zone connectors should not be included in screenline analyses under most circumstances.
- Each screenline should ideally cross between three and seven road facilities. The report recommends ten facilities as the practical maximum.
- Screenlines should only be long enough to capture the recommended number of roadways. Lengths of up to five miles may be appropriate in low density areas, while two miles is recommended in denser urban areas.
- Screenlines should be located between major roadway interchanges (or every two miles). This will allow for checks on the changes in volume along the length of individual facilities.

Base Year Checks

To determine if any adjustments to the model assignment forecasts will be necessary, modeled volumes from a base-year scenario should be compared to actual traffic counts from that year. The comparisons should be done for each identified screenline by totaling the volumes on each parallel facility. NCHRP 255 establishes guidelines for the "maximum desirable deviation" between the modeled volumes and the traffic counts for each screenline. The allowable percentage of deviation decreases as the volumes moving through a screenline increase. If the base year volumes exceed the maximum desirable deviation, several corrective actions may be taken, including:

- Check for and correct errors in the model itself and then re-run the model.
- Extend the screenline, making sure that the additional facilities captured serve the same travel market as those traversed by the original screenline.
- Factor the screenline volumes based on the difference between the base year assignment and the base year traffic counts.

The report contains maximum desirable deviation curves for both individual count locations and screenlines.

Modeled Volume Adjustments

NCHRP 255 provides detailed procedures for adjusting modeled assignment volumes for links on screenlines with larger than desirable deviations. These procedures adjust the volume on each link of a screenline in order to realize forecasted volumes that are closer to the actual traffic counts. These procedures balance volumes on each link while accounting for future changes including increased capacities on specific facilities.

Current MPO Practice

The number, type and location of screenlines vary between MPOs based on the size and geography of the urban area. The Atlanta Regional Commission (ARC) recently increased its number of screenlines from 16 to 22 in conjunction with the expansion of its modeled region to meet conformity requirements. A list of the ARC screenlines and the results of their recent Year 2000 validation is shown in Table 6-1 below. A map of the ARC screenlines is shown in Figure 6-4. The maximum desirable deviation standards are taken from the curves in NCHRP 255 and the calculated deviation values (based on ARC's regression lines that fit sections of the NCHRP curves) applied using a TP+ script.

Table 6-1: Atlanta Regional Commission Screenlines and Year 2000 Validation Results (Source: Atlanta Regional Commission)

Screenline	Assigned Volume	Traffic Count	Volume / Count Ratio	Percent Deviation	Maximum Desirable Deviation (+/-)
Chattahoochee River	1,371,296	1,299,756	1.06	5.50%	8.27%
Inner Rail Ring	1,537,742	1,489,970	1.03	3.21%	7.83%
Outside of I-285	2,769,930	2,778,592	1.00	-0.31%	6.13%
South Atlanta - East/West	683,089	656,921	1.04	3.98%	10.83%
North Atlanta - East/West	621,623	571,619	1.09	8.75%	11.44%
Central Atlanta - north of I-20	1,358,561	1,313,793	1.03	3.41%	8.23%
Corridor south of Marietta	402,583	374,168	1.08	7.59%	13.52%
I-20 Corridor east of Douglasvill	144,106	137,303	1.05	4.95%	20.09%
I-75 Corridor north of Jonesboro	305,607	303,223	1.01	0.79%	14.69%
I-85 Corridor north of Norcross	401,463	375,894	1.07	6.80%	13.50%
GA 400 Corridor north of Buckhead	148,050	159,613	0.93	-7.24%	18.93%
I-20 Corridor east of I-285	226,171	230,181	0.98	-1.74%	16.38%
GA 400 Corridor in Roswell	157,985	197,000	0.80	-19.80%	17.42%
SR 20 Corridor west of Cumming	36,448	32,894	1.11	10.80%	35.33%
I-85 Corridor south of Fairburn	135,972	115,800	1.17	17.42%	21.49%
Lake Lanier	82,811	88,276	0.94	-6.19%	23.92%
I-985 South of Gainesville	82,862	66,900	1.24	23.86%	26.69%
West Region N/S	149,852	133,013	1.13	12.66%	20.35%
East Region N/S	129,794	133,146	0.97	-2.52%	20.34%
I-75 South of Locust Grove	150,441	137,733	1.09	9.23%	20.07%
Alcovy River	153,929	159,708	0.96	-3.62%	18.93%
Flint River	218,479	196,692	1.11	11.08%	17.43%
Totals	11,268,784	10,952,194	1.03	2.89%	3.56%

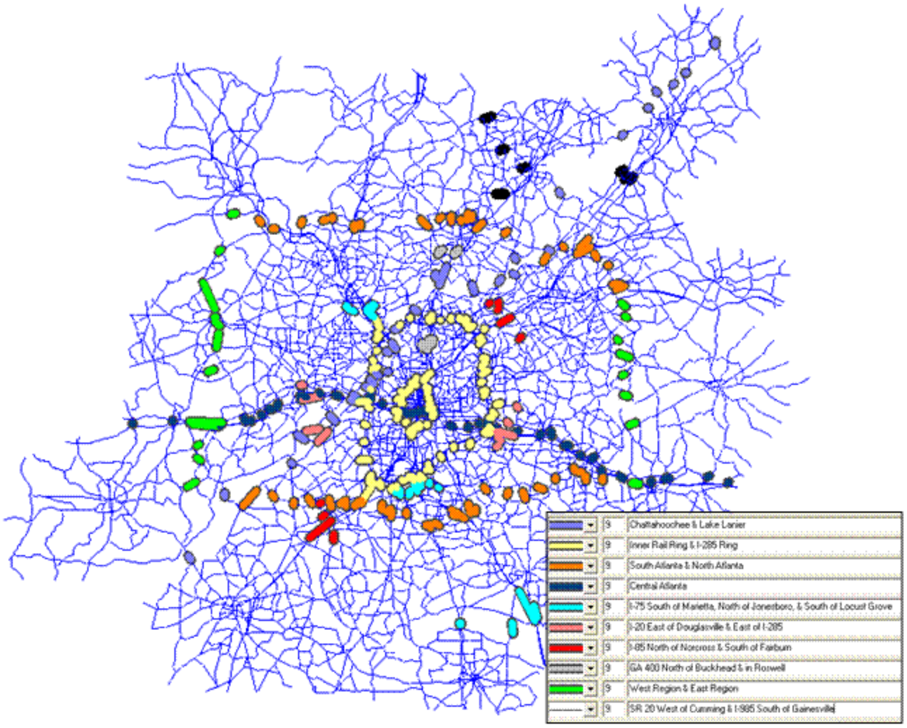

Figure 6-4: ARC Screenlines (Source: Atlanta Regional Commission)

The Baltimore Metropolitan Council (BMC) uses 52 screenlines which are divided into four categories:

- 12 Baltimore City screenlines follow the city borders, the limits of the core area and a few major corridors within the city.
- 24 circumferential screenlines capture traffic flows entering and leaving the city at various distances.
- 11 corridor screenlines capture traffic flows in major corridors throughout the region.
- 5 Local Area Cordons capture traffic leaving and entering secondary urban centers in the Baltimore region (Columbia, Towson, Westminster, Bel Air, and Annapolis).

Maps of the BMC screenlines are shown in Figures 6-5 through 6-8.[38]

[38] The BMC validation report clarifies their use of the word screenlines, stating that "the term "screenline" as used by BMC staff refers to an imaginary line that intersects one or more roads which is used to evaluate traffic flows in an area. Most screenlines used by BMC staff are technically called "cutlines" or "cordon lines." BMC's use is similar to that of TPB.

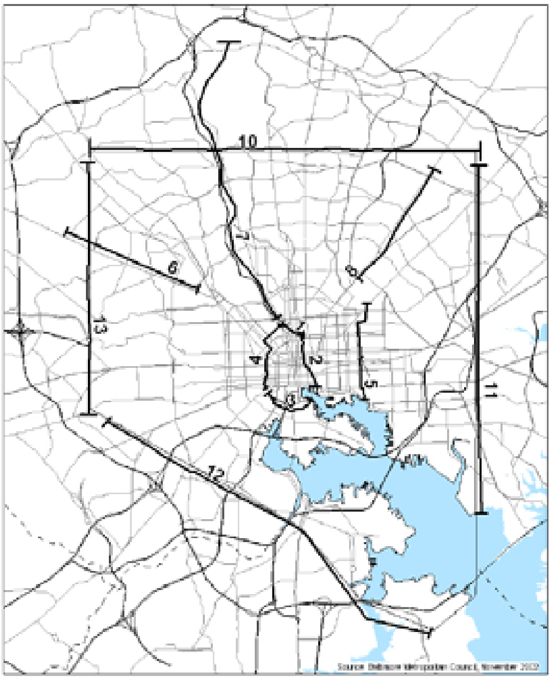

Figure 6-5: BMC City Screenlines (Source: BMC)

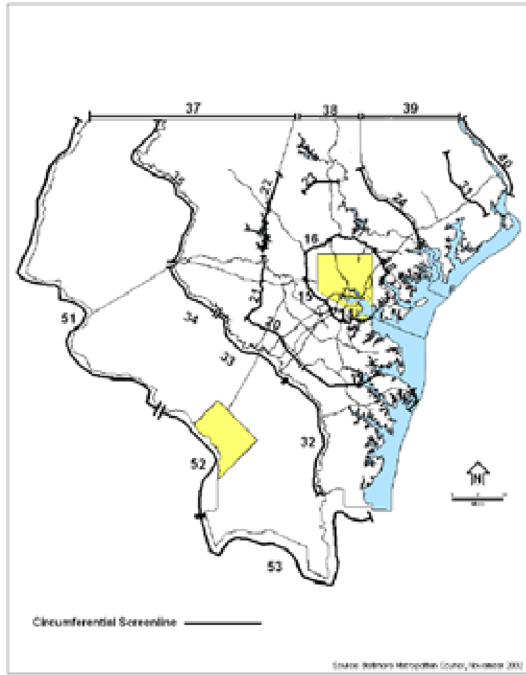

Figure 6-6: BMC Circumferential Screenlines (Source: BMC)

The Regional Transportation Commission (RTC) of Southern Nevada, the MPO for the Las Vegas metropolitan area, uses two sets of screenlines. The first, called "k-factor screenlines" are located on the boundaries of the 18 k-factor districts. These districts form 27 screenlines that are used to measure the flow between adjacent districts of the city. An additional 44 screenlines are used to measure corridor flows on major facilities. The New York Metropolitan Transportation Council (NYMTC) uses an extensive three-tiered system of screenlines that includes volume counts on over 2200 links. 26 screenlines divide the region along county borders, and additional screenlines are used to divide each county into quadrants and sub-quadrants. The Metropolitan Transportation Commission (MTC), the MPO for the San Francisco Bay Area, places screenlines at all county borders and some intervening screenlines within certain counties based on regional travel markets and the level of urbanization. MTC also includes a separate screenline for the eight bridges crossing San Francisco Bay and its tributaries.

The Denver Regional Council of Governments' (DRCOG) 2001 validation of its trip-based model used eight regional screenlines and cordons around downtown Denver and the City of Boulder. The Puget Sound Regional Council (PSRC), the MPO for the Seattle-Tacoma region, uses 71 screenlines for model validation. The Denver and Seattle screenlines are shown in Figure 6-9 and Figure 6-10, respectively.

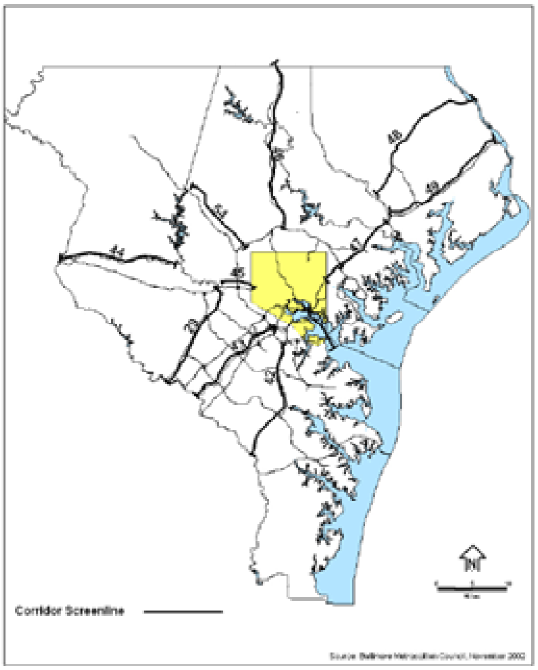

Figure 6-7: BMC Corridor Screenlines (Source: BMC)

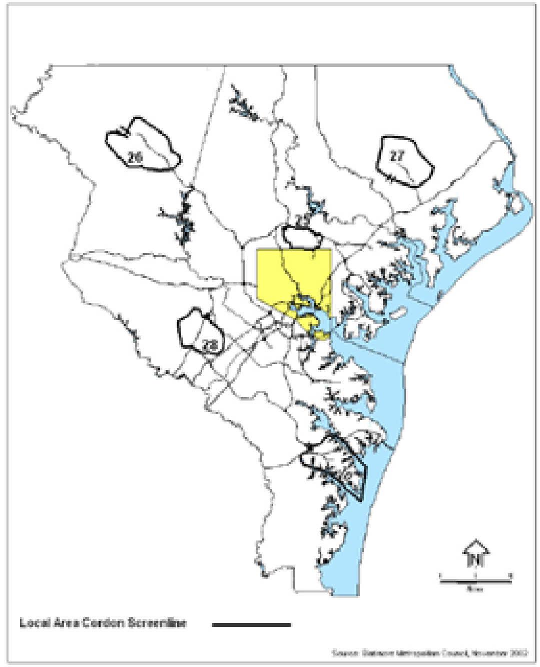

Figure 6-8: BMC Local Area Cordons (Source: BMC)

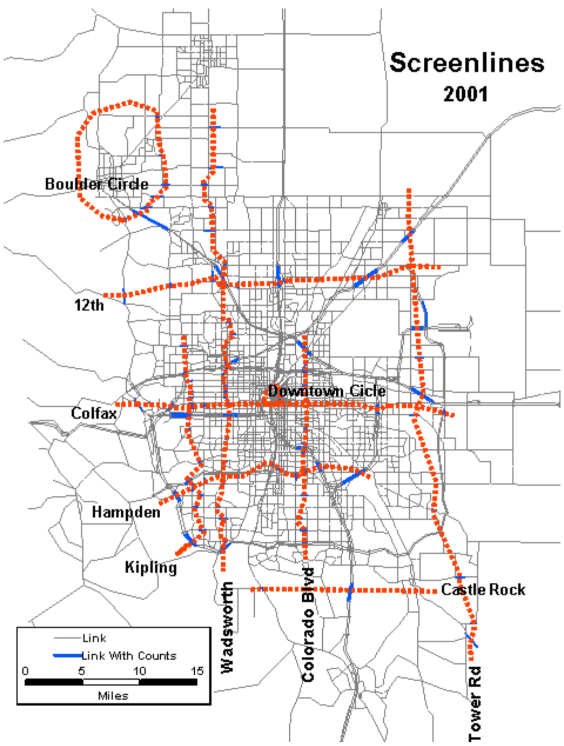

Figure 6-9: DRCOG Screenline System (Source: DRCOG)

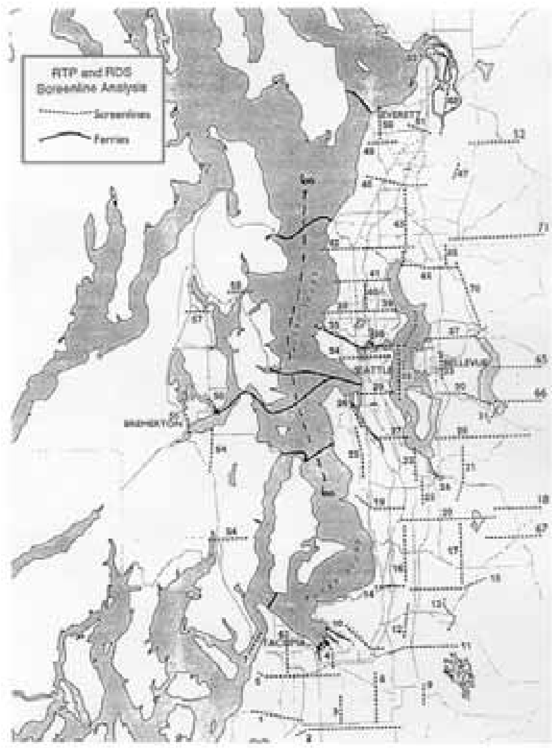

Figure 6-10: PSRC Screenline System [39]

The Southern California Association of Governments (SCAG), the Los Angeles MPO, used 16 regional screenlines for its year 2000 model validation. The Maricopa Association of

[39] Source: Dailey, *et al* (2002)

Governments (MAG) in Phoenix is using a new system of 74 screenlines for its upcoming validation. The recently validated version 4.0 of the Central Florida Regional Planning Model (CFRPM), which covers District 5 (Orlando / Cocoa / Daytona area) of the Florida Department of Transportation (FDOT), used 54 regional cutlines over a nine-county area.[40] The SCAG and MAG systems are shown in Figure 6-11 and Figure 6-12 and the cutlines by area in the CFRPM model are shown in Figures 6-13 through 6-19.

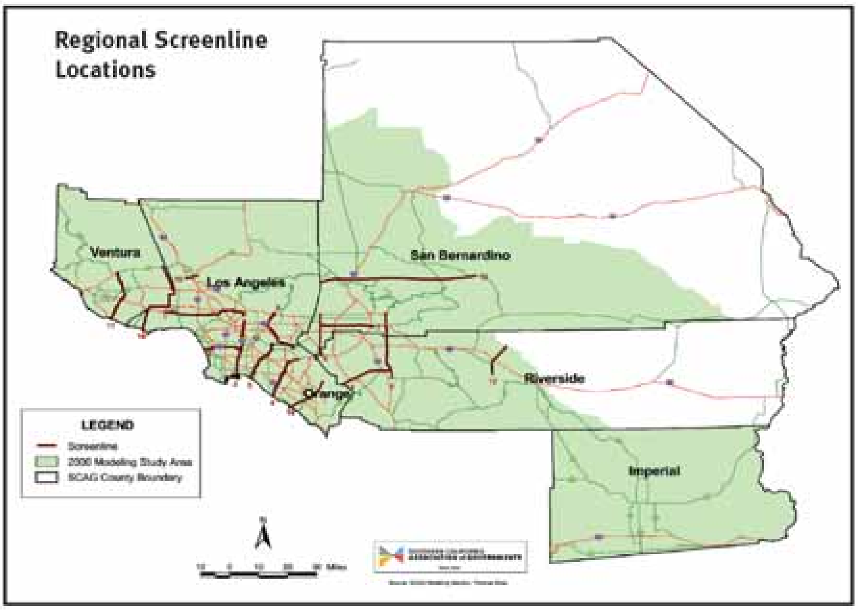

Figure 6-11: SCAG Screenline Locations (source: SCAG)

Figure 6-12: MAG Screenline Locations (source: MAG)

[40] The CFRPM is based on the Florida Standard Urban Transportation Model Structure (FSUTMS).

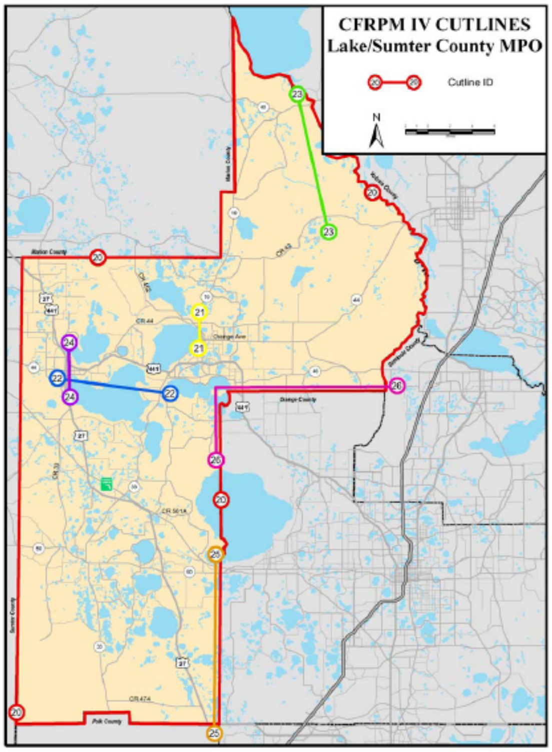

Figure 6-13: CFRPM Lake/ Sumter County Cutlines[41]

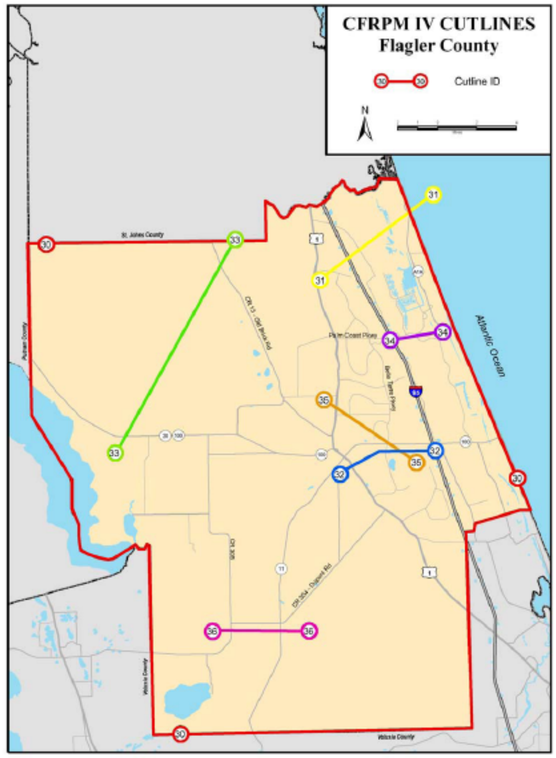

Figure 6-14: CFRPM Flagler County Cutlines

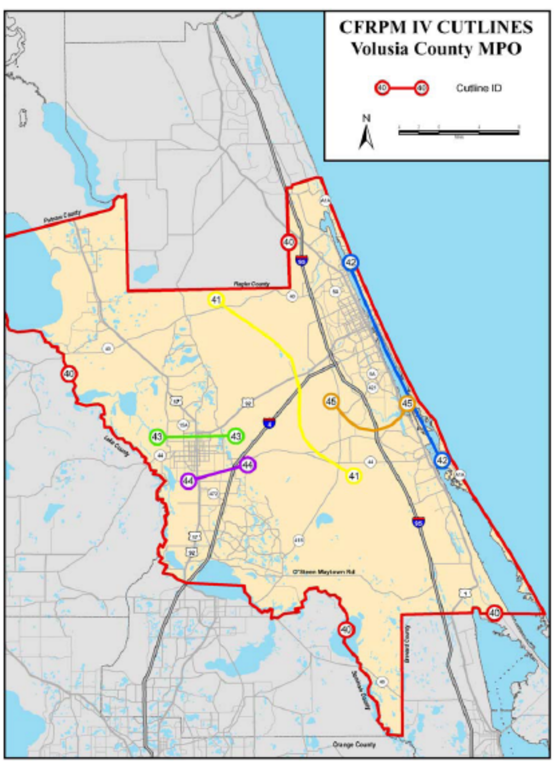

Figure 6-15: CFRPM Volusia County Cutlines

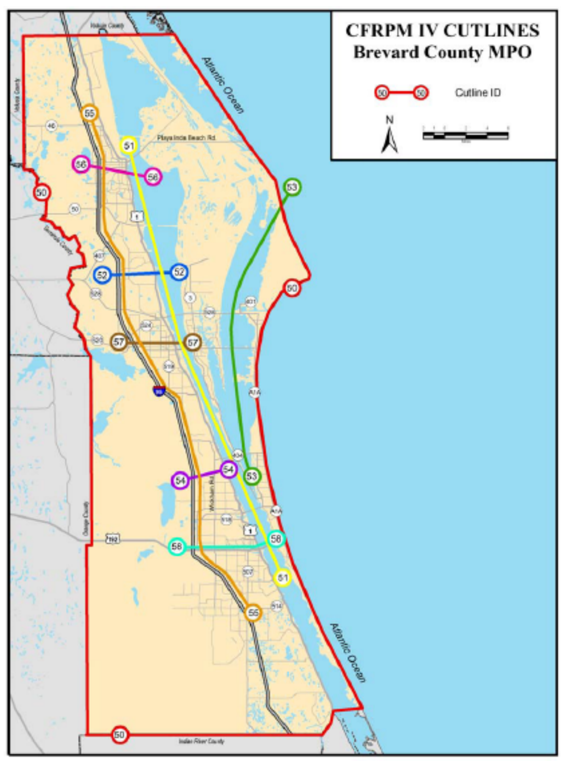

Figure 6-16: CFRPM Brevard County Cutlines

[41] Source for figures 12-19: HNTB (2006).

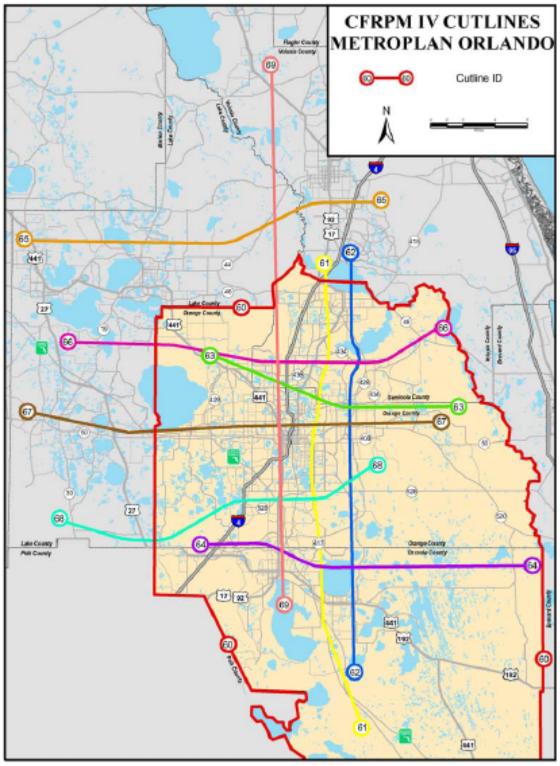

Figure 6-17: CFRPM Metroplan Orlando Cutlines

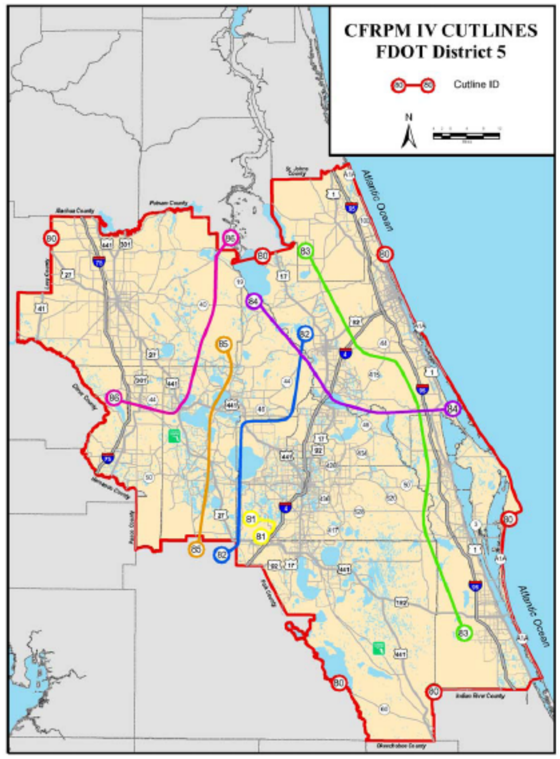

Figure 6-18: CFRPM FDOT District 5 Cutlines

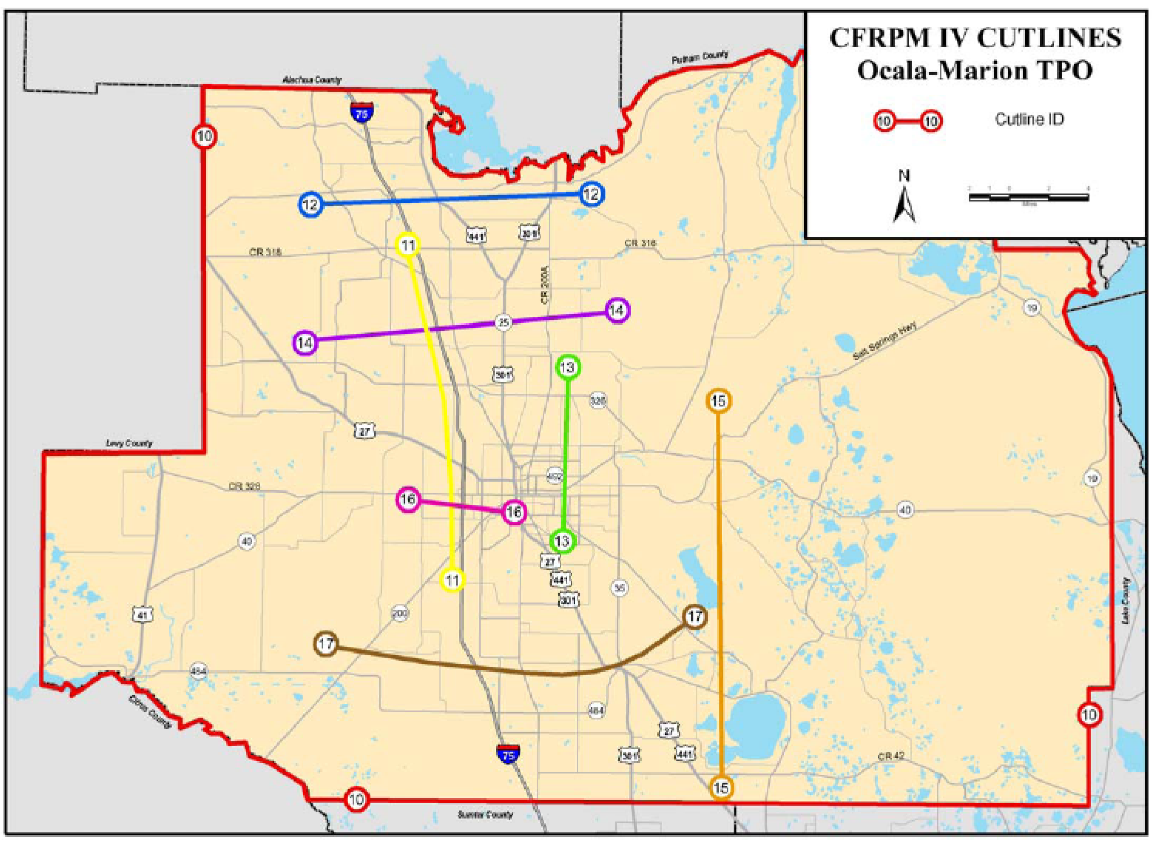

Figure 6-19: CFRPM Ocala / Marion Counties Cutlines

As noted previously, all of the above mentioned MPOs use the volume adjustment procedures and the “maximum desirable deviation” methodology outlined in NCHRP 255 (and refined in federal and state model validation manuals) for validation of screenline volumes.

Proposed New TPB Screenlines

Introduction

The following methodology was used to identify the proposed new screenlines for the TPB region:

- Review existing screenlines.
- Consider changes in regional and sub-regional travel markets based on growth / shifts in population and employment.
- Overlay existing screenline system on 2006 Constrained Long Range Plan (CLRP) projects as secondary measure of where future analysis may be needed.
- Professional judgment.
- Check proposed new screenlines against NCHRP 255 guidelines.

This evaluation process can and should be repeated periodically, particularly with the Transportation Improvement Program (TIP) to serve as a check on model performance and to ensure that the data needed to support project planning studies will be available, especially if new traffic counts need to be taken. Table 6-2 below lists the new screenlines, their location, and their reason for recommendation. Figure 6-20 below shows the 23 proposed new screenlines and Figure 6-21 overlays the proposed new screenlines on the existing screenlines.

Table 6-2: Proposed New TPB Screenlines

Screenline Number	Location	Justification
39	Western Loudoun	Population / Employment Growth
40	North / West of Leesburg	Population / Employment Growth
41	East of Leesburg	Growth; potential future studies of VA 7 and Dulles Greenway
42	West of City of Frederick	Extra-regional growth in Washington County; emergence of Frederick County as employment / shopping destination
43	North of City of Frederick	Extra-regional growth in Pennsylvania; emergence of Frederick County as employment / shopping destination
44	South / East of City of Frederick	Supplement for studies in I-270 and I-70 corridors
45	Germantown	Supplement for project planning studies in I-270 corridor
46	Extension of Screenline 12 to District of Columbia line	Capture east-west flows across Rock Creek inside the Capital Beltway
47	Wheaton / Fairland	Demographic changes in this section of Montgomery County
48	Ten Mile Square NW (Arlington / Fairfax Section)	Supplement to Screenline 3; easier boundary to manage
49	Ten Mile Square NW (Montgomery / DC Section)	Supplement to Screenline 2; easier boundary to manage

Screenline Number	Location	Justification
50	Ten Mile Square NE (Montgomery, Prince George's, DC)	Supplement to Screenline 2; easier boundary to manage
51	West of MD 295	Few crossing streets
52	Ten Mile Square SE (Prince George's / DC)	Supplement to Screenline 4
53	Ten Mile Square SW (Fairfax / Alexandria / Arlington)	Supplement to Screenline 3; better capture movements within Alexandria
54	Annandale / US 50	Better capture movements to east-west travel corridor inside Beltway in Northern Virginia
55	Extension of Screenline 37	Growth in area
56	North-South Screenline for SE Loudon and NW Fairfax	Better capture travel between VA 267 and US 50 / I-66 corridors
57	Burke / Clifton	Supplement to Screenline 17; better capture travel from south to I-66 / US 50 corridor
58	2nd ring, west of I-95	Nearby transportation improvements
59	2nd ring, east of I-95	Fort Belvoir / improvements
60	I-95 north of VA 234	Nearby transportation improvements; growth in Prince William County
61	Manassas West	Nearby transportation improvements; growth in Prince William County
62	Manassas East	Nearby transportation improvements; growth in Prince William County

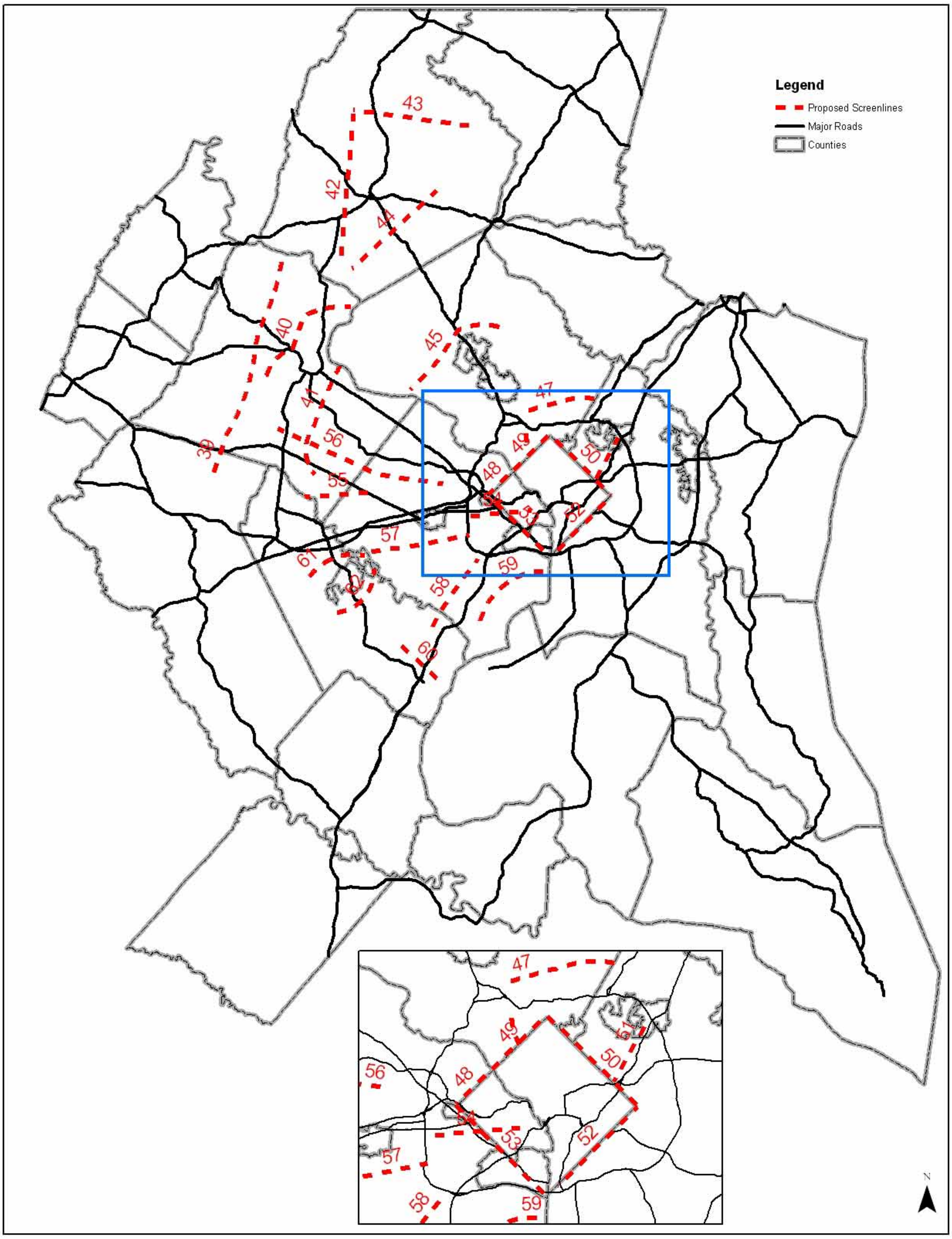

Figure 6-20: Proposed New Screenlines (Source of Base Data: TPB GIS)

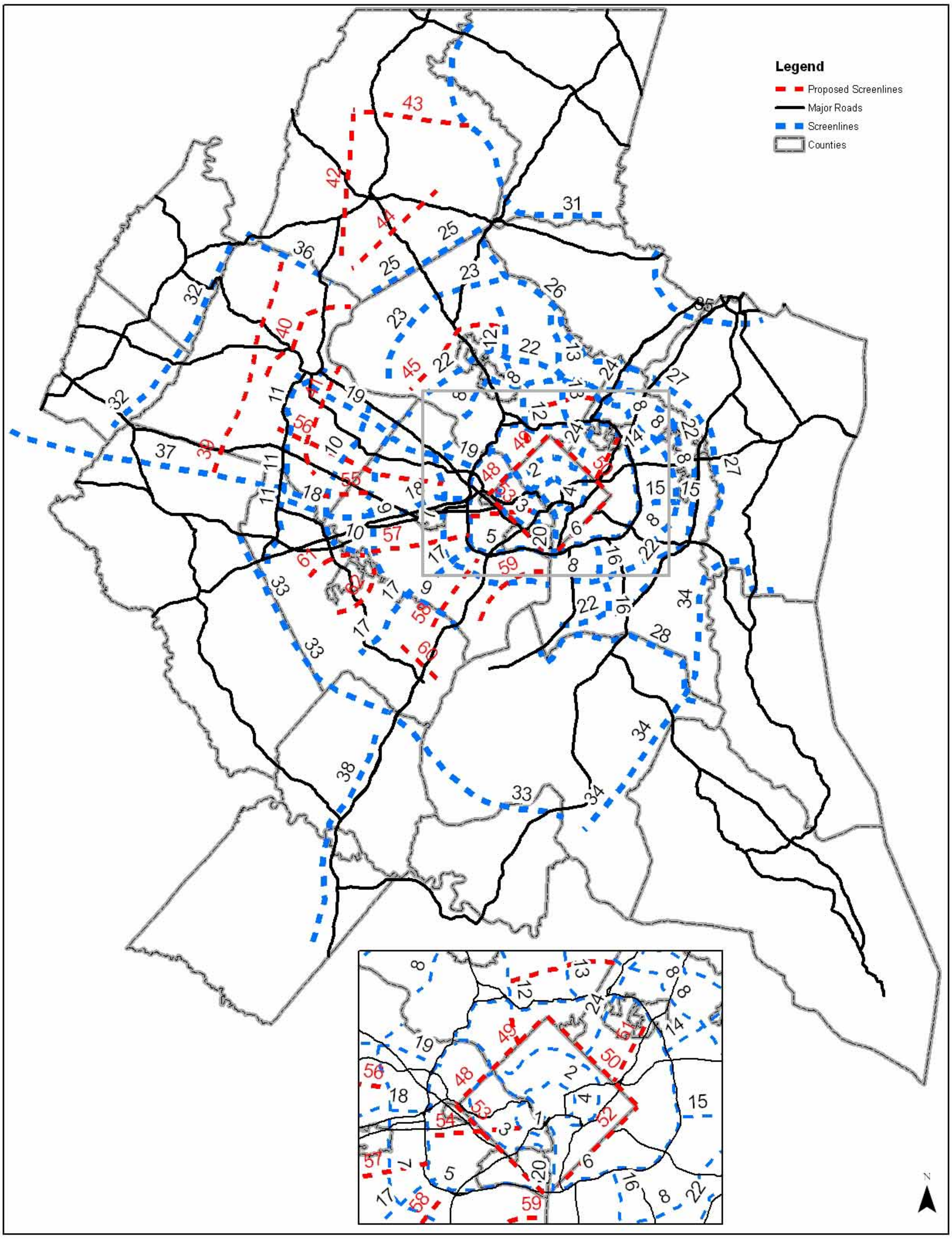

Figure 6-21: Existing TPB Screenline System with Proposed New Screenlines (Source of Base Data: TPB GIS)

Some screenlines such as 42 and 43 may appear to duplicate information obtained from the external cordon count, but by placing these screenlines TPB can compare estimated volumes with observed data collected by Maryland SHA instead of having to collect its own data. TPB staff may consider the location of some of the new proposed screenlines as guidelines for moving / adjusting the existing screenlines, but our recommendation is that the new screenlines should be added to provide the maximum number of locations for comparison of estimated and observed data during model validation. The screenlines may also be subdivided into cutlines for use in subregional validation for project planning studies.

It is also important to note that several of these new screenlines are intended to be *multimodal* screenlines. In fact, TPB should be treating all screenlines as multimodal when they traverse transit routes (both rail and bus). Certain new screenlines such as number 46 and 51 have specific transit-supportive purposes – to assist with project planning studies for the segments of the Purple Line. It may also be possible in the future to use screenlines for non-motorized travel modes, at least for project planning studies (these modes still have too small shares to really be considered during a regional validation).

Results of 2005 Model Run and Estimated / Observed Volumes Comparison Along Selected Existing and New Screenline Locations

In order to test the validity of the proposed new screenlines, a 2005 model run (using Version 2.1D #50) was completed and the screenline results compared with observed traffic data in selected locations. The observed data set contains counts from the Maryland State Highway Administration (traffic count website) and the Virginia Department of Transportation (website, NoVA traffic engineering database). Two study corridors were analyzed: the I-270 corridor in Frederick and Montgomery counties, extending to the District of Columbia line, and the I-66 corridor in Loudoun and Fairfax counties, extending to the Arlington County line. The initial results for each study corridor are reported in Table 6-3 and Table 6-4. Proposed new screenlines are shown in *italics.*

Table 6-3: Estimated vs. Observed 2005 Screenline Volumes, I-270 Corridor

Screenline / Location	Estimated Volume	Observed Volume	Deviation	Maximum Desirable Deviation
44 Southern Frederick	*119,126*	*107,450*	*11%*	*23%*
25 Montgomery / Frederick Line	115,290	121,176	5%	22%
23 Clarksburg / Northern Montgomery	26,670	36,632	27%	39%
45 Germantown	339,014	309,775	9%	14%
22 Gaithersburg (W of Screenline #12)	344,556	351,462	2%	12%
8 Rockville	303,988	342,863	11%	12%
6 Beltway Cordon	209,789	219,858	5%	17%
49 Ten-Mile Sq NW (Montgomery / DC Line btw Screenlines 46 and Potomac River Screenline)	185,222	132,475	40%	21%

Table 6-4: Estimated vs. Observed 2005 Screenline Volumes, I-66 Corridor

Screenline / Location	Estimated Volume	Observed Volume	Deviation	Maximum Desirable Deviation
11 US 15 / Eastern Loudoun	192,406	181,000	6%	19%
41 East of Leesburg	142,522	126,000	13%	22%
10 Riding	91,460	69,600	31%	29%
9 Chantilly	492,958	417,200	18%	10%
7 E of Fairfax City	473,868	494,000	4%	7%
5 Beltway Cordon	395,312	431,000	8%	9%
48/53 Ten Mile Sq NW / SW	231,714	221,600	5%	17%

The initial results suggest that some refinement along the proposed new screenlines may be necessary, such as further QA/QC of the observed data sets. In particular, time series analysis of the AADT figures to confirm the overall validity of the 2005 numbers and checks against hourly counts (where available) would be beneficial. These checks may require new or additional data collection.

Conclusions / Recommendations

TPB should consider placement of the recommended screenlines for its next model validation. Easily obtaining reliable observed data at screenline crossings is still a potential problem, so TPB staff may wish to consult with member jurisdictions to prioritize the list of new screenlines and possibly phase them into the validation tests over time. In terms of observed data, accessing the VDOT traffic engineering count database and eventually the freeway data archives for northern Virginia will provide two previously unused and robust observed data sets, but even more data are needed going forward, particularly if TPB puts an even greater emphasis on the use of smaller area screenlines. Greater segmentation of the roadway links for AADT data will be needed. There is an even greater need for improved access to observed transit data in order to make the screenline validation truly multimodal.

However, it is important to remember that validation to small cutlines compared with using regional screenlines is pulling the TPB model in two different directions, and there needs to be a balance between efforts for macroscopic and mesoscopic modeling, using the appropriate tools for each level. The creation and use of specific cutlines and subsequent validation at the beginning of a project planning study will never go away completely; there are simply too many potential study locations to be covered during a typical regional validation cycle. The need to perform screenline checks using the NCHRP 255 methodology will continue as well; TPB staff should consider expanding the sample work performed in this memo for the I-270 and I-66 corridors to the entire regional modeled area and the new screenline system as it is implemented.

Regarding the use of screenlines and cutlines during model sensitivity testing, nothing suggests that the procedures outlined in this memo cannot be applied during sensitivity testing. Placement of screenlines / cutlines does not change with sensititvity testing. However, cutlines located near the network or demographic input changes applied for the sensitivity tests will show amplified results during the test. The sensitivity tests must examine cutline volumes further upstream and downstream of the modifications in order to dampen the amplification and provide a full accounting of the model's response to the test scenario. In addition, care must be taken when conducting the higher magnitude tests (e.g., adding 10,000 or more jobs or households to a single TAZ) that the changes are not significantly altering the travel markets being captured by the cutline or introducing new or secondary travel markets that require the placement of additional cutlines to accurately check the model's response.

Finally, recall that cordons or screenlines usually cover "major" regional travel patterns, but as major destinations become more dispersed, the major travel patterns also become more dispersed, and at that point cutlines may be employed to look at particular locations and the use of local cordoned areas as employed by BMC may be necessary. BMC reports that their local area cordons are included in their regular count program, which cycles through all screenline locations over three years. There may be value in designating areas like Tysons Corner, Bethesda, and others with a local cordon.

References

Barton-Aschman Associates and Cambridge Systematics. Model Validation and Reasonableness Checking Manual. Prepared for Travel Model Improvement Program, Federal Highway Administration, February 1997.

Neil J. Pedersen and D.R. Samdahl (for JHK Associates). National Cooperative Highway Research Program Report Number 255: Highway Traffic Data for Urbanized Area Project Planning and Design. Prepared for National Research Council, Transportation Research Board, December 1982.

Email from Guy Rousseau, Atlanta Regional Commission.

Email from Erik Sabina, Denver Regional Council of Governments.

Email from Vladimir Livshits, Maricopa Association of Governments.

Email from Deng Bang Lee, Southern California Association of Governments.

Email from Lavanya Vallabhaneni, Maricopa Association of Governments.

Email from Matthew de Rouville, Baltimore Metropolitan Council, November 2007.

Baltimore Metropolitan Council. Baltimore Region Travel Demand Model Version 3.3, 2000 Validation. Task Report 07-8, January 2007. Available at www.baltometro.org/reports/ValidationV3point3.pdf.

HNTB Corporation with AECOM Consult. Central Florida Regional Planning Model (CFRPM) Version 4.0 Technical Memorandum Number 2: Model Calibration and Validation. Orlando: prepared for FDOT District 5, July 2006.

Metropolitan Transportation Commission. 2000 Base Year Validation of Travel Demand Models for the San Francisco Bay Area (BAYCAST-90), Technical Summary. Oakland, CA: MTC Planning Section, May 2004.

Southern California Association of Governments. Year 2000 Model Validation and Summary, Regional Transportation Model. Los Angeles, SCAG, April 2003.

D.J. Dailey, D. Meyers, I. Pond, and K. Guiberson. Traffic Data Acquisition and Distribution (TDAD). Prepared for Washington State Transportation Commission by University of Washington / Washington State Transportation Center, May 2002.

Telephone conversation with Larry Blain, Puget Sound Regional Council.

This page intentionally left blank

Task 7 -- Review Other MPOs' Experience with Speed Feedback and Nested Logit

This page intentionally left blank

TPB requested that VHB review the use of feedback loops coupled with nested logit mode choice models within travel demand models in use at MPOs in the United States. Using a nested logit structure for mode choice modeling has been identified as the State of the Practice, particularly among large MPOs. Some MPOs using nested logit mode choice also feed the results back into their trip distribution models. There are perceived advantages and disadvantages to this process that TPB needs to understand for its forecasting process. This memo summarizes our review; Table 7-1 summarizes the responses from individual MPOs.

Table 7-1: Results of MPO Responses

MPO	Region	Mode Choice Formulation	Feedback	Unusual Results
ARC	Atlanta	Nested Logit	Highway and Transit to Generation; Distribution; Mode Choice	None reported
CTPS	Boston	Nested Logit	Highway and Transit to Distribution and Mode Choice	None reported
NCTCOG	Dallas-Fort Worth	Nested Logit (work trips); Multinomial Logit (non-work)	Highway and Transit to Distribution and Mode Choice	None reported
Metro Council	Minneapolis/ St. Paul	Nested Logit	Highway and Transit to Distribution and Mode Choice	None reported
SANDAG	San Diego	Nested Logit	Highway and Transit to Distribution and Mode Choice	Some VMT dampening; issues upstream and downstream of HOT facility
PSRC	Seattle	Multinomial Logit with non-motorized nest	Highway and Transit to Distribution and Mode Choice (work trips) Highway only to Distribution (non-work trips)	None reported
DRCOG	Denver	Multinomial Logit	Highway and Transit to Distribution and Mode Choice	None reported; but presence of feedback loop makes finding answers “difficult” when checking strange results
MTC	San Francisco	Nested Logit	Highway and Transit to Auto Ownership; Generation; Distribution; Mode Choice (but no logsum interaction between Mode Choice and Distribution)	As much as 50% fluctuation in distribution; unusual effects from introducing multimodal improvements or large single-mode improvements

MPO	Region	Mode Choice Formulation	Feedback	Unusual Results
H-GAC	Houston	Nested Logit (being recalibrated)	Not presently; part of 2007 model improvement will examine use of feedback by other MPOs and consider implementation at H-GAC for either all purposes or just work trips	N/A
MAG	Phoenix	Nested Logit	Highway only to Land Use Model; Distribution; Mode Choice	None reported
WFRC	Salt Lake City	Nested Logit	Highway only to Distribution; Mode Choice run post-convergence	None reported
SCAG	Los Angeles	Nested Logit	Highway only to Generation; micro-loops of auto times and home-based work logsums between distribution and mode choice	None reported
DVRPC	Philadelphia	Binary Logit, nested by mode-of-approach	Full feedback	None reported

Background: Nested Logit and Feedback as State of the Practice and Current TPB Process

The recent TRB survey of MPOs identified the use of a nested logit mode choice structure as the State of the Practice in travel demand forecasting. Nearly 75% of large MPOs (peers to TPB) use a nested logit mode choice model for home-based work trips and nearly 60% use nested logit for other trip purposes.[42] MPOs (such as TPB) that provide forecasts for Federal Transit Administration (FTA) New Starts projects, highway corridor studies, or that model toll lanes are more likely to employ nested logit for mode choice than MPOs that do not engage in these three activities.[43] The use of a feedback loop was also identified as the State of the Practice. Over 80% of large MPOs feedback highway and transit travel times to trip distribution and mode choice.[44] As with the use of nested logit, feeding back travel times to distribution and mode choice is more likely among MPOs that have New Starts/Small Starts programs, that conduct corridor studies, or that model toll lanes.[45]

TPB's current production model (Version 2.1 D #50) uses a sequential multinomial logit structure for mode choice and feeds back peak period and off-peak highway and transit times to trip generation, trip distribution, and mode choice (peak period transit accessibility to jobs is fed

[42] F. Spielberg and P. Shapiro. Metropolitan Planning Organization Travel Forecasting State of the Practice. Presentation to TRB 2007 Annual Meeting. *Note: the Final Report of the TRB Committee has not been published; therefore, minor details of the data are subject to change but the overall findings are valid.*
[43] Ibid.
[44] Ibid.
[45] Ibid.

back to the demographic submodels; peak and off-peak composite times are fed back to trip distribution).[46] There are six iterations of the feedback loop executed for a typical model run.[47]

A primary component of TPB's model development program for FY 2007 is the development and implementation of a nested logit mode choice model. TPB has been working with AECOM Consult to develop the model. Previously, AECOM had developed a three-purpose, 15-mode nested logit (NL) mode choice model that has been used in project planning studies, and is applied as a post-process to the COG/TPB travel model. Building off AECOM's work, TPB staff has modified this NL mode choice model so that it has four purposes (and still 15 modes). Using the version 2.2 travel model as a base, TPB staff has replaced the existing five-mode multinomial logit (MNL) mode choice model with its four-purpose, 15-mode nested logit mode choice model and is currently working on a year 2002 calibration.

Literature Review

VHB conducted a review of major professional and academic publications and other sources for articles on the issue of feedback and its application with a nested logit mode choice model. Following the feedback requirement of ISTEA, most of the available literature covers the application of speed feedback generally and does not directly address the interaction of feedback with nested logit mode choice specifically; however, nearly all the specific models considered in the literature employ either multinomial logit or nested logit for mode choice.

Boyce *et al.* (1994) evaluate a number of feedback algorithms and provides the theoretical basis and proof of the Method of Successive Averages (MSAs) algorithm used for establishing equilibrium across the model components iterated during feedback.[48] TPB employs the MSAs algorithm in its feedback loop implementation. Boyce's algorithms were being introduced while peer review panels funded by USDOT were recommending that MPOs in Atlanta, Cincinnati, and Hartford employ feedback in their models.[49] Johnston and Ceerla (1996), working with the Sacramento regional model, found that with a full feedback loop using the MSAs algorithm a no-build solution appeared more favorable than building HOV lanes.[50] Lan (2003) reviews the feedback process implemented in the standardized model set used by MPOs in Florida and notes that "*proper calibration of the friction factor governing trip length and other important model parameters seems to bear more significance with regard to affecting the model accuracy than the feedback process itself.*"[51] More recently, Boyce (2002 and 2004) has been pushing for a shift away from the sequential modeling paradigm to an integrated model, arguing that 1) the implementation of feedback only partially overcomes the deficiencies inherent in the sequential

[46] M. Moran. Memo to VHB 7/19.2007.
[47] R. Milone, TPB Travel Forecasting Model Version 2.1D #50 User's Guide, 2004.
[48] Boyce, D.E., Y-F. Zhang, and M.R. Lupa. "Introducing 'Feedback' into Four-Step Travel Forecasting Procedure Versus Equilibrium Solution of Combined Model", Transportation Research Record 1443, TRB, National Research Council, Washington, D.C., 1994, pp. 65–74.
[49] SG Associates. Summary of Comments Prepared by Travel Forecasting Peer Review Panels, Federal Transit Administration, 1994.
[50] Johnston, R.A. and R. Ceerla. Travel Modeling with and without Feedback to Trip Distribution. UCTC Report Number 431. University of California at Berkeley, Journal of Transportation Engineering, January/February 1996, pp. 83-86.
[51] Lan, C.J. Incorporating Feedback Loop into FSUTMS for Model Consistency, Volume 1. Summary of Final Report BC-791. University of Miami (FL), September 2003.

process and 2) the widespread acceptance and use of the sequential method is largely a result of historical happenstance.[52]

A 2001 TRB report lists implementation of feedback as an improvement for air quality modeling.[53] Previous USDOT guidance notes the use of feedback loops as a best practice.[54] A recent peer review of the San Diego MPO model set noted the use of feedback; the review concluded that the model set was consistent with the state of the practice.[55] A 2003 survey of approximately 30 MPOs (including most of TPB's peers) concluded that 64% of large MPOs fed back congested times to distribution and mode split; that usage was much higher among MPOs in non-attainment areas; and that several MPOs were planning to implement a feedback loop as part of model improvements.[56] The previously noted data from the recent TRB survey shows that this figure has increased for all responding MPOs.

Forecasting work performed in 2005 to model a toll facility in the Minneapolis/St. Paul area necessitated modifying the existing feedback loop in the MPO model after initial tests with the full feedback loop yielded some unusual results: overall network travel times worsened with the implementation of toll lanes when compared with the base case.[57] Similarly, documentation from a 2004 corridor study in the Salt Lake City area discusses the effects of adding feedback to distribution and mode choice to the MPO model on the forecasting results.[58]

MPO Contacts

VHB contacted representatives of several MPOs to obtain information about their use of feedback in travel demand forecasting. MPOs were selected primarily because they are peer agencies of TPB in at least one of the following characteristics: size of modeled area; network complexity; similar regional issues; sophistication of modeling approach. Several of the MPOs contacted currently use multinomial logit rather than nested logit for mode choice, but had enough other characteristics in common with the TPB region that their information was useful. Each MPO was specifically asked if they had experienced any usual model results or outcomes, either transit or highway, with the use of their feedback loop. The discussion with each MPO is summarized below.

---

[52] Boyce, D. "Is the Sequential Travel Forecasting Paradigm Counterproductive?" Journal of Urban Planning and Development, December pp.169-182. See also Boyce, D. Forecasting Travel on Congested Urban Transportation Networks: Review and Prospects for Network Equilibrium Models. Paper for TRISTAN V: the Fifth Triennial Symposium on Transportation Analysis, Le Gosier, Guadeloupe, June 13-18, 2004.
[53] Cambridge Systematics, Inc. National Highway Cooperative Research Program Report 462: Quantifying Air-Quality and Other Benefits and Costs of Transportation Control Measures. TRB, National Research Council, Washington D.C., 2001.
[54] Barton-Aschman Associates, Inc. and Cambridge Systematics, Inc. Model Validation and Reasonableness Checking Manual. Federal Highway Administration, Travel Model Improvement Program. Section 4.1. http://tmip.fhwa.dot.gov/clearinghouse/docs/mvrcm/ch4.stm
[55] Volpe National Transportation Systems Center. Report on Findings of the Peer Review Panel of the San Diego Association of Governments Travel Demand Model. Cambridge, MA, December 2005.
[56] Walker, W.T., A White Paper on Metropolitan Planning Organization Land Use, Transportation, and Air Quality Modeling Needs in the New Federal Transportation Bill. National Association of Regional Councils in association with the Delaware Valley Regional Planning Commission, February 2003. TPB was not one of the participating agencies in this study.
[57] Cambridge Systematics, Inc. with URS Corporation. MnPass System Study, Technical Memorandum #3, Travel Demand Forecasting Approach. Prepared for Minnesota Department of Transportation, Feburary 3, 2005 (revised). The feedback loop was modified to iterate only through mode choice rather than distribution and mode choice. This resulted in fixed trip tables coming out of distribution, similar to FTA requirements for New Starts forecasts.
[58] Federal Highway Administration and U.S. Army Corps of Engineers. Draft Legacy Parkway Supplemental Environmental Impact Statement/Reevaluation and Draft Section 4(f), 6(f) Evaluation, Appendix B—2020 Travel Demand Analysis, December 2004. The model improvements incorporating the feedback loop were made between the initial draft EIS and the initial final EIS.

Atlanta (Atlanta Regional Council [ARC])[59]

ARC uses TP+/Cube as its modeling platform and has a nested logit mode choice model. In its early implementation ARC used "lots" of feedback loop iterations, and the results did not converge. To address this issue, ARC in spring 2004 implemented the MSAs algorithm for its equilibrium assignment and changed the closure criteria, and the results improved. At the same time ARC modified its feedback loop to include an additional midday highway assignment, for a total of two midday assignments within the loop. The midday assignments are performed regardless of the number of loops required for the AM assignment. In winter 2005 ARC corrected its highway skim procedure so the high-occupancy vehicle (HOV) skims for the AM and midday assignments in the feedback loop used congested times rather than uncongested times. At this time the feedback loop closure criteria were modified as well. These changes required recalibration of the HBW gravity model used for trip distribution; the model was updated using gamma functions for each income group.

ARC's feedback loop includes trip generation as well as distribution and mode choice, and both highway and transit speeds/times are part of the loop; bus speeds are generated by a separate bus speed model that was independently calibrated. A typical model run iterates five to ten feedback loops. The assignment gap tolerance is 0.001 and there are usually 30 iterations, although the number of assignment iterations varies by model year and time period. ARC did not report any issues with its current implementation. ARC considered changing their gap to 0.0001 several years ago during a model upgrade but found that the resulting increase in model run-time did not yield significant improvements to their assignment results; however, they are now again considering this change because of the performance improvements offered by Cube Cluster.

Boston (Central Transportation Planning Staff [CTPS])[60]

CTPS currently uses EMME/2 as its modeling platform but is planning to change software within the next year and is looking at both Cube Voyager and TransCAD. The current model uses a nested logit formulation for mode choice and feeds back composite impedances for all trip purposes back to distribution and mode choice. There are typically one to two feedback loops for a regular application, although more are used during calibration. CTPS did not report any significant issues.

Dallas/Fort Worth (North Central Texas Council of Governments [NCTCOG])[61]

NCTCOG uses TransCAD as its modeling platform and feeds back congested times to trip distribution and mode choice (nested logit for work trips, multinomial logit for non-work trips). Their base model run contains two feedback loops. In the initial iteration and the first feedback iteration trips are assigned for the AM peak period and off-peak. There are 30 assignment iterations for each time-of-day assignment: 15 for the first feedback loop, 15 for the second feedback loop, and 30 for the final iteration (no feedback). A base model run typically takes 13-

[59] Telephone conversation with Guy Rousseau, Atlanta Regional Commission. See also Atlanta Regional Commission, Travel Forecasting Model Set for the 20-County Atlanta Region, 2006 Draft User's Guide and The Travel Forecasting Model Set for the Atlanta Region, May 2006.
[60] Telephone conversation with Karl Quackenbush, CTPS.
[61] Telephone conversation with Arash Mirzaei, NCTCOG. See also North Central Texas Council of Governments, Dallas-Fort Worth Regional Travel Model Description, Draft, September 2006.

15 hours using parallel processing and multiple-processor hardware. Large shifts in trip distribution are indicative of problems with the feedback loop; NCTCOG did not report encountering any of these problems.

NCTCOG limits the number of feedback loops both for run-time minimization as well as for “reasonable” convergence. They are able to replicate results between model runs and have not experienced “random” results; the process is designed to encourage consistency without overstated accuracy but while maintaining a “reasonable” level of precision. NCTCOG typically performs four model runs per week, about 200 runs per year. While acknowledging that they currently need terabytes of storage for model space and completed runs, NCTCOG indicated they are likely to run more assignment iterations in the future, although not necessarily more feedback loops.

Minneapolis/ St. Paul (Metropolitan Council of the Twin Cities)[62]

The Twin Cities’ MPO uses TP+ as its modeling platform and has a nested logit mode choice model. The Metro Council model usually reaches closure (less than 2% change in AM peak period VMT between iterations) in three feedback iterations; each highway assignment step is capped at 30 iterations (using the TP+ *maxiters* parameter), although staff reports the model reaches equilibrium in 14 to 18 iterations during the most congested hours. The Metro Council uses the TP+ default convergence tests during assignment. The MPO reported no known issues with its feedback loop; see the previously cited MnPass technical memo (Cambridge Systematics [CS] 2005) for feedback loop issues associated with toll modeling.

San Diego (San Diego Association of Governments [SANDAG])[63]

SANDAG uses TransCAD as its modeling platform and has a nested logit mode choice model. The SANDAG model feeds back composite impedances to distribution and mode choice. Initially there was a single feedback loop, now multiple loops are employed. The feedback loops do not start with free-flow conditions; this reduces model run-time and is representative of the observed network – where free-flow conditions are rarely observed. One of the issues SANDAG noted was the selection of the input impedances (logsums). They indicated that the calibration process is tedious, and that the feedback loop is currently disconnected while improvements are being made to the mode choice model. It was also noted that the loop iterations drive up run-time, that they have sometimes experienced VMT dampening, and that they have sometimes noticed unusual results upstream and downstream of their HOT facility. Nonetheless, SANDAG continues the use of its feedback loop because it is consistent with the State of the Practice.

Seattle (Puget Sound Regional Council [PSRC])[64]

PSRC has a multinomial logit mode choice model with a single nest for non-motorized travel. Their feedback loop to distribution and mode choice uses composite impedances from the mode

[62] Telephone conversation and follow-up email with Mark Filipi, Metropolitan Council of the Twin Cities.
[63] Telephone conversation with Bill McFarlane, San Diego Association of Governments. See also Volpe Center (2005) report on SANDAG TMIP Peer Review and associated SANDAG staff presentation.
[64] Telephone conversation with Larry Blain, Puget Sound Regional Council.

choice logsums for work trips but feeds back auto times only for non-work trips. Their process begins with a free-flow assignment (assignment zero) for the first iteration. Iterations two through four use the full feedback loop and the MSAs algorithm for assignment before updating the skims. The fifth iteration proceeds through distribution and mode choice to the final assignment. PSRC uses a generalized cost assignment with five time-of-day segments; the number of time-of-day assignments plus the feedback loops does increase run-time—a typical model run is about 14 hours. PSRC has found that in their model trip distribution stabilizes fairly well at four feedback loops. Overall, they have found most of their forecasting results to be explainable. The PSRC model has been used by Cambridge Systematics and Mirai Associates to support toll modeling for the Washington State Department of Transportation (WSDOT).

Denver (Denver Regional Council of Governments [DRCOG])[65]

DRCOG uses TransCAD as their modeling platform and has a multinomial logit structure for mode choice. If DRCOG continued to significantly upgrade their trip-based model, they would implement a nested logit structure; however, the focus of model improvement is the development and implementation of an activity-based model. The current Compass trip-based model includes a feedback loop to distribution and mode choice. DRCOG did not report any encountering any unusual issues with their feedback loop; however, they did note that the presence of the loop can make finding "the answer" a complex search when addressing questions about forecasting results, either expected or unusual. DRCOG intends to include a feedback loop in their new activity-based model.

San Francisco (Metropolitan Transportation Commission [MTC])[66]

The MTC model uses a nested logit formulation for mode choice and feeds back auto and transit times (zone-to-zone) to their auto ownership model (which drives trip generation) as well as distribution and mode choice; however, there is no interaction between the mode choice logsum and trip distribution. MTC reported that these so-called "grand feedback loops" lead to a lot of fluctuation, and that they have seen as much as a 50 percent fluctuation in work trip distribution from the effects of the feedback loop. Introduction of multimodal or large single-mode improvements (for example, a long BART extension or a new crossing of San Francisco Bay) can create strange, difficult to explain effects. MTC urged looking at the results of feedback applications at a sub-regional level, as problems can be masked by aggregation.

Ultimately, MTC agreed that feedback is useful, but modelers need to pay a lot of attention to what is going on in the forecasting results and the decision whether or not to employ feedback is in part a consideration of whose "rules" are in place for the study (FTA, air quality agencies, etc.). MTC uses fixed trip tables when preparing forecasts for long-range planning studies to avoid bizarre results from the use of full feedback. For shorter studies, MTC sometimes uses their feedback loop, depending on the amount of linkage required with their land use model; for example, is it useful to feed year 2010 travel times to year 2015 land use allocation? They also

[65] Telephone conversation with Erik Sabina, Denver Regional Council of Governments. See also PB Consult and Gallop Corporation, The Integrated Regional Model Project, Vision Phase Final Report. Produced for Denver Regional Council of Governments, March 2005.
[66] Telephone conversation with Chuck Purvis, Metropolitan Transportation Commission.

utilize “micro” feedback loops that iterate three to five times between mode choice and assignment only.

Finally, MTC noted that a major contributor to the issues with feedback to trip distribution is the oversensitivity of the gravity model (used by nearly all MPOs for distribution) to travel time changes. Using a destination choice model rather than a gravity model for distribution may improve the performance of feedback, although k-factors and trip length correction factors may still be needed. In a destination choice model there could be empirical estimation of factors such as crossing impedances for a major travel barrier (e.g., a Potomac River coefficient). Within an existing gravity model, MTC suggested looking at attraction balancing and specifically how tightly closure is forced for non-work trips as these criteria may also contribute to wide shifts during feedback applications.

Houston (Houston-Galveston Area Council [H-GAC])[67]

Although a 2005 review of MPO models for DRCOG indicated that the H-GAC model fed back travel times to trip generation, H-GAC indicated that they currently have no feedback loop in their EMME/2 model. H-GAC is in the process of implementing significant model improvements: converting their software platform to Cube Voyager and recalibrating their nested logit mode choice model. As part of this work H-GAC and their consultants will be evaluating the use of feedback to determine if their model update should feedback all trip purposes or just work trips to distribution and mode choice. H-GAC expects this effort to be concluded by September 2007.

Phoenix (Maricopa Association of Governments [MAG])[68]

MAG uses EMME/2 for modeling and has a nested logit formulation for mode choice. MAG is considering conversion of its software platform to TransCAD. The MAG model feeds back congested speeds to a separate land use model, as well as to distribution and mode choice. Transit speeds are not fed back due to transit’s low regional mode share (less than one percent); however, there is a transit nest and corresponding impedances within mode choice. The feedback loop requires a minimum of five iterations to achieve equilibrium (root mean square error less than or equal to five percent for AM peak period trip table and link volumes). MAG reported that a typical model run executes ten iterations of their feedback loop, depending on the type of applications fewer iterations can be used. MAG also reported very long model run-times (multiple days for one run), which they are hoping to improve with their software conversion (and subsequent hardware upgrade). MAG did not report any unusual model results.

Salt Lake City (Wasatch Front Regional Council [WFRC])[69]

WFRC’s model uses nested logit (local bus, express bus, BRT, LRT, commuter rail) mode choice and feeds back highway times to distribution only. WFRC stated that their model iterates

[67] Telephone conversation with Chris Van Slyke, Houston-Galveston Area Council. See also PB Consult and Gallop for DRCOG (2005).
[68] Telephone conversation with Vladimir Livshits, Maricopa Association of Governments. See also Maricopa Association of Governments, Draft Model Documentation, 2007.
[69] Telephone conversation with GuiLin (Andy) Li, Wasatch Front Regional Council.

only between assignment and distribution to save model run-time. They run a minimum of four iterations but typically require five loops for convergence. Mode choice is run once convergence has been achieved; the final assignment follows mode choice. WFRC did not report any unusual model results.

Los Angeles (Southern California Association of Governments [SCAG])[70]

SCAG's validated year 2000 TRANPLAN model feeds back auto times to trip generation (which includes an accessibility model). There are five "grand" feedback loop iterations; there are also micro-loop feedback iterations of auto times and the HBW logsums between distribution and mode choice. SCAG uses a nested-logit mode choice model. Each assignment step runs thirty iterations. The SCAG model employs a modified MSAs algorithm (average of averages) and reaches stability within three iterations. SCAG is currently undergoing a model conversion to TransCAD, where they will retain the feedback structure but are considering Caliper's new assignment algorithm to improve performance and stability. SCAG did not report any unusual model results.

Philadelphia (Delaware Valley Regional Planning Commission [DVRPC])[71]

DVRPC's TRANPLAN model passes through the full model chain 24 times and uses the Evans algorithm to implement equilibrium assignment. The mode choice model is binary logit with nests by mode-of-access. DVRPC did not report any unusual results.

Conclusions

In their technical memo for the 2005 Minnesota toll study, CS provides an excellent summation of both the advantages and disadvantages of the use of feedback:

One of the effects of using composite impedances (or other multimodal feedback mechanisms) in trip distribution/destination choice is that when transportation improvements are made for any mode, the distribution of trips within the region is altered. This is probably more behaviorally accurate than the fixed trip table assumption, but it makes the user and systemwide benefits of the improvements more difficult to understand.[72]

The inability of sponsors to adequately explain the effects of feedback loops on their forecasting results and the desire to easily isolate project user benefits is why FTA requires the "disconnection" of feedback loops (i.e., models must use fixed trip tables) for New Starts forecasts. Nonetheless, the use of feedback can provide forecasting models with a good tool to get closer to modeling the "reality" of travel decision-making – congestion and improvements alter travelers' trip-making behavior. That is why the use of feedback is considered the State of the Practice and why its use is required by law for conformity analysis.

---

[70] Telephone conversation with Guoxing Huang, Southern California Association of Governments. See Chapter 4 for further discussion of the SCAG TransCad model.
[71] Walker, W. Thomas and Thomas Rossi. A Practioner's Guide to DVRPC's Evans Congestion-Equilibrium Travel Simulation. Presentation to 2007 TRB Planning Applications Conference, Daytona Beach, FL, and follow-up.
[72] Cambridge Systematics and URS Corporation 2005, Ibid.

Ultimately, the available literature and comments of the surveyed MPOs using or considering the use of feedback echoes the themes of the CS text – MPOs employ feedback loops in their travel forecasting tools because it is good modeling practice; yet they acknowledge some issues that can be difficult to understand or explain to decision-makers. Nothing in the literature or heard from MPOs suggests stopping or curtailing the use of feedback in conjunction with nested logit mode choice; rather, it recommends continuing feedback as part of the model chain and acknowledges the role of proper calibration and validation as echoed by the 2003 FSUTMS report previously cited. Rigorous examination of the results of each calibration run should identify any unusual model results that necessitate correction before validation and acceptance with a determined optimal number of feedback loop iterations. When examining the results of the feedback process careful attention should be paid to localized impacts as well as regional impacts to assure that unexplainable results in one area are not masked by regional impacts. Further research may be needed to determine an optimal number of feedback iterations to be used with the TPB model.