Application of Dynamic Traffic Assignment to Advanced Managed Lane Modeling

Final Report

Contract BDK80-977-30

Prepared for

Florida Department of Transportation

By

Florida International University

Lehman Center for Transportation Research

URS Corporation

CITILABS

November 2013

Disclaimer

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the State of Florida Department of Transportation.

Metric Conversion Chart

APPROXIMATE CONVERSIONS TO SI UNITS

SYMBOL	WHEN YOU KNOW	MULTIPLY BY	TO FIND	SYMBOL
LENGTH				
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
AREA				
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
VOLUME				
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: volumes greater than 1000 L shall be shown in m^3				
MASS				
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
TEMPERATURE (exact degrees)				
^oF	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^oC
ILLUMINATION				
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
FORCE and PRESSURE or STRESS				
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Application of Dynamic Traffic Assignment to Advanced Managed Lane Modeling		5. Report Date November 2013 6. Performing Organization Code
7. Author(s) Mohammed Hadi, Shaghayegh Shabanian, Halit Ozen, Yan Xiao, Michael Doherty, Cesar Segovia, and Heejoo Ham		8. Performing Organization Report No.
9. Performing Organization Name and Address Lehman Center for Transportation Research Florida International University 10555 W. Flagler Street, EC 3680, Miami, FL 33174		10. Work Unit No. (TRAIS) 11. Contract or Grant No. BDK80-977-30
12. Sponsoring Agency Name and Address Office of Research and Development State of Florida Department of Transportation 605 Suwannee Street, MS 30, Tallahassee, FL 32399-0450		13. Type of Report and Period Covered Draft Final Report June 2011 – December 2013 14. Sponsoring Agency Code
15. Supplementary Notes Mr. Vladimir Majano of the Systems Planning Office at the Florida Department of Transportation served as the project manager for this project.		

16. Abstract

In this study, a demand estimation framework is developed for assessing the managed lane (ML) strategies by utilizing dynamic traffic assignment (DTA) modeling, instead of the traditional approaches that are based on the static traffic assignment (STA). The framework includes methods for calibrating the network (supply), demand, and assignment modeling parameters. The methods were extensively tested using real-world data to assure that the DTA modeling of managed lanes reflected real-world conditions and responded reasonably to different congestion management scenarios, such as variable pricing policies and different levels of willingness to pay. This framework is standalone and independent of the utilized DTA modeling tool.

17. Key Word	18. Distribution Statement		
dynamic traffic assignment, simulation models, advanced assignment methods, managed lanes	Unrestricted		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 186	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Executive Summary

Background

Transportation agencies have realized that in many cases, adding lanes is no longer a feasible or effective solution to cope with ever-increasing traffic congestion problems. Rather, these agencies have applied advanced strategies and tools for optimal utilization of the existing capacity in time and space. Demand management, access management, active traffic management, incident management, smart work zone applications, and advanced traveler information systems are examples of the types of strategies that are designed to get the most out of the existing physical capacity. Advanced technologies are needed to implement these strategies for the constant monitoring of traffic conditions, effective analysis of the traffic data in offline and online applications (for planning and operation), and active response to different traffic situations. Managed lanes (ML) are increasingly being considered one of the most promising strategies to address transportation system problems.

Intelligent transportation systems (ITS) have provided a solid platform for deploying the abovementioned strategies. The advances in ITS technologies and strategies have made collecting and archiving traffic data more efficient and affordable. This data can be used to closely monitor and analyze traffic conditions, in both real-time and offline applications, as well as to correspondingly plan, operate, and manage the facility.

ML has evolved based on the notion of "actively" operating freeway facilities. A managed lane is a lane (lanes) within an existing freeway that can be dynamically managed to constantly meet preset criteria, such as acceptable levels of service or minimum speeds. Advanced applications of managed lanes involve traffic management centers (TMCs) dynamically adjusting their operation parameters via controlling access to the ML, changing eligibility of vehicle occupancy, and varying the toll values to regulate the demands and keep the facility in optimal operational condition. The definition of "optimal condition" has varied over time and between different agencies. Examples of criteria that dictate the operation policy of ML include keeping a certain level of service on the ML, congestion mitigation on the parallel general purpose lane (GPL), revenue maximization, pollution reduction, improving trip reliability, encouraging transit use, and improving safety.

There are high transportation demands in Florida's urban areas due to heavy commuting traffic, in addition to tourist trips and freight transportation, resulting in severe congestion of the transportation network. The ML was first implemented in Florida on a congested segment of I-95 in Miami, Florida. The dynamic toll values are set depending on the congestion levels in the ML to maintain a guaranteed reliable trip for the ML users. Since this implementation has been very successful and supported by the public, more ML implementations are being studied or developed in Florida. In South Florida, the ultimate goal is to connect these facilities within a seamless regional network.

Dynamic Traffic Assignment (DTA) Application in Managed Lane Modeling

Effective planning and implementation of ML strategies require the utilization of advanced modeling methods to allow for more accurate assessments of the impacts of changes in associated traffic flow conditions and operations. Macroscopic, mesoscopic, and microscopic analyses and simulation have been used in assessing ML strategies. Mesoscopic simulation modeling has been proposed as a level of modeling detail between macroscopic and microscopic modeling since microscopic simulation is expensive to apply and calibrate, and macroscopic analysis is not capable of capturing the dynamics of traffic flow, particularly under congested conditions with breakdown and queue spillback effects. Dynamic traffic assignment (DTA), combined with mesoscopic simulation and in some cases, microscopic simulation, has been increasingly utilized to evaluate traffic management strategies. Compared to traditional methods that normally utilize static traffic assignment (STA) and simple analytical traffic flow equations, simulation-based DTA better captures the dynamics of traffic operations by modeling time-variant system measures (including queuing and travel times), demand, advanced management strategies, and the associated responses of travelers.

After almost thirty years of research and practice, DTA is maturing and has been used for demand forecasting and performance assessment. However, there is still a need to understand its capabilities and limitations, particularly when implemented to assess advanced strategies such as ML. The challenges that DTA users face include choosing the right DTA tool based on specific needs, understanding the time and data requirements for modeling and calibration, network and demand size limitations, and training needs. Of particular interest to this research is the need for the development and assessment of methods to utilize DTA in managed lane modeling.

The Florida Department of Transportation (FDOT) is developing a standard approach for managed lane demand forecasting applications in the Florida Standard Urban Transportation Model Structure (FSUTMS) as a three-phase project. Each phase will produce a toolbox for managed lane applications that varies in the level of resolution and sophistication between the different phases. Phase I addresses the use of static assignment in managed lane modeling and analysis. In Phase II, the choice between GPL and ML is formulated via a logit model in the mode choice step of the traditional four-step demand forecasting procedure. Phase III will take this approach to the next level and integrate the framework with activity-based models (ABMs) and more disaggregated demand data, which in turn, will generate more detailed studies. The combination of DTA and ABM will also be explored in this phase. This research developed methods to support the development, calibration, assessment, and use of DTA in modeling managed lanes and subsequently took advantage of the presence of the detailed traffic data from ITS implementations. It provides a critical input to the Phase III effort in developing the methodologies required to use DTA in managed lane modeling.

Goals and Objectives

The goal of this project was to develop and assess procedures for ML modeling utilizing simulation-based DTA to meet modeling requirements that cannot be achieved when using procedures that are based on static assignment. The specific objectives of this research were as follows:

1. Develop procedures for utilizing detailed data from multiple sources to calibrate the supply and demand sides of DTA models for use in modeling ML.
2. Develop procedures for the use of simulation-based DTA to model ML.
3. Assess the performance of the use of DTA compared to STA in ML modeling.

4. Test the ability of a DTA tool (Cube Avenue from Citilabs, Inc.), which is a strong candidate for use in Florida to model ML.
5. Demonstrate the application of the developed procedures to a ML implementation in Florida.

The research team worked closely with the Cube Avenue platform developers to improve the prototype and increase its reliability. The contribution of the research included developing procedures for data acquisition and validation, optimally utilizing ITS data and setting performance measures to improve supply and demand calibration, developing a procedure for demand estimation, extensively testing the performance of the proposed methods, and making recommendations to improve managed lane modeling based on the developed procedures.

Methodology Overview

Evaluation of DTA implementation and its ability to replicate travelers' behaviors in the real world required that the network and demand be carefully calibrated. Demand data could be obtained by integrating data from several sources. The approach used for calibrating the network and demand was an iterative process. This study also examined the benefits of replacing STA with DTA. A prototype of DTA implementation for ML applications was developed, tested, and evaluated versus real-world data, and a comprehensive sensitivity analysis was performed to evaluate the capability of the model in demand forecasting and in route choice replication.

Data

Data acquisition and preprocessing is a crucial step required for DTA model development, validation, and calibration. In this study, there was a unique opportunity to access GPL and ML data, particularly from the ITS deployment. To support the development, calibration, and evaluation of the models, the methodology was tested on a corridor in Miami, Florida, with managed lane in operation for about four years, detailed traffic measurements, and available toll data. It should be mentioned, however, that if the data is not carefully validated and preprocessed, its use may adversely affect the modeling and calibration. Removing non-representative days and time intervals, as well as detector erroneous data, and checking consistency in time and space are examples of processes developed in this research to address this concern.

Supply Calibration

Supply calibration, sometimes referred to as network calibration, as applied to the mesoscopic simulation model used in this study, included the estimation of capacities, free-flow speeds, and traffic flow model (TFM) parameters. The process used in this study was to conduct calibration utilizing a systematic sequential manner with an increasing spatial scope of the calibrated network. First, bottleneck spots were isolated and calibrated separately to make sure that the breakdown in the real world can be replicated at the right time and space. In the next step, corridor segments containing connected bottlenecks were calibrated. Finally, the estimated parameters were fine-tuned for the whole area. It was found that the estimated capacity based on detector data is noticeably lower than those suggested by the Highway Capacity Manual (HCM), as well as those coded in the regional demand forecasting model. The calibration of this capacity was found to have a significant impact on the study's results.

Demand Estimation

Demand estimation aims to estimate the origin-destination (OD) table for each 15-minute interval of the modeling period. This process was accomplished based on an initial matrix obtained from a planning level demand forecasting model, with a model period of three hours and a set of traffic counts obtained from detected data. This part itself was broken into three steps, with an ascending level of details and data requirements. The first step was a simple split of the initial matrix using "time-slicing factors" that distribute the three-hour matrix during 15-minute intervals based on variations observed in the detected link counts (and/or home survey results). These matrices were used as inputs to the next step, which is a least-square optimization method combined with a static traffic assignment model. In this step, the ODs were estimated in such a way that when loaded onto the network by the assignment module, they produce link volumes that are similar to the counts captured by the traffic detectors, with limited deviations from the initial trip table.

Since a static assignment runs over a single model period, each 15-minute interval must be run separately to estimate the OD matrix for the associated interval. The most important concern with using the STA in this process is its inability to capture queue spillback in space and time. In the current study, this problem could only be partially addressed by utilizing heuristics to account for queue presence.

In order to overcome the abovementioned limitations, the best approach is to utilize DTA instead of STA as part of the least-square optimization to better account for traffic dynamics and travelers' behaviors. However, limitations were identified with an existing tool developed for this purpose, and modifications were proposed to improve the performance of this approach. The demand calibration process presented above was performed in an iterative manner with the demand calibration steps mentioned earlier.

Assignment

The ability of STA and DTA to replicate the behavior of travelers in using ML was evaluated. Different strategies for modeling the route choice behaviors were investigated. Sensitivity analysis was performed to assess different approaches for route choice and assignment processes. It was found that DTA could better replicate the observed data, while STA tends to underestimate the demands and the portion of travelers that are willing to use managed lanes.

Findings

Simulation-based dynamic traffic assignment (DTA) has been increasingly utilized to evaluate traffic management strategies, including managed lane (ML). Compared to traditional methods that normally utilize static traffic assignment (STA) and simple analytical traffic flow equations, simulation-based DTA better captures the dynamics of traffic operations by modeling time variant system measures (including queuing and travel times), demands, advanced management strategies, and the associated responses of travelers. The following can be stated regarding DTA applications in general and their use in ML modeling in particular, based on the findings from this study.

- A variety of modeling approaches have been proposed to assess managed lane implementations. These approaches range from high-level sketch planning tools to micro-level modeling of individuals' behaviors and traffic operations. Simulation-based DTA should be considered an effective modeling approach to support the planning and operation of ML.
- The modeling network can be extracted as a subarea network from the regional models. For this purpose, the subarea boundary can be specified using the Cube Polygon feature or a GIS tool. However, it was found that the modeled network geometry needs to be updated to better represent the existing real-world network geometry, since the details and accuracy of modeling the network in demand forecasting models are not sufficient for DTA applications.
- Advanced modeling tools such as DTA required more detailed and higher quality data to ensure that the developed model accurately replicates real-world conditions. This study successfully and extensively utilized detector data collected from an existing ITS system operated by the regional traffic management center, combined with portable traffic monitoring sites (PTMS) ramp counts and measurements from other sources of data to satisfy the DTA data needs. However, significant efforts were needed to process, fuse, and validate the data to allow for use in the modeling processes.
- Managed lane modeling was successfully implemented and tested in Cube Avenue using two approaches. The first approach involved adding the equivalent value of time of the toll cost value to the travel time function within the assignment, resulting in a generalized cost that considered the ML toll. In this approach, referred to as the "generalized cost function" approach in this study, vehicle use of ML was solely governed by the user equilibrium (UE) assignment procedure, based on the generalized costs of the competing paths. The second approach was referred to as the "willingness-to-pay curve" approach. In this method, prior to the assignment, travelers were divided into two groups: a group that will choose not to pay the toll and is limited to using the general purpose lane (GPL),

and a group that is willing to pay and use the ML. The latter group is eligible to use the ML lanes based on willingness to pay, but the final decision to use a managed lane depends on its origin and destination and on the difference in the generalized costs between ML and alternative routes according to the UE process.

- In this research, several issues were identified that limited Cube Avenue's modeling abilities, in particular, as it relates to ML modeling using the willingness-to-pay curve approach. Citilabs, developer of Cube Avenue, participated in this research and addressed the identified issues and updated Cube Avenue during the course of the project to reflect the project's findings. Although these problems were solved and reasonable results were produced with the final version of the model, it is recommended that the user examine the results and report any issue to the program developer.
- A sequential procedure that iterates among network (supply) calibration, demand estimation, and route choice parameter estimation was recommended in this study. Despite the existence of mathematical formulations and solutions for simultaneous supply and demand estimation, their implementations in the real world were not straightforward and were not executed.
- Supply or network calibration in Cube Avenue entails estimating capacity, free-flow, and traffic flow model parameters for each link in the network. These parameters affect the travel time, congestion time, queue formation, and queue spillback when the demand was loaded. A systematic multilevel approach to network (supply) calibration was recommended in this study, with an increasing calibration scope in each level. The process started at the level of separated bottlenecks where the capacity was estimated by various methods based on field data. The network was gradually extended to connect the bottlenecks, and then to the whole corridor and subarea coverage.
- The supply calibration performed in this study illustrated the importance of coding capacity based on detector measurements in DTA tools, particularly when there was evidence that the modeled corridor capacity was lower than the HCM-based estimates. In the case explored in this study, it was found that the free-flow speed and more importantly, the capacity, were overestimated by the HCM procedures, resulting in incorrect travel times and congestion when used in the DTA model.

- One of the important congestion spots in the modeled network was caused by a spillback from an off-ramp that caused low speeds in the two left lanes (the I-95 Northbound off-ramp to the Turnpike). Since the utilized DTA tool (Cube Avenue) does not support lane-by-lane modeling, it is not possible to correctly replicate that location, because the queue in the model first fills up the whole segment (including 5 lanes) before backing up to the upstream link. In the real-world, only the two left lanes are blocked. If replicating the congestion at such locations is important to a study, a tool that better handles this situation or multi-resolution analysis should be considered.
- Dynamic traffic assignment requires trip matrices specified for short time intervals (e.g., 15 minutes or 30 minutes). The derivation of these matrices was performed in this study using a sequential process that started from matrix factorization based on count data, followed by static assignment-based OD matrix estimation, and finally followed by dynamic assignment-based OD matrix estimation. However, identified limitations, tool immaturity, and the results of this study indicated that at the current stage, the dynamic OD estimation process in Cube Analyst should be used with caution until further enhancements and testing of these enhancements are completed so as to confirm that the tool is able to produce good results.
- During the matrix estimation process with the currently available tools, several manual adjustments and iterations were required to ensure joint calibration of demand, supply, and route choice behaviors. Adjustments and fine-tunings were also needed to avoid unrealistic deviation from the initial matrix and trip pattern.
- Important specific enhancements to the OD estimation process in Cube Avenue are recommended, as listed in Chapter 6.
- When calibrating supply, demand, and assignment parameters, a distance function between simulation outputs and field measurements is minimized. This function can include different measures, such as link volumes, OD demands, link speeds and/or densities, etc. Limiting the function to replicating link volumes, as is the case in many studies, can be misleading and fail to produce the correct demands or congestion patterns. Most OD matrix estimation methods are based on link traffic volumes and initial OD matrices. If enough data on speeds, densities, queue lengths, OD routes or zonal trip end

rates are available, they should be incorporated into the calibration process to better replicate real-world traffic conditions.

- Calibrating the toll curve, value of time, and willingness-to-pay curve parameters are important aspects of DTA utilization for ML modeling.
- There was evidence that the value of time used in the SERPM model ($12.60 per hour) is low. The value identified in a previously conducted study of the I-95 Express ($22.00) produced better results.
- There was evidence that motorists perceived additional benefits of ML not accounted for by the raw value of travel time. Thus, a factor was used in this study, as a multiplier of the saved travel time, to improve the results by magnifying the saved travel time, originally obtained from skimming toll and toll-free routes. This was to account for other factors initially not accounted for, such as reliability, comfort, and safety.
- The findings from this study highlighted the shortcomings of utilizing static assignment for assessing managed lanes, even when the measured capacity values and correct volumes were coded, illustrating the need to utilize DTA modeling for such assessments. The calibrated DTA model was able to produce results that are similar to real-world results. However, the Cube Voyager static assignment module was not able to replicate real-world data.
- For the case study of this project, it was found that the generalized cost approach and the willingness-to-pay approach produce comparable results, although the generalized cost approach is much simpler to implement, calibrate, and converge.
- For the case study of this project, the generalized cost approach appears to be able to achieve converged and stable solutions. However, the willingness-to-pay approach was not able to achieve converged and stable solutions.
- Initial attempts were made to expand the network to a larger subarea, compared to the linear model used in this study. It was found that a linear facility with small-size trip matrices can work well and are easier to replicate than larger networks. In addition, the original release of Analyst Drive (dynamic OD estimation) did not work for the larger area.
- Cube Voyager, Cube Analyst, and Cube Avenue require separate licenses if they will be used either alone or together in a model application. This project required all of these

applications to run together, which is quite new in Florida's modeling practices, hence, the emergence of licensing conflicts and the need for a solution to this problem.

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1 Background Statement

Transportation agencies have realized that in many cases, adding lanes is no longer a feasible or effective solution to cope with ever-increasing traffic congestion problems. Rather, these agencies have applied advanced strategies and tools for optimal utilization of the existing capacity in time and space. Demand management, access management, active traffic management, incident management, smart work zone applications, and advanced traveler information systems are examples of the types of strategies that are designed to get the most out of the existing physical capacity. Advanced technologies are needed to implement these strategies for the constant monitoring of traffic conditions, effective analysis of the traffic data in offline and online applications (for planning and operation), and active response to different traffic situations. Managed lanes (ML) are increasingly being considered one of the most promising strategies to address transportation system problems.

Intelligent transportation systems (ITS) have provided a solid platform for deploying the abovementioned strategies. The advances in ITS technologies and strategies have made collecting and archiving traffic data more efficient and affordable. This data can be used to closely monitor and analyze traffic conditions, in both real-time and offline applications, as well as to correspondingly plan, operate, and manage the facility.

ML has evolved based on the notion of actively operating freeway facilities. A managed lane is a lane (lanes) within an existing freeway that can be dynamically managed to constantly meet preset criteria, such as acceptable levels of service or minimum speeds. Advanced applications of managed lanes involve traffic management centers (TMCs) dynamically adjusting their operation parameters via controlling access to the ML, changing eligibility of vehicle occupancy, and varying the toll values to regulate the demands and keep the facility in optimal operational condition. The definition of optimal condition has varied over time and between different agencies. Examples of criteria that dictate the operation policy of ML include keeping a certain level of service on the ML, congestion mitigation on the parallel general purpose lane (GPL), revenue maximization, pollution reduction, improving trip reliability, encouraging transit usage, and improving safety.

There is high transportation demand in Florida's urban areas due to heavy commuting traffic, in addition to tourist trips and freight transportation, resulting in severe congestion of the transportation network. The managed lane (ML) was first implemented in Florida on a congested segment of I-95 in Miami, Florida as a countermeasure for optimal utilization of the facility. The toll values are set depending on the congestion levels in the ML to maintain a guaranteed reliable trip for ML users. Since this implementation has been very successful and supported by the public, more ML implementations are being studied or developed in Florida. In South Florida, the ultimate goal is to connect these facilities within a seamless regional network. MLs are also being considered for implementation at other locations in Florida and around the nation.

Effective planning and implementation of ML strategies require the utilization of advanced modeling methods to allow for a more accurate assessment of the impacts of changes in traffic flow conditions and the impact of operation strategies. Macroscopic, mesoscopic, and microscopic analyses and simulation have been used in assessing managed lane strategies. Mesoscopic simulation modeling has been proposed as a level of modeling detail between macroscopic and microscopic modeling since microscopic simulation is expensive to apply and calibrate, and macroscopic analysis is not capable of capturing the dynamics of traffic flow, particularly under congested conditions with breakdown and queue spillback effects. DTA, combined with mesoscopic simulation and in some cases, microscopic simulation, has been increasingly utilized to evaluate traffic management strategies. Compared to the traditional methods that normally utilize static traffic assignment and simple analytical traffic flow equations, simulation-based DTA better captures the dynamics of system operations by modeling time variant system measures (including queuing and travel times), demands, advanced management strategies, and the associated responses of travelers.

After almost thirty years of research and practice, DTA is maturing and has been used in demand forecasting and performance assessment. However, there is still a need to understand its capabilities and limitations. Challenges that DTA users face include choosing the right DTA tool based on specific needs, understanding the time and data requirements for modeling and calibration, network and demand size limitation, and defining proper performance measures. Of particular interest to this research is the need for the development and assessment of methods to utilize DTA in managed lane modeling.

The current research develops methodologies to support the development, calibration, assessment, and use of DTA in modeling managed lanes, and subsequently takes advantage of the presence of the detailed traffic data from intelligent transportation system (ITS) implementations.

The remaining sections of this chapter provide a basic understanding of managed lane strategies and modeling approaches. DTA is briefly introduced, and the benefits and limitations of its utilization in managed lane modeling are discussed. Goals and objectives are clearly identified, and the structure of the document to address these objectives is outlined.

1.2 Managed Lane

With the escalated challenges of congestion and constraints in building new roads, such as construction costs and right-of-way limitations, transportation agencies are increasingly implementing advanced operational strategies to maximize the performance of the existing infrastructure. The implementation of ML has been accepted as a successful strategy that regulates demand, decreases turbulence by separating traffic streams, and utilizes unused capacity in time and space. Managed lanes can be defined as "freeways within freeways" that are proactively operated via pricing, access management, and occupancy restrictions to better utilize the existing capacity (FHWA, 2008c).

The key feature that distinguishes ML from traditional capacity improvements is the operational flexibility to actively respond to the current situation, and continuously keep the facility in optimal condition. The optimization criteria can include the following: Preserving a certain level of service in ML, maximizing the revenue, supporting environmentally-friendly vehicles, improving trip reliability, improving safety, and encouraging the use of public transit. (FHWA, 2008c). Although the objective of the strategies utilized in existing managed lane applications is mainly to maintain an acceptable level of service of the priced lanes, studies show that travelers in general purpose lanes also benefit from managed lane deployments. (Safirova et al., 2003; and Janson and Levinson, 2013)

Combinations of access control, pricing, and vehicle eligibility define different types of ML. High occupancy toll lanes (HOT lanes) and high occupancy vehicle lanes (HOV lanes) are the most commonly deployed types of ML across the country. HOV was the preliminary form of ML that was first implemented in the northern state of Virginia in the late 1960s. During rush hour, HOV lanes can only be used by transit or eligible high occupancy vehicles (a minimum of two or three passengers). In many congested urban areas, however, HOV lanes no longer function as intended. Either they become as congested as GPL lanes, or there is an unused capacity available when GPL lanes suffer from severe traffic jams. HOT lanes take advantage of the excess capacity on ML by allowing non-eligible vehicles, such as Single Occupancy Vehicles (SOVs), to use it by paying a toll. The benefits of ML may not be limited to providing a reliable mobility option for ML users, but also include moderating the spreading of demands and peaks due to the implemented pricing strategies.

In addition to HOV and HOT lanes, other forms of managed lanes include express lanes, truck-only-lanes, bus-only-lanes, reversible lanes, and ramp metering, coupled with priority access. Very few truck-only and bus-only lanes are under operation and construction in the U.S.

Successful implementations of ML in California, Florida, Virginia, Texas, Minnesota, Washington, and New York have formed a basis for spreading the interest in management lanes across the country. More ML projects are under development in Georgia, Oregon, Maryland, North Carolina, and other states around the country.

Managed lanes are separated from general purpose lanes by different buffer types to reduce the friction between managed lane traffic and the adjacent general purpose lanes. This separation can also control the access to the managed lanes. The buffer can simply be painted markers on the pavement, a removable plastic barrier, or a fixed concrete barrier. The degree of separation between managed lanes and general purpose lanes affects the ML performance, as discussed in Section 4-2.

To allow the payment of tolls, most existing ML facilities are equipped with Electronic Toll Collection Systems (ETC) to provide an uninterrupted flow of ML users. ML policies greatly vary in different regions and over time, and should be tailored to local traffic conditions. In basic applications, the toll values are fixed. In some cases, however, it varies by time of day, and in more advanced models, it is adjusted dynamically based on the congestion levels in time intervals as short as three minutes. In an advanced, flexible operation policy, different toll values may be applied to different user groups based on different criteria. Most current applications of managed lanes use dynamic pricing policies.

An occurring nationwide trend is to convert under-performing HOV lanes to HOT lanes to better utilize capacity in time and space. Another trend connects ML facilities within a regional network of ML. A good example for both trends is the ML implementation in South Florida that is extending the existing 8-mile HOT lanes on I-95 to a 21-mile facility that connects I-395 in Miami-Dade County to I-595 in Broward County. This ML facility will be connected to other priced roads in the region within a seamless and interconnected regional network.

1.3 Managed Lane Modeling Structure

A robust and reliable ML model is necessary for the planning, design, and operation of ML. Such a model should be able to replicate the real-world measures of traffic for current conditions and reasonably forecast changes in demand, congestion patterns, performance measures, and revenues in response to changes in operation policies, strategies, and traffic growth. An ultimate model should have a sound structure and well-calibrated parameters, and should be sensitive to pricing policies at different levels, including trip generation, vehicle occupancy, destination and departure time choice, mode choice, and route choice.

Most of the existing modeling efforts of managed lanes are designed within the framework of the traditional four-step demand forecasting model, although this is sometimes followed by microscopic traffic simulation of the resulting demands with no further reassignment of traffic. These efforts usually include modifications in the assignment step to account for the binary route (toll vs. no-toll) choice. Some models consider the effects of ML implementation on the mode choice or trip distribution. Although such modeling may be adequate for uncongested freeways with limited variations in demands during peak periods, it is generally not sufficient for evaluating ML alternatives in congested urban areas where most ML applications are expected.

In frameworks that explicitly model the binary route choice as part of the assignment process (whether STA or DTA), it is crucial to consider several user classes to capture the heterogeneity of travelers that stem from differences in trip purpose, income level, and vehicle occupancy (NCHRP, 2012). If the effect of ML operation is considered in other steps, such as mode choice or trip distribution, the consistency between these steps, proper feedback structure, and achieving system equilibrium (i.e., a global convergence for the whole demand forecasting model) is crucial (NCHRP, 2012).

The Florida Department of Transportation (FDOT) is developing a standard approach for managed lane demand forecasting applications in the Florida Standard Urban Transportation Model Structure (FSUTMS) as a three-phase project. Each phase will produce a toolbox for managed lane applications that varies in the level of resolution and sophistication between the different phases, as described below.

Phase I focuses on the use of static assignment in developing route choice in managed lane modeling and analysis. In the absence of DTA, variable demands and a dynamic toll that reflects real-world implementations are modeled by dividing the model period into 15-minute intervals, and running the modeling process separately for each interval. The input demand for the entire peak period is obtained from sub-area extractions from regional demand forecasting models. The model developed in this phase does not change the split between different user groups, including SOV, HOV, and transit users, as estimated by the regional demand forecasting models. The developed model can be used to determine the proportion of drivers willing to use ML, given the charged tolls and the difference in performance between ML and GPL lanes, based on a static assignment procedure combined with a willingness-to-pay curve (Ruegg et al., 2013).

In Phase II (Parsons Brinckerhoff, 2013), the choice between GPL and ML is formulated via a logit model in the mode choice step of the traditional four-step demand forecasting procedure. A trip distribution step is then executed, and the resulting origin-destination (OD) table is used as input into a static assignment module. The toll and percentage of drivers willing to use ML are updated by the assignment model and fed back into the mode choice model. The iterative process is run until a desired level of system convergence is achieved. The Phase II model is able to estimate the impacts of managed lane on mode choice, (split between SOV, HOV, and transit), in addition to the proportion of travelers using ML.

Phase III will take this approach to the next level and integrate the framework with activity-based models (ABM) and more disaggregated demand data, allowing for more detailed studies. The combination of DTA and ABM will also be explored in this phase. This research provides critical input into the Phase III effort in developing the methodologies required to use DTA in managed lane modeling.

The growing trend in ML modeling is to apply ABM to better capture the travelers' behaviors, based on disaggregated demand data. As the name suggests, activity-based models can reflect the effect of ML operation in different levels of individuals' activities and decisions. Capturing changes in vehicle occupancy, destination, and departure time choice are more feasible in ABM models, compared to the conventional trip-based models. More advanced prototypes implement DTA instead of STA, either in trip-based or activity-based frameworks. Integration of ABM and DTA in a coherent framework is the ultimate goal in ML modeling. The next section identifies the benefits of applying DTA in ML analysis and forecast.

1.4 DTA Implementation in Managed Lane Modeling

With the restrictions on adding lanes to highways, transportation agencies are recognizing the need to operate their existing infrastructure more efficiently and intelligently by implementing advanced strategies that can eliminate unnecessary trips, shift trips out of congested routes and peak hour periods, and encourage transit trips and carpooling. Ramp metering, Advanced Travel Information System (ATIS), Advanced Traffic Management Systems (ATMS), Integrated Corridor Management (ICM), incident management, and smart work zone applications are examples of advanced operational strategies that optimize the utilization of the transportation supply. Congestion pricing strategies such as ML can be categorized as an advanced traffic and demand management (ATDM) strategy that regulates demand, along with improving the operation of the utilized road. In many cases, ATDM can also generate enough revenue to operate the highway and expand the development of transportation systems.

The impacts of advanced strategies such as ML are particularly significant when the facility is operating near its capacity. Applying these strategies is time-dependent and highly sensitive to small changes in traffic and/or demand. Therefore, these applications require more advanced and detailed modeling frameworks, compared to the approaches used in traditional demand forecasting.

The use of simulation-based DTA was proposed as an alternative to provide more realistic and detailed analyses of ML. Simulation-based DTA tools utilize mesoscopic or microscopic simulation to assess traffic performance after each iteration assignment. Mesoscopic models generate and track individual vehicles, as is the case in microscopic simulation, however, the interaction between them must be modeled through a macroscopic traffic flow model (TFM), rather than a microscopic traffic modeling. Microscopic simulations are useful tools for traffic analysis. However, they are extremely demanding in the data and time needed for the correct modeling and sound calibration of traffic flow. Such models are not appropriate for regional networks. Macroscopic simulations, on the other hand, are too aggregated for operational analysis purposes, and many are unable to capture vital features of congested networks like bottlenecks.

DTA is a modeling approach that captures the dynamic interaction between demand and network, and advanced strategies and associated parameters. It models the period demand over short-time intervals, with a traffic assignment in each interval, which is affected by the network condition resulting from the previous interval assignment. This means that for each OD pair, vehicles that depart in different time intervals can use different paths and may experience different travel times. The core engine that assigns the demand to eligible routes in most static and dynamic assignment tools stems from the user equilibrium (UE) concept. Equilibrium means that for each OD pair, the experienced travel time on different routes are the same, and no traveler can improve his/her travel time by switching the routes. In DTA, dynamic user equilibrium is to be achieved for every departure time interval.

To better understand the difference between STA and DTA, it is necessary to first understand the main components of traffic assignment procedures that run sequentially and iteratively seeking a convergence. These three main components are:

- Shortest path identification (also referred to as tree-building): This includes the identification of a set of attractive paths (routes) between each OD pair. In DTA, this component is time-dependent and includes updating the set of attractive paths given the estimated travel times of the paths during the previous assignment process.
- Assignment of the trip demands to the identified attractive paths: This component results in the estimation of link flows by assigning the demands to the competing attractive paths. In DTA, the proportions of demands assigned to each path are calculated for each assignment time period. In general, a time period of 15-30 minutes is most widely used.
- Network loading: This component refers to the representation of the movement of vehicles on the network as they travel from origins to destinations. Network loading allows the estimation of performance measures for use in the assignment, such as route travel time between origins and destinations. In DTA models, network loading procedures can be classified as analytical procedures or simulation procedures. Due the complexity of traffic operations, particularly with the presence of congestion and traffic control, simulation-based procedures are the most widely used types of procedures at the present time (Hadi et al., 2012).

The discussion above indicates that unlike STA, which defines the shortest paths and allocates all of the traffic to these paths at once for the whole peak period, DTA conducts the traffic assignment and reaches equilibrium for each time interval far shorter than the model period.

This is preferred in two aspects, as follows: 1) DTA can model time variant demands, time variant operational strategies (such as those applied in ML), associated travelers' responses, dynamic variations in network performance, and dynamic events such as lane blockage incidents; and 2) Simulation-based DTA can model queue building and dissipation and queue spillback due to exceeding link capacity or downstream link queuing capacity, as it occurs in the real world. Therefore, DTA provides a more realistic representation of travelers' behaviors and traffic conditions, and provides a better approach for assigning traffic and estimating travel cost and time, resulting in better demand and performance measure forecasting.

Despite the potential benefits of utilizing DTA, there are some concerns and issues hindering its use. The most common concerns identified by modelers and planners include 1) the excessive data and time needed to model and calibrate DTA networks, 2) the required time and cost for training, and 3) the time required to integrate DTA with other transportation analysis tools such as demand forecasting models, multi-resolution modeling, and ABM modeling. In particular, integration of DTA with activity-based or choice models is difficult to converge.

In April 2009, the Transportation Research Board (TRB) Network Modeling Committee conducted a DTA user survey through the Federal Highway Administration (FHWA) Travel Model Improvement Program (TMIP) mail list, which shows that more than 70% of the 85 respondents plan to apply DTA tools within two years (Tung and Chiu, 2011). On the other hand, the respondents also clearly identified the following top five technical and institutional barriers:

- DTA requires more data than current availability or accessibility (47%)
- Setting up a DTA model takes too many resources (44%)
- Cost/benefit is unclear (45%)
- DTA tools take too long to run (35%)
- Modeling approaches are unclear (35%)

Another survey was conducted in 2010 by the Florida modeling community, related to their views of DTA applications and limitations. Forty-seven responses were received from private sectors, metropolitan planning organizations (MPOs), and state agencies. Thirty-six percent of responders believed that there is a lack of data for DTA applications at this stage of development; 24% mentioned lack of experience as an obstacle for DTA implementation, 22% were concerned about calibration and validation requirements, and 21% named computational time as a DTA drawback compared to the traditional regional models. The need for training, complexity of the process, and the cost of software were also confirmed as issues when considering implementation of DTA (Hadi et al., 2012).

Convergence of DTA models should also be an important area of consideration by modelers. In static user equilibrium, the convergence of the solution is theoretically provable. However, in simulation-based DTA tools, the convergence is not theoretically guaranteed. Therefore, arbitrary performance measures were introduced as convergence criteria, with no agreed-on acceptance levels.

1.5 Goals and Objectives

The main goal of this project was to develop and assess procedures for managed lane forecasting procedures that utilize modeling based on DTA's ability meet modeling requirements that would otherwise not be achieved when using procedures that are based on static assignment. The specific objectives of this research were:

1. Develop procedures for utilizing detailed data from multiple sources to calibrate the supply and demand sides of DTA models for use in modeling ML.
2. Develop procedures for the use of simulation-based DTA in conjunction with advanced toll lane modeling for demand forecasting of modeling ML.
3. Assess the performance of the use of DTA compared to STA in ML modeling.
4. Test the ability of Cube Avenue, which is a strong DTA tool candidate for use in Florida to model ML.
5. Demonstrate the application of the developed procedures to a potential ML implementation in Florida.

The research team worked closely with the Cube Avenue platform developers to improve a prototype of the developed procedure and increase its reliability. The contribution of the research includes the following: 1) developing procedures for data acquisition and validation; 2) optimally utilizing ITS data and setting performance measures to improve supply and demand calibration; 3) developing a procedure for demand estimation; 4) extensively testing the performance of the proposed methods versus detector data; and 5) making recommendations to improve ML modeling based on the developed method.

An essential objective of this research is network and demand calibrations based on available data from multiple sources (specific objective number 1 mentioned above). This step requires significant data acquisition and preprocessing efforts. These efforts include network and demand extraction and processing from regional demand forecasting models, acquisition and integration of data from different real-world data acquisition sources, data reduction and archiving, and data validation and consistency checking. The demand and supply calibration requires defining proper performance measures and an effective procedure for the calibration.

The network and demand calibration process developed in this study is an iterative process; when it is complete, it results in a well-calibrated network and a highly reliable trip table. Although there are approaches for simultaneous calibration of network and demand, it was found that in practice, an iterative process works better and is more tractable. Thus, such a procedure is used in this study.

The calibrated network and demand-produced procedure is used in conjunction with proposed prototypes for ML modeling to assess the performance of these prototypes. Based on extensive testing of the proposed prototypes and related sensitivity analyses, recommendations are proposed with regard to the assessed prototypes and the associated parameters.

1.6 Document Organization

This section includes a description of the remaining chapters of this document. Chapter 2 presents a review and assessment of past research related to the objectives and tasks of this study. First, a review is presented of existing ML modeling frameworks that were found to vary in their levels of details and complications. Next, current practices and research in supply and demand calibration, separately or jointly, are reviewed, leading to the selection of an iterative-joint approach to supply and demand calibration for use in this project. Lastly, existing literature on convergence are reviewed, illustrating that achieving a stable and equilibrated model is important to the ML modeling that requires assessing different strategies relative to one another.

Chapter 3 describes the procedures needed to prepare the demand and network from a regional demand model for the purpose of this project. Additional modifications are needed to prepare a network for DTA applications. There was a unique opportunity in this project in terms of accessing the ITS data-rich environment. Removing non-representative day and time intervals, removing detector erroneous data, and checking special and temporal consistency are crucial tasks to preprocessing and validating this data.

Chapter 4 summarizes the development of a managed lane modeling prototype based on DTA similar to the FDOT Phase 1 approach based on STA. The prototype was developed for the Cube Avenue DTA tool. Several issues were identified with the Cube Avenue's ability to implement the prototype. These issues were resolved by the Cube Avenue developer.

Chapter 5 describes the procedure for network/supply calibration. The goal is to estimate capacity and traffic flow model parameters for network links. This process starts with replicating isolated bottlenecks and is extended gradually to cover a larger network.

Chapter 6 includes the framework for demand estimation consisting of sub-elements that can run sequentially in an ascending level of detail and complexity. In this section, proper performance measures are set to assure a reliable, reasonable estimation of demands. Estimation of demands based on dynamic OD is explored based on DTA, and heuristics are investigated to improve demand estimation based on STA.

Chapter 7 is dedicated to evaluating two approaches for utilizing DTA to assess managed lane modeling. The first approach is to incorporate the toll in the objective function. The second approach is similar to the FDOT phase 1 approach, but with the use of DTA instead of STA. Static and dynamic assignments are also compared in terms of replicating travelers' behaviors in choosing priced lane versus detector data.

Chapter 8 demonstrates the results of the model application in a real-world network consisting of ML, as well as a competing parallel path and arterial system.

Chapter 9 summarizes the findings of this research on demand and supply calibration, and the assignment module in the context of managed lane modeling.

2. Literature Review and Assessment

This chapter presents a review and assessment of past research related to the objectives and tasks of this study. First, a review is presented of existing ML modeling frameworks that were found to vary in their levels of details and complications. Next, current practices and research in supply and demand calibration, separately or jointly, are reviewed, leading to the selection of an iterative-joint approach to supply and demand calibration for use in this project. Lastly, existing literature on convergence are reviewed, illustrating that achieving a stable and equilibrated model is important to managed lane (ML) modeling that requires assessing different strategies relative to one another. Without assuring a stable and well-converged network, it is not possible to differentiate between differences in performances that are due to changes in inputs and policies and those that are due to model noise and randomness because of the non-converged models.

2.1 Managed Lane Modeling Frameworks

A variety of modeling approaches have been proposed to assess managed lane implementations. These approaches range from high-level sketch planning tools to micro-level modeling of individuals' behaviors and traffic operations.

The Federal Highway Administration (FHWA) developed an open source sketch planning tool (POET-ML) to perform a quick evaluation of ML functionality and pricing policies. The input into the spreadsheet includes eligibility policies such as occupancy restrictions; physical characteristic such as the lengths and numbers of managed lane (ML) and general purpose lane (GPL) lanes, median types, and buffer types; and demand information such as the peak hour volumes on ML and GPL facilities. The user can change the current policy according to the results produced by the tool, and can also review the potential impacts on travel demand, revenue, mobility, and the environment (FHWA, 2008a).

TRUCE 3.0 and TRUCE-ST are similar tools developed by FHWA that allow the user to quantify the impacts of congestion pricing on urban highways at the State level. The input includes aggregated traffic data from urban mobility reports (Schrank, and Lomax, 2007) and socioeconomic data from census for the desired study area. The tool allows for the evaluation of the effect of different congestion pricing policies on traffic condition, air quality, and revenue. (FHWA, 2008b).

FITSEVAL is another sketch planning tool developed for the Florida Department of Transportation (FDOT) by Florida International University in Miami, Florida, to evaluate and assess ITS alternatives in Florida within the Florida Standard Urban Transportation Model Structure (FSUTMS) framework. This tool evaluates the effects of intelligent transportation system (ITS) applications, including ML on network performance measures such as Vehicle Mile Traveled (VMT) and Vehicle Hour Traveled (VHT), average speed, and fuel consumption (Xiao et al., 2010).

When utilizing the four-step demand forecasting to model ML, the most straightforward approach is to add a toll term to the generalized path cost in the assignment module and assign a dollar value to travel time in the generalized cost function. Recently, travel time reliability was also added to the generalized cost function of the assignment. More advanced models apply a binary route choice (toll vs. non-toll), either within the assignment or externally, tying this binary choice to the assignment in an iterative manner. Recent applications have modeled the travelers' behaviors in choosing ML by utilizing probabilistic approaches, such as using a logit model based on a derived utility function or a willingness-to-pay distribution based on traveler surveys.

An essential component of the managed lane choice, whether implemented in the generalized cost function of the assignment process or as a separate logit model, is the Value of Time (VOT). VOT is a measure of a driver's willingness-to-pay for travel time savings. VOT is a means of capturing dissimilarities between different classes of drivers in route choice; more specifically, either in mode choice, route choice, or within assignment. These dissimilarities are caused by several socioeconomic and trip factors. Chiu (2012) compared modeling VOT dissimilarity in discrete choice model versus multi-class stratified assignment. In the discrete choice model, every traveler makes decisions of choice based on a generalized utility function (GPL vs. ML), while in the multi-class stratified assignment, predefined classes of travelers (stratified based on

VOT distributions) are assigned separately. The first approach is time-efficient and easier to implement, but difficult to converge. The second approach is more time-consuming, but produces a more stable solution; therefore, it is more appropriate for scenario comparison applications. The toll choice procedure in the Southeast Regional Planning Model (SERPM) is conducted utilizing the standard multimodal logit model, and conducted separately for each trip purpose and each vehicle occupancy category (FDOT, 2013).

More comprehensive models consider the toll and VOT (and potentially the value of reliability or VOR) in the utility function of the mode split and the impedance function in trip distribution. In these cases, linking different steps (assignment, mode split, and/or distribution) is essential to ensure consistency between their outputs. After partitioned toll and non-toll trips are calculated and loaded into the network in the assignment module, the travel time is skimmed and fed back into the mode choice and trip distribution steps. Mutual consistency should exist between different levels. For example, global convergence problems were reported for cases where the toll cost is modeled in the assignment generalized cost function and in the mode choice, but not in the impedance function used in the trip distribution (NCHRP, 2012).

Boyce et al. mentioned that the travel time input to trip distribution and mode choice should be equal to the travel time (cost) obtained from the equilibrium assignment in the next step. The author performed several computational experiments of how to incorporate the feedback into demand forecasting models. It was found that the direct (naïve) feedback is not efficient, and a type of averaging is needed. In comparing different alternatives of what to average and how to average, it was recommended to average the trip matrices with fixed weights (e.g., weights do not change by iterations). The converged solution will produce a matrix, that when loaded onto the network by the assignment module, generates route travel times that if fed back to trip distribution and mode choice step, would reproduce the same matrix. The same feedback procedure is applied in the enhanced demand forecasting model in Florida to overcome inconsistency issues between trip distribution/mode choice and assignment (FDOT, 2013). Global or system convergence is challenging and essential. Measures and methods may be used to evaluate the system convergence and consistency between different steps.

The FDOT enhanced the modeling of ML in the SERPM model in regard to the implemented ML. In the older released version (date of previous version release), the toll was converted to travel time by means of value of time, and was added to the travel time in the generalized cost function. However, in the improved version, a binary route choice between managed lane (ML) and general purpose lane (GPL) is performed within the assignment (FDOT, 2013).

There has been a recent interest in utilizing activity-based modeling in managed lane studies. It has been argued that traditional trip-based models are unable to respond to pricing policies in trip generation, departure time, and occupancy choices. Activity-based models are ideal for this purpose. In both approaches (trip-based and activity-based), there is a growing trend of combining mode, occupancy, and binary route choice in a multi-level nested logit model structure when modeling ML (Vovsha et al., 2013)

Recently, a survey was conducted as a partial effort to incorporate toll modeling into the existing Phoenix metropolitan area demand model by URS (URS, 2011). This survey aimed to identify the best practices of toll modeling; 17 agencies responded to this survey. Sixteen of these agencies currently use the four-step demand forecasting model, with six agencies planning to replace the conventional trip-based model with activity-based models (ABMs). Nine metropolitan planning organizations (MPOs) incorporate the toll cost in the impedance function in the trip distribution step. The Nested Logit model is the most commonly used mode choice model. Nine agencies partition the trip table between toll and toll-free users, either in their mode choice or assignment model. Almost all MPOs include VOT in combination with the mode choice, trip distribution, and trip assignment steps.

A majority of agencies responding to the abovementioned survey used a feedback loop, from trip assignment to trip distribution, or to mode choice. In almost all cases, the assignment method is static user equilibrium. Seven MPOs consider both travel time and toll cost to calculate the shortest paths. The modeling route choice, both as a sub-element of mode choice hierarchy and in the assignment step, has the advantage of sensitivity to socioeconomic characteristics. A calibrated logit-type model or willingness-to-pay distribution (diversion curve) can be used in the assignment module to define the route choice behavior. The final prototype proposed by URS was an advanced highway assignment with a customized route choice that feeds back to trip distribution and mode choice. The utility function takes into account the income levels and bias factors. A bias coefficient accounts for unknown factors that affect single occupant vehicle (SOV) decisions, such as perceived improved trip reliability, safety, and comfort.

Value of Time

A key step in ML design and modeling and predicting the associated demand and revenue is to estimate the VOT (and potentially VOR) and associated factors that affect a traveler's decision to choose or avoid paying a toll. VOT is generally referred to as the monetary toll value divided by the saved time, or equivalent "perceived" benefit for using ML. The heterogeneity of travelers is a crucial property to be captured. The necessary level of model detail requested by the National Environmental Policy Act (NEPA) for traffic and revenue (T&R) analysis includes: four to five major travel purposes, three to four income groups, and three to four time-of-day periods. Vovsha et al. (2013) recommended considering the length of trips and congestion levels in VOT estimation. It was found that drivers perceive every minute in congestion as 1.5 to 2 minutes of free-flow driving.

Recent findings recommend including travel time reliability as a decision factor in the assignment process, and subsequently, VOR was introduced in the generalized cost/utility functions. Two general approaches are introduced in measure travel time reliability. The first approach relates reliability to variability, meaning the higher variability in travel time (measured as trip travel time variance or similar concepts) is equivalent to a less reliable trip. The second approach measures reliability as a portion of success or failure against pre-established thresholds, such as proportion of trips with a delay less than a predefined threshold (Cambridge Systematics, Inc., 2012).

In order to obtain travel time reliability from stated and revealed preference surveys, the Resource Systems Group (2012) associated travel time reliability with travel time distribution entropy. It is assumed that travelers will pay to reduce the entropy. The entropy is calculated as a function of the mean and standard deviation of the travel time distribution. The value of reliability is in dollar per unit of entropy.

Minnesota was the first state to implement a fully dynamic pricing algorithm that updates the toll based on HOT lane density and density variability every three minutes, with a goal to keep the LOS at C (Janson and Levinson, 2013). By implementing different toll policies and analyzing the flow on the ML, a counter-intuitive positive correlation between pricing and ML demand was observed. The authors believed this contrary behavior is because drivers perceive the toll value as an indication of GPL congestion level. Similar results were observed in Burris et al. (2012). The authors performed data analysis on two HOT lane facilities in Minnesota and California, which revealed that in Minnesota, SOVs will pay up to $116/hour, and in California up to $54/hour to use HOT lanes during the afternoon peak, and slightly less for the morning peak. The authors interpreted these high values are not only paid toward time saving, but also for improvements beyond time saving, such as trip reliability. Alvarez's (2012) research at Florida International University showed that based on historical ITS data, people occasionally chose to pay toll during the AM peak, while the parallel GPL had a lower travel time. The Resource Systems Group (2012) showed that the saved time in ML is overestimated by travelers, by comparing joint stated and revealed preference surveys with historical data.

Much lower values are used in practice and are recommended for modeling as default values (NCHRP, 2012; Vovsha et al., 2013). These values vary between $7 per hour for SOVs, to $18 per hour for vehicles with three occupants for work trips in the PM peak. Past studies have shown that two groups of factors affect SOV decisions to use ML: 1) Trip-related factors such as trip length and purpose, trip time of day, travel time savings, improved trip reliability, safety, and comfort; and 2) Socioeconomic factors such as income level, age, gender, and household composition. It was found that the income level and trip purpose are the most influential factors (Burris et al., 2012). An Investment Grade Traffic and Revenue Study (WilburSmith, 2011) reports a range of $6/hour to $18/hour of VOT, with an average of $14.31/hour for the US 36 Corridor in Colorado.

The calibration of the ML model within the SERPM framework required the use of a value of time of $1 equal to 5.1 minutes ($11.75/hour) for VOT, and a range of $0.00 to $2.99 for VOR. This value is based on state and revealed preference surveys from fall 2011 (Resource Systems group, 2012). Calibrating models based on state and revealed preference surveys for the Florida Turnpike's tolling framework has resulted in a VOT ranging from $3/hour to $13.50/hour, based on trip purpose and income level (Dehghani et al., 2003). A customer satisfactory survey conducted by the Florida Turnpike in 2005 shows that 91% of the responders perceived the benefit of paying the toll in terms of service, safety, and convenience (Florida's Turnpike Enterprise, 2005). Nava et al. (2013) selected a VOT of $15.50/hour for SOV and HOV users and a VOT of $46.50/hour for commercial trucks. In their methodology, the toll value update mechanism is internally implemented within a dynamic user equilibrium framework, which implies mutual consistency and convergence between toll value and route choice.

Choosing ML versus GPL is a learning process for commuters. Studies show that the learning process that leads to a high correlation between saved time and ML selection takes about 60 days. In other words, it takes 60 days of adjustment prior to choosing ML over GPL, based on the saved travel time (Alvarez, 2012).

Sometimes, constants are also included in the utility functions that account for unobserved factors that lead travelers toward ML or away from it. These parameters are hard to measure and are estimated through model calibration and fine-tuning tasks. In the South Florida SERPM model calibration, bias against the HOT lane choice is inserted in the utility function for off-peak periods. For peak periods, a bias toward HOT lane is included to replicate the observed volume on the HOT lane. In addition, saved travel time is exponentially increased with congestion level to reflect how travelers "perceive" the benefit of using HOT versus GPL when the road is heavily congested. This effect was revealed in travel surveys and stems from better safety, comfort, and reliability when using HOT (FDOT, 2013).

It should be noted that the revealed VOT for ML might be different from the VOT for toll facilities when the entire facility is tolled. This is because with ML, drivers can decide at the last moment which route to take based on dynamically changing traffic conditions and tolls.

Moreover, with ML, usually a small portion of ML capacity can be purchased by SOVs, therefore, only SOVs with relatively high VOT will divert compared to toll facility users.

2.2 Supply/Network Calibration

Supply calibration includes the estimation of parameters associated with traffic operations in the network. These parameters vary depending on the type of the model (macroscopic, microscopic, or mesoscopic) and the specific tool under consideration. The parameter used in mesoscopic simulation tools generally include segment capacities, free-flow speed, queuing density and/or jam density, and/or other parameters used in the macroscopic traffic flow model used to move the vehicles onto highway segments. The performance of the system with the selected parameters is evaluated by comparing the model results to real-world measures of traffic flow, such as queue formation and spillback, density, and travel time on each link.

2.2.1 Mesoscopic Simulation Supply Calibration

Kunde (2002) calibrated the network supply of the DynaMIT model through a sequential process at increasing levels of aggregation. The process starts at the level of separated bottlenecks where capacity is estimated by various methods based on field data. The network is gradually extended to connect the bottlenecks, and then model the whole corridor. The parameters from the previous steps are fine-tuned, and the supply-demand calibration runs iteratively until a desirable convergence is achieved. The most disaggregated level is the individual segment level, at which the speed-density relationship and capacity are calibrated. At this stage the interactions between adjacent segments is ignored. Due to the lack of data and large number of variables, network segments were first grouped into 11 representative clusters. All segments in a cluster were set to have the same traffic flow model (TFM) parameter values. The next stage is to perform calibration at the sub-network level where the origin-destination flows can be reasonably estimated solely from the sensor counts, because the probability of a second alternative route choice between each origin and destination is zero or negligible. This way, the impacts of errors in demand estimation on supply calibration are avoided. The last step is the network-wide calibration, which takes into account all of the interactions between various segments and any errors due to demand estimation. Stochastic optimization is used to calibrate the supply at the whole network level.

To estimate the macroscopic TFM model parameters in Dynasmart-P with the modified Greenshields model as utilized TFM, Mahmassani et al. (2004) rewrote the model formula in the natural logarithmic form, whereby the relation between speed and density becomes linear. The authors estimated the parameters by performing multiple runs of regression analysis. In each run, they set one of the parameters as fixed and systematically changed the other parameters within a reasonable range to determine the optimum combination that replicates detector data.

Wang et al. (2009) applied Kalman filtering to continuously estimate the state of the traffic based on real-time data. Capacity and TFM parameters were calculated within a stochastic nonlinear macroscopic TFM framework by an adaptive estimator. This method does not require an initial estimation of the parameters; it automatically adapts to changes in the model due to changes in external conditions and can recognize interruptions due to incidents. The drawback of this method is that the output cannot be related to the theoretical aspects of traffic flow.

The Highway Capacity Manual (TRB, 2000; TRB, 2010) is used as the authoritative source of defining and estimating capacity in the United States. A procedure is presented in the HCM that allows estimating freeway capacities based on free-flow speed. The procedure allows adjusting the capacity estimates to account for deviations from default conditions, considering a limited number of factors. However, many other parameters affecting capacity are not considered in the adjustment. Thus, the HCM encourages measuring capacity in the field to consider the differences in geometry and driving characteristics between different regions and facilities.

The remaining subsections of Section 2.2 discuss in more detail the specific aspects of the calibration process, including bottleneck identification, free-flow estimation, capacity, and TFM parameter estimation.

2.2.2 Bottleneck Identification

In a congested network with recurrent bottlenecks, the most crucial part of network calibration is to replicate bottlenecks as they happen in the real world, in time and space, and correctly estimate the capacity and impacts of the bottleneck.

A bottleneck is defined as a point upstream of which a queue is formed, with the traffic flowing at free-flow speed at downstream locations (Bertini et al, 2008). Bottlenecks can be active or hidden. A hidden bottleneck is a potential one that is a result of geometric or demand features but cannot be observed because the approaching traffic demand is metered by another upstream bottleneck. An active bottleneck is the only location where capacity can be measured based on field data. Chen et al. (2004) identified bottlenecks based on the speed differences between adjacent detectors, where the speed at the upstream detector is below a particular threshold (e.g., 40 mph), and the speed drop is above a particular threshold (e.g., 20 mph). The required parameters, including the maximum speed threshold, minimum speed difference between adjacent detectors, and data aggregation levels were recommended to be site-specific. Zhang and Levinson (2004) identified bottlenecks based on the occupancy differences between adjacent detectors. Hall and Agyemang-Duah (1991) used the occupancy-to-flow ratio as a bottleneck identification criterion. Bertini and Myton (2005) used cumulative vehicle counts and cumulative occupancy graphs to identify bottleneck activations without the need to set speed or occupancy thresholds.

2.2.3 Free-Flow Speed

Free-flow speed (FFS) is a crucial parameter in the HCM capacity estimation procedure for uninterrupted facilities. The HCM provides a free-flow speed estimation procedure that incorporates reduction factors to account for deviations from basic conditions. Reductions in free-flow speed will implicitly drop the capacity, according to the HCM procedure.

Equations 2-1 and 2-2 show the relationship between the basic and adjusted free-flow speed to account for the deviations from basic conditions in the HCM 2000 and HCM 2010 respectively.

$$FFS = BFFS - f_{LW} - f_{LC} - f_N - f_{ID} \tag{2-1}$$

BFFS= base free-flow speed (75 mph for rural freeways and 70 mph for urban freeways),
f_{LW} = adjustment factor for lane width (mph),
f_{LC} = adjustment factor for right shoulder lateral clearance (mph),
f_N = adjustment factor for number of lanes (mph), and
f_{ID} = adjustment factor for interchange density (mph), and

$$FFS = 75.4 - f_{LW} - f_{LC} - 3.22TRD^{0.84} \tag{2-2}$$

TRD = total ramp density (ramp/mi).

The HCM encourages users to measure FFS in the field as the average of all vehicle speeds when the volume is less than 1000 pc/ln/hr. Chao et al. (2005) used the average of speeds when occupancy is below 10 percent. Dervisoglu et al. (2009) estimated FFS by fitting a straight line to the uncongested part of the fundamental diagram.

2.2.4 Capacity Definition and Estimation

The HCM defines freeway capacity as the maximum sustained 15-minute flow rate that can be accommodated by a uniform freeway segment under prevailing conditions. As mentioned earlier, the HCM recommends values of capacity based on free-flow speed, and provides a few adjusting factors to account for deviations from prevailing conditions. However, there is evidence that these adjustments are not enough to reflect the significant differences between locations due to geometry, demand, and driving characteristics (Washburn et al., 2010). Given a determined FFS and weather condition, heavy vehicle and driver population are the only factors used to adjust the capacity. The heavy vehicle percentage can be obtained by detectors that classify vehicles, or by manually counting vehicle classes. However, the driver population, which is the percentage of non-commuters that are not familiar with the analyzed highway, is very difficult to estimate.

To account for site specifications, direct measurements of capacity were recommended. In absence of a recommended method by the HCM, researchers proposed a number of approaches for these measurements. Dervisoglu et al. (2009) estimated capacity as the maximum observed 5-minute flow rate over several days. Chao et al. (2005) estimated the capacity as the maximum hourly flow observed during a 30-day period. Jia et al. (2010) estimated capacity as the average of the top one percentile of a 15-minute flow rate over several days, which turned out to be similar to values estimated by the HCM.

Van Arem and Van der Vlist (1992) estimated capacity by determining the maximum occupancy in the uncongested part of the fundamental traffic flow diagram and the associated volume. Bassan and Polus (2010) approximated the capacity by fitting data into parabolic speed-flow and flow-occupancy models. Similarly, Wang et al. (2009) used the apex of a flow-density curve as capacity. Rakha and Arafeh (2010) performed an automated fitting procedure of a quadratic speed-flow function to loop detector data. This function combines the microscopic Pipes car-following model and the single regime Greenshields model. The automated model calibration yields an estimated number of key parameters, including capacity.

Researchers have also argued that capacity is not constant, even under identical external conditions (Elefteriadou et al., 1995; Minderhoud et al., 1997). These researchers recommended a paradigm shift in capacity calculation, from a deterministic value to a stochastic value, and proposed statistical methods to measure capacity. In most of these studies, capacity is tied to the notion of traffic breakdown. The most common proposed values as capacity representatives are queue discharge flow and the maximum flow before breakdown. The queue discharge rate is defined as the long-run average of flow over the breakdown period. Pre-breakdown flow was measured using different time intervals before breakdown, such as 5 minutes and 15 minutes (Elefteriadou and Lertworawanich, 2003; Hall, and Agyemang-Duah, 1991).

Based on a lane-by-lane analysis of breakdown, Dehman (2012) pointed out that in some cases, the flow increased after the breakdown and explained that this mainly happened because of lane changing between underutilized and fully utilized lanes. Brilon et al. (2005) found that a freeway operates at the highest expected efficiency only if it is loaded to 90% of the conventionally estimated capacity.

There are no guidelines on whether to use pre-breakdown, queue discharge, or a weighted combination of both as values representing capacity (Zhang and Levinson, 2004). It has been reported, however, that queue discharge is lower than the pre-breakdown flow by 2 to 26 percent in different studies, mostly due to a change of driving behavior to stop and go status (Yeon et al., 2007; Hall, and Agyemang-Duah, 1991). The HCM 2010 also recognizes this phenomenon; however, it does not consider it in its procedures and does not recommend any specific percentage of capacity reduction after traffic breakdown.

A freeway facility HCM computational engine was developed to implement the HCM 2010 Chapter 10 procedure, so as to estimate freeway capacity when queue exists. In this engine, called FREEVAL, oversaturated conditions are followed by a user-defined drop in capacity, reflecting the queue discharge rate during these conditions. The National Cooperative Highway Research Program (NCHRP) project 3-96 also aimed to develop methods for the performance assessment and capacity analysis of managed lanes compatible with HCM procedures. The result of this project is the development of additional features in FREEVAL, resulting in the FREEVAL-ML package that allows modeling of the GPL and the parallel ML (Wang et al., 2012).

In more recent studies, to account for the probabilistic nature of capacity, some researchers recommend calculating it as a percentage of the breakdown probability distribution. The most common utilized probability functions are the normal and Weibull distributions (Hall, and Agyemang-Duah, 1991; Elefteriadou and Lertworawanich, 2003; Brilon et al., 2005). Minderhoud et al. (1997), which state that given a true distribution of capacity, one can obtain the capacity value by choosing the average, median, or 90th percentile of the distribution. This choice so far has been arbitrary and supported by the results from testing the local data goodness-of-fit. There is no consensus on which point of the breakdown distribution should be used to estimate capacity.

Lin (2009) used bi-level linear programming to exclusively calibrate capacity in a DTA model. The upper level problem minimizes the deviation of simulated and observed occupancy data, and the lower level runs a simulation-based cell transition assignment.

In summary, the HCM is regarded as the most reliable source for estimating capacity for different facility types. However, the HCM procedures allow for the use of a number of factors to reflect local conditions. In some cases, however, this adjustment may not be sufficient, and direct measurement of capacity is needed. A variety of surrogate measures have been proposed for capacity measurements. It is worth mentioning that in some studies, the measured capacities reported as being lower than those estimated by the HCM (Washburn et al., 2010).

2.2.5 Traffic Flow Model Parameter Estimation

Dynamic traffic assignment (DTA) tools use mesoscopic simulation models to generate and track individual vehicles, but move vehicles according to macroscopic relationships that are subject to link capacity and link storage limits. Depending on the specific model under consideration, the utilized macroscopic relationships could include the Bureau of Public Roads (BPR) relationship, the modified Greenshields model, the Van Aerde model, or the Akcelik model.

BPR is the most common model in traditional static traffic assignment (STA)-based demand forecasting applications. It has also been used in the Cube Avenue DTA tool (Citilabs, 2013). Different values have been suggested by practitioners to calibrate the BPR curve parameters to better replicate observed performance measures such as speed, volumes, total Vehicle Miles Traveled (VMT), and Vehicles Hours Traveled (VHT). In some applications, the parameters are set based on facility type and design speed. In more advanced applications, a volume/capacity (v/c) threshold is selected to divide the BPR curve into two different regions with different coefficients to reflect the difference in traffic dynamics between these two regions. The v/c values of 1, 2, and 4 have been used as thresholds in different studies (Spiess, 1990a; Singh, 1995; Dowling, 1997 and Hansen, 2005).

Saberi (2010) compared the results from the HCM empirical speed-density curves, BPR formula, and Davidson formula (Davidson, 1966 and 1978) and its descendent, the Akcelik formula (2003), and assessed their abilities to replicate the observed speed-density curves. The author recommended the use of the BPR curve for v/c <1 and Akcelik formula for v/c >1, since this formula accounts for the presence of queue. The author incorporated the probability distribution of capacity into the speed-density relationship to account for the stochastic nature of capacity.

Huntsinger and Rouphail (2011) improved the accuracy of the BPR, Conical, Akcelik and HCM traffic flow models by replacing the volume with the estimated demand in these TFMs. The demand is calculated as the summation of volume at capacity and queue at the bottleneck location. The authors optimized the parameters of the abovementioned TFMs to fit the demand/capacity versus travel time observations.

Dervisoglu et al. (2009) presented an automated empirical calibration approach of TFM parameters for a cell transmission model. The TFM is formulated as a triangular relation between flow and density. Capacity is estimated as the maximum 5-minute flow rate over several days. This value of flow on the flow-density curve is then projected horizontally to meet the free-flow speed line (a line from the origin of the diagram with a slope equal to free-flow speed) to establish the tip of a triangular fundamental diagram. This point corresponds to the critical density, above which the flow is considered to be congested.

Van Aerde and Rakha (1995) performed an automated fitting of a quadratic speed-flow function. This function combines the microscopic Pipes car-following model (applied in CORSIM) and a macroscopic single regime model (the Greenshields model). Speed and volume (and density if available) measurements from detector data were used to calibrate four parameters that define the relation between speed and density.

Chiu et al. (2010) introduced a vehicle-based mesoscopic model called the Anisotropic Mesoscopic Model. Instead of using the conventional TFMs that assume the same speed for all vehicles on a link at a given time step, vehicles on a link can travel at different speeds. In this model, the speed is affected by the presence of leading vehicles within a neighborhood, called the speed influence region (SIR).

Loudon (2007) pointed out that the traffic characteristic is quite different in ML, compared to GPL. In particular, the observed ML speeds were found to be lower than the original estimations, depending on the degree of separation between ML and GPL. This is due to the interaction between ML and its adjacent, more congested GPL lanes. This effect is referred to as "side friction," the degree of which depends on the separation type. The most significant effect was observed with marker painting buffers, and the least significant was observed with concrete barriers.

It is not feasible to estimate the capacity for every link when estimating the capacity in the field; first, because capacity can only be observed at critical link locations. This requires grouping road segments, which significantly reduces the size of the parameter estimation. Clustering can simply be based on geographical features such as number of lanes, horizontal/ vertical curve, and closeness to ramps (Balakrishna, 2007; Kunde, 2002), or through machine learning approaches such as the k-means algorithm.

2.3 Demand Estimation

Time-dependent origin-destination matrices are essential input to trip-based DTA models. Because of the very high cost of travel surveys, possible errors with these surveys, such as misreporting the trips and the need for a fine-grained demand matrix covering short-time intervals, methods must be developed to estimate reliable fine grained trip origin-destination (OD) matrices based on initial seed OD matrices obtained from demand forecasting models. Although seed OD matrices are very important in the estimation process, other sources of data, such as traffic counts and possibly partial OD matrices measured using Automatic Vehicle Location (AVL) or Automatic Vehicle Identification (AVI) data, are needed to improve the accuracy of the estimated matrices. The OD estimation methods can be categorized as assignment-based and non-assignment-based. Non-assignment-based methods apply traffic conservation relations between entrance, exit and mainline volumes. These methods are mostly limited to road facilities without signals and without any queues. Other sources of information, such as AVI, are also difficult to incorporate into the models.

In general, the problem of OD estimation is underspecified, which means that the number of equations based on traffic counts on links are far less than the number of unknowns (OD table cells). Thus, different combinations of OD pairs can produce the same set of link volumes if loaded onto the network. To circumvent the problem of underdeterminacy, researchers may aggregate ODs over longer time intervals, compare them to surveillance data time intervals, or alternatively, disaggregate the surveillance data into shorter time intervals. (Tavana, 2001; Gupta, 2005)

Assignment-based models utilize traffic assignment to map OD matrices to link volumes, allowing for the minimization of the deviation between model outputs and observed or estimated measures (such as initial OD matrices and measured traffic volumes) (Chi, 2010). Different sources of data are easy to incorporate into assignment-based models. In addition, if dynamic assignment is used, queues and signal delays are modeled by the DTA simulator. Thus, they are more appropriate to use than non-assignment-based estimation. However, the quality of the results of assignment-based models depends on the availability of high quality initial OD matrices (Lin, 2006). A main interest of this study was the current work being performed on the assignment-based OD estimation processes and the factors affecting this estimation.

Traditionally, assignment-based OD estimation is modeled as an iterative bi-level optimization, where the upper level minimizes the deviation between observed and simulated quantities, and the DTA simulator at the lower level produces a link-flow proportion matrix as a result of loading the OD over network links. The simplest structure for this approach is depicted in Equation 2-3.

$$D = \arg\min \sum_l \sum_t f\left(c_{(l,t)} - \hat{c}_{(l,t)}\right)^2 \qquad (2\text{-}3)$$

This is subject to the constraint:

$$\hat{c} = \hat{p} * D$$

and other sets of constraints, as discussed below.

In the equation above, D is demand, and c and $\hat{c}$ are observed and estimated traffic counts. The link with the detector measurement is 1, t is the time interval with traffic data, and $\hat{p}$ is link-flow proportion matrix that indicates which portion of each OD pair travels on a certain link. This matrix is usually obtained as a result of DTA modeling. The objective function is not limited to minimizing the deviation between simulated and observed counts. It can be extended to consider the deviation between simulated and observed speed, density, queue length, or the distance between an initial set of demands (seed OD matrices) and the estimated demands. Constraints also include, but are not limited to, non-negativity constraints, initial values, link capacities, cordon line counts, fixed OD flows, and/or production/attraction counts. Even route choice probabilities can be used as constraints if these parameters are to be fixed.

Tavana (2001) modeled the upper level of the OD estimation problem as a generalized least square (GLS) optimization to minimize the discrepancy between the estimated and measured link volumes. In favor of GLS, Brandiss (2001) pointed out that GLS formulation allows the incorporation of information about the reliability of measurements in terms of a weighting matrix. Alternatively, maximum likelihood and maximum entropy methods can be used instead of GLS. To incorporate information from historical OD matrices, Tavana (2001) included a Bayesian inference that updates demand based on the results from the bi-level optimization. Alternatively, the distance between the estimated and target OD matrices could be incorporated into the objective function, as in Gupta (2005).

The upper level of the OD matrix estimation problem in Zhou (2004) is a weighted minimization of the deviation between the observed and simulated demand and link flows. Weights can be used in the upper level optimization function to reflect the level of reliability that the user wants to apply on demand or link flow measurements. Similar to Zhou (2004), Chi (2010) used adaptive weights on different components of the objective function. At the beginning of the estimation, higher weights were assigned to traffic measurements such as counts, speeds or travel times, since at the beginning of the process, these data are more reliable than the OD matrices from the demand model. As the system converges, a better estimation of OD is obtained, and the weight of the observed link counts is reduced in the optimization since they are not error-free.

The adaptive weights can also mitigate the problem of overfitting of the observed counts. The optimum value of the weight can be obtained through least square estimation, or the model user can arbitrarily set them based on local knowledge. Another issue is that in congested networks, the volume is not an incremental function of demand; therefore, Chi (2010) proposed detecting congested segments temporally and spatially, and using density instead of volume in the objective function for congested segments, which is a better representative of traffic conditions. Mahmassani et al. (2004) carried out the supply and demand estimation tasks in a sequential manner. They first calibrated the network as described in Section 2.2, and then used a bi-level optimization to estimate the OD matrices, similar to Tavana (2001). The authors investigated two different alternatives for the optimization part. The first approach was a linearly constrained GLS approach that minimizes the deviation between the estimated and observed link flows. The second approach was a weighted objective function whereby a higher weight was allocated to the links that carried more flow. In both approaches, weights were allocated to the objective function components, as discussed in Zhou (2004) and Chi (2010). The authors mentioned that using sparse matrix structure and decomposing the OD matrices into sequential sub-matrices can alleviate the problem of scalability. Fixing the OD cells that have no or little effect on traffic conditions and restarting the estimation with fewer variables increased efficiency.

Other approaches that were used to demand estimation are the Bayesian Inference and state-space framework, which are described below.

Bayes Theorem can be stated as Equation 2-4:

$$P\,(A|B) = \frac{P(A)P(B|A)}{P(B)} \qquad OR \qquad P\,(B|A) = \frac{P(B)P(A|B)}{P(A)} \tag{2-4}$$

Considering A as network conditions and B as travel behaviors, the formula above can be interpreted as predicting network (supply) behavior, given the demand (P($A|B$)). Equivalently, it can be interpreted as predicting demand behavior, given the network conditions (P($B|A$)). In the joint supply-demand calibration, the mutual relationship between supply and travel behavior can be modeled through the Bayesian Inference, as shown in Figure 2-1.

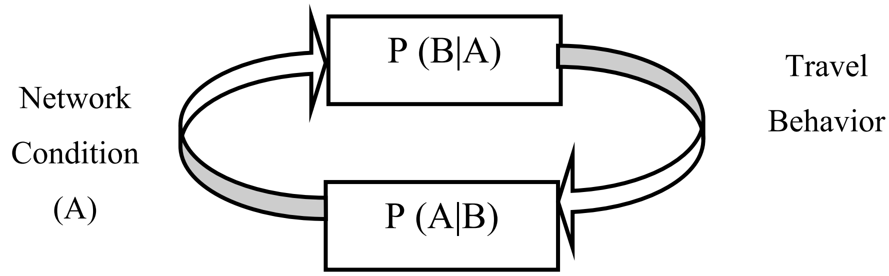


Figure 2-1 Interaction between demand and supply

For a linear dynamic system, the state-space framework can be summarized, as shown in Equations 2-5 and 2-6 (Chen, 2003):

$$x_{h+1} = f * x_h + w_h \tag{2-5}$$

$$y_h = g * x_h + v_h \tag{2-6}$$

Equation 2-5 is called a "state or transition formula," and shows how a state vector (x) evolves over time by evaluating P (xh+1| xh), the probability of xh+1, given xh. The state vector can be OD flows, travel behavior parameters, speed-density relation parameters and so on. Equation 2-6 is called the "measurement equation" and maps the observation vector (y) to the unobserved state vector (x), or describes the probability P (yh|xh). The model coefficients, f and g, need to be estimated, and w and v are model noises. The detector data, such as volume and speed, are examples of y. A well-known solution for the state-space model is Kalman filtering, in which model noises (w and v terms in Equations 2-5 and 2-6 are assumed to be a normal distribution with a mean of zero.

Ashok and Ben-Akiva (2002) and Lin (2006) modeled the relationship between demand and link flow as a state-space formula. It should be noted that in congested areas with capacity constraint, link flows do not represent the demand. Capacity plus queue at the link can be used to approximate the demand. Hu and Chen (2004) estimated OD and travel time simultaneously through extended Kalman filtering. Zhou (2004) defined the true demand be estimated as a combination of regular pattern, structural deviation from the mean pattern, and random fluctuations, and applied Kalman filtering to capture these components. Kalman filtering was used as an external controller to inspect the adjusted OD before sending the OD estimation output to the DTA simulator.

The growth of ITS implementation is very promising in collecting full or partial trajectory data. With commercialized connected-vehicle devices mounted on cars, more trajectory information will be available in the future. Zijpp (1997) and Zhou (2004) were able to reduce OD estimation errors by combining AVI and count data. Dixon and Rilett (2002) deployed GLS and Kalman filtering to show the benefits of the incorporation of origin-destination and travel time information from AVI data.

Doblas and Benitez (2005) pointed out a practical aspect of OD estimation that was ignored in related studies. The preservation of the structure and pattern of initial OD should not be sacrificed to replicate traffic counts. Traffic counts reported by detectors are not error-free. Moreover, the information in the initial OD (usually from surveys) is very valuable and expensive, and deviation from initial OD structure should be constrained. The authors modified the gradient-based algorithm of Spiess (1990b) implemented in commercial DTA tools to control the adjustment of the OD matrices by preserving the number of production and attraction trips for each zone. To optimally use the available data, Nguyen (1982) incorporated production/attraction data from a historical OD matrix to a maximum entropy formulation.

In summary, the assignment-based OD matrix estimation problem that is of interest to this study was formulated using a number of methods, including bi-level optimization (utilizing a GLS or maximum likelihood approach), state-space problem, or Bayesian inference. The latter two methods can also be used in conjunction with the bi-level optimization problem to update the OD matrices based on the results from the optimization, in an iterative process. They can also be used as an external controller to limit the deviation of the estimated OD matrix from the initial or historical matrix. Depending on the source and quality of the initial or historical OD matrix, certain features of the matrix may be necessary to keep. For instance, some or all of the attraction production rates or some OD pairs might be kept constant during the estimation process.

2.4 Combined Supply and Demand Calibration

It is logical to suspect that there is a relationship between the supply calibration discussed in Section 2.2 and demand calibration discussed in Section 2.3. Supply calibration requires a good estimate of demand, and demand calibration requires a well-calibrated network. Doan (1999), Antoniou et al. (2007) and Vaze (2007) showed that joint supply-demand calibration is superior to the sole use of calibrating demand. There are two main approaches to demand-supply DTA calibration: sequential process that can be performed iteratively (Balakrishna, 2002; Mahmassani, 2004), or simultaneous estimation of all parameters (Balakrishna, 2007; Vaze, 2007).

Antoniou et al., (2007) utilized a nonlinear state-space model to jointly calibrate supply and demand in an online framework. Ashok and Ben-Akiva (1993) used the deviations of the model parameters from the best estimated parameter instead of the parameters themselves, as part of a joint supply-demand calibration process. This way, all available information (obtained from estimation in previous steps) would indirectly be incorporated into the model structure. The network was composed of 45 segments of a mainline freeway and associated ramps (no route choice behavior was involved). The author decomposed the problem and sequentially calibrated supply and demand parameters. Segment capacities were estimated according to the HCM methodology, and the TFM parameters were found by fitting the modified Greenshields model to sensor data for three grouped segments. Utilizing a similar approach, Vaze (2007) calibrated all network parameters, route choice parameters, and OD matrix elements in DynaMIT through state-space modeling, as well as through stochastic optimization modeling.

Chi (2010) conducted a weighted bi-level optimization to calibrate the supply parameters and estimate OD demands in a freeway system. The network (supply) was calibrated once before the OD estimation by fitting observed data to the modified Greenshields model, and once afterward to fine-tune the parameters obtained from the previous stage. Fine-tuning of the TFM parameters was carried out through bi-level optimization. The author also showed that the incorporation of an initial OD estimate can improve the overall performance of the estimation. In absence of historical OD estimates, a gravity model was used to produce an initial OD matrix.

This matrix was then improved using a static OD estimation module that utilizes a maximum likelihood framework.

Balakrishna (2007) estimated all parameters of the supply and demand sides through stochastic optimization. Following Kunde (2000) and Vaze (2007), he utilized Simultaneous Perturbation Stochastic Approximation (SPSA) to simultaneously estimate hundreds of parameters on the network. Although this method is theoretically elegant, it has not been implemented successfully in real-world applications.

Interrelation between supply and demand was carried out through sequential and simultaneous processes. Simultaneous estimation of all parameters, although asserted to be more efficient, complicates the problem and limits the user's ability to monitor and control the change of parameters. Moreover, incorporating local knowledge about the network or the demand is difficult in this approach, since a large part of the optimization is automated. No successful application of this approach in the real world has been reported so far.

2.5 Convergence

Another issue that will be explored in this study is the quality of the traffic assignment solution, as measured by convergence. By definition, the user equilibrium (UE) is achieved when travelers cannot improve their travel times by selecting alternate paths, given their departure time. This implies that every used path between an origin and destination is a minimum cost path and that there are no changes in flow patterns or experienced travel times between assignment iterations after the convergence is approached. Convergence of the user equilibrium assignment is necessary to ensure the integrity of the resulting solution and to ensure that the model can be used in assessing alternative designs and operational strategies.

A number of approaches were proposed to solve the static and dynamic assignment problem. Some of these approaches are heuristic approaches, and others involve more rigorous mathematical programming (Ortúzar and Willumsen 2001). The mathematical programming approaches express the assignment problem as an objective function subject to constraints representing traffic flow properties.

The mathematical assignment methods generally allow the proof of optimality and uniqueness and produce superior solutions to those obtained utilizing the heuristic approaches. However, due to the complexity of the dynamic network loading functions required for DTA, the traffic flow models in DTA problems are generally non-differentiable. Therefore, heuristic algorithms that do not require derivative information are used for simulation-based DTA. Although with heuristic assignment, no formal convergence proof can be given, as is the case with mathematical solutions, measures of gap similar to those used in static equilibrium assignments that are based on mathematical solutions can be used to assess the quality of a solution. Still, heuristic approaches with simulation-based DTA fail to guarantee optimality and convergence.

Boyce et al. (2002) pointed out that a relative gap of 0.01% (0.0001) is required for static assignment so as to ensure sufficient convergence to achieve link-flow stability. There is no positive agreement on what represents an acceptable value of the relative gap in DTA. It was realized, however, that it is much more difficult to achieve a small relative gap in simulation-based DTA compared to static assignment, particularly for congested conditions (Chiu et al. 2011). The dynamic nature of traffic flow, particularly during congested conditions and the heuristic nature of the UE problem in DTA, makes it more difficult to achieve convergence in DTA, compared to STA.

A widely used measure for calculating convergence is called the “relative gap,” which measures the difference between the current iteration solution and the ideal solution. The ideal solution is loading the whole volume on the single shortest path (Chiu et al., 2011). This concept was applied with slight differences in the formulation in different studies. Link-based measures versus path-based measures have also been suggested by researchers, with recent discussions on the subject indicating that path-based (also referred to as trip-based) measures might be more meaningful (Chiu and Bustillos, 2009). Path-based or trip-based measures exploit disaggregate and tractable information of trips instead of aggregated link volumes. In addition, path-based criteria provide additional information that allows utilizing heuristics targeting those trips, travelers, households, or market segments that have the most impeding convergence to achieve better solutions (Resource Systems Group, 2010).

The relative gap should not be considered an ultimate qualification for the UE solution. A well-known problem of UE is that although it produces a unique set of link volumes, there can be multiple route solutions associated with these volumes. This can be a serious issue in problems such as select link analysis and subarea analysis. It is possible to define the unique desirable UE path set by setting some extra constraints on the assignment solution to avoid violating the conditions of stability and proportionality.

Bar-Gera et al. (2010) pointed out that even if the link flow reaches convergence, a main issue with route flows is that they are not uniquely determined by the UE conditions. Reaching path flow convergence is particularly important for applications, such as multi-class assignment, select link analysis, estimation of origin-destination flows from link flows, derivation of OD flows for a subarea of a region, average travel time and average distance per OD in a generalized cost assignment, and so on. It was found that among all possible UE routes, there was just one that maximized the entropy, which should be considered the unique solution. It is proven that this solution also meets the proportionality condition. The proportionality requirement is defined by Bar Gera et al. (2010) in that the proportions of travelers on each of the two alternative segments should be the same regardless of their origin or their destination.

Lack of convergence can also affect the consistency and stability of the resulting solutions. Consistency is defined as the contribution of all eligible routes to the UE solution. This means that all routes should be included in the UE solution, unless there is a good reason for not being considered, like having a high generalized cost. Lu and Ni (2010) defined stability as the solution ability to accordingly respond to perturbation, meaning that if small changes in the network or demand are made, the model should respond to it with reasonable changes. On the other hand, Chiu and Bustillos (2009) and Peeta et al. (2011) state that a network is stable when link volume does not fluctuate, and a network is consistent when it responds appropriately to small perturbation.

A small relative gap does not assure a credible dynamic assignment solution. Lu and Ni (2010) showed that even with a very small relative gap (10^{-12}), misleading results that look reasonable may be obtained, yet respond unreasonably to small perturbation. For instance, a 10% decrease in capacity of a secondary road might cause serious congestion in another part of the network. Consistency, proportionality, and stability are needed to check for the evaluation of alternative treatments of the transportation system, and for applying methods such as select link analysis, select zone analysis, and subarea analysis. This is also very important to ensure unique solutions of multi-class assignments, particularly in ML where preferential treatments of some of the classes are applied (Boyce et al. 2010).

2.6 Conclusions Based on the Literature Review

Managed lanes were accepted as effective countermeasures against freeway congestion. These facilities are proactively operated in response to traffic situations, by means of access management, variable toll policies, and vehicle eligibility constraints. Choosing a managed lane versus choosing a free alternative road was modeled variously as part of trip-based or activity-based frameworks. Assignment is a critical step in demand forecasting that reflects the effect of congestion pricing on drivers' route choice behaviors. A model can only replicate real-world observations when supply and demand calibrations are completed.

Supply or network calibration entails estimating capacity and traffic flow model parameters for each link in the network. These parameters affect the travel time, the congestion time, queue formation and queue spillback when the demand is loaded. Demand calibration, also referred to as trip table estimation, is used to estimate a trip table that produces observed link counts and congestion patterns when loaded onto the network.

Joint calibration of supply, demand, and route choice parameters is confirmed to be superior in the process of separately calibrating these components. Two different approaches, sequential and simultaneous calibration, have been used by researchers and practitioners. Despite well-established mathematical formulations and solutions for simultaneous supply and demand estimation, their implementations in the real world are not straightforward and have not been executed. Once the supply, demand, and route choice parameters are selected as described above, additional fine-tuning of the parameter may be needed to adjust local variables to produce the observed queues and operations

It should be noted that replicating traffic volumes does not guarantee a well-calibrated network. Temporal-spatial congestion patterns should be reasonably replicated. Estimated OD matrices should also be consistent with other sources of data, such as zonal information from the production/attraction step or from the trip distribution step and certain attributes of the historical OD matrices. Simulated queue length and/or density are other measures that should be checked against the estimated values from field observations when the network is congested and the demand is not easy or possible to obtain. In the objective function used to estimate OD matrices, adjustable weights on different components can reflect the level of confidence in the data and improve the performance of the estimation. These weights can also reflect the importance of individual segments of interests, such as bottlenecks or locations with volumes that better replicate the changes in demand patterns.

Ranking links based on their contribution in updating OD routes reduces computational time. Also, OD elements that do not significantly affect the assignment can be fixed to reduce the size of the OD estimation problem. OD matrices can be aggregated into longer time segments, compared to the observed data time interval, so as to alleviate the problem of underdeterminacy. Origin-destination survey data is very valuable if available, and a structural deviation from it should be avoided at all costs. Different logic and reasonableness criteria should be devised into the OD estimation procedure as a feedback process to avoid error propagation.

Different aforementioned methods of OD estimation should be empirically tested to determine which method can better preserve the historical OD pattern, which is the most computationally efficient, and which can better replicate congested network conditions. Investigating the optimal modeling of the supply-demand joint calibration also requires empirical testing.

Assignment convergence and joint calibration convergence should be properly addressed and checked. Assignment convergence should be checked for each time interval and for each OD pairs. Producing a converged network; however, to assure model credibility, stability, consistency and proportionality should also be checked, as explained in Section 2.5.

3. Data Acquisition and Validation

Advanced modeling tools, such as dynamic traffic assignment (DTA), demand more detailed and higher-quality data to ensure that the developed model accurately replicates real-world conditions. Compared to static traffic assignment (STA) models, DTA requires more detailed and refined network representation and additional data details, both temporally and spatially. Moreover, congestion data such as queue presence and queue length should be incorporated into DTA calibration, while such data is generally not used in STA-based tools. Traffic control details are also needed if the impacts of traffic control are to be accurately modeled.

In this study, the network and an initial estimation of the associated trips were extracted from a regional planning model. The performed network editing efforts and refinement of the initial demand for use in DTA are discussed in this section. This chapter also describes the collection of traffic detector data that provides estimates of measures, which are essential to the development and calibration of simulation-based DTA tool applications. These measures include traffic volumes that can be used in the derivation of time-variant trip tables, as described in Chapter 5 of this document. In addition to these data, volume, speed, and occupancy measurements were obtained for use in the calibration of the applications. Detector data requires careful examination and a significant amount of time for filtering and processing to exclude and/or correct suspect data.

3.1 Network and Demand Data Extraction

3.1.1 Subarea Network and Matrix Extraction

The study area was extracted as a subarea network from the ECRLRTPMDL020110 (S65TODMDL) version of the Southeast Regional Planning Model (SERPM), as depicted in Figure 3-1. The subarea boundary can be specified using the Cube Polygon feature or a GIS tool. The Cube can then be used to extract the subarea network from the SERPM model network by using this predefined subarea boundary. The results of this extraction are a subarea network and associated trip tables for multiple users.

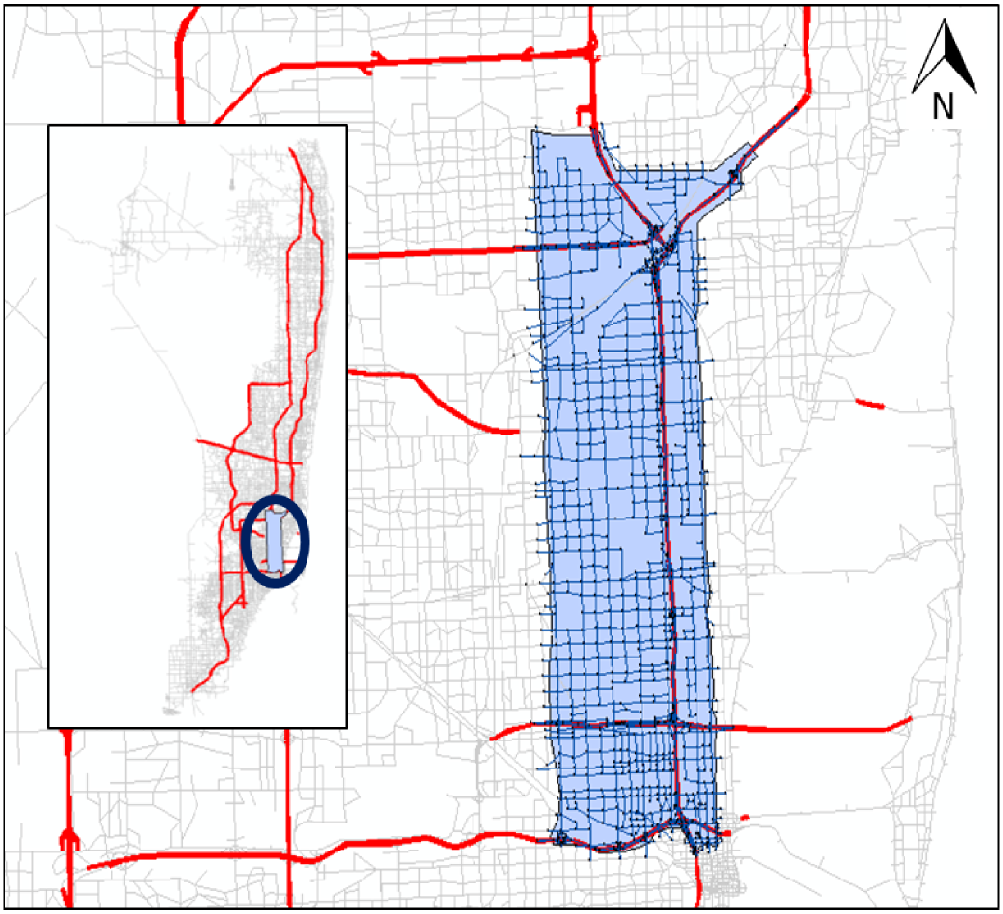


Figure 3-1 The extracted subarea from the SERPM model

The extracted subarea contains new node and zone numbers, which are different from the original numbers. The Cube stores the association between the old numbers (in the original network) and the numbers in the new network (in the subtracted network) in two new node features in the subtracted network. These two features are OLD_NODE and SUB_TYPE. The OLD_NODE attribute provides the old node number used in the whole network representation. The SUB_TYPE can have one of four values as explained below:

- Code 0 indicates that this is an existing node in both the whole model network and the extracted subarea network, and that the location of the node is the same in the two networks.
- Code 1 indicates an existing zone centroid in both the whole model network and the extracted network; location of the zone centroid does not change in the two networks.

- Code 2 indicates a zone centroid that does not exist in the whole model network, but is created in the subarea network as a new zone centroid at the subarea boundary to account for the demand that is entering or exiting the subarea network.
- Code 3 indicates a node that is connected to a new zone centroid created for the subarea network (i.e., connected to a zone centroid with SUB_TYPE "2").

When identifying the subarea boundary, the analyst must be careful with regard to crossing the links and connectors. As depicted in Figure 3-1 The extracted subarea from the SERPM model 3-2, subarea network extraction method A creates two new centroids (zone centroid number 2 and 7) that are connected to each other. This will increase the total number of zones and thus computational time, and also it will produce unrealistic assignment results, as the centroid connector that connects these two zones is the shortest path and vehicles between this OD pair will be assigned to this centroid connector instead of using physical links. Such problems can be avoided by using subarea network extraction method B, as shown in Figure 3-2.

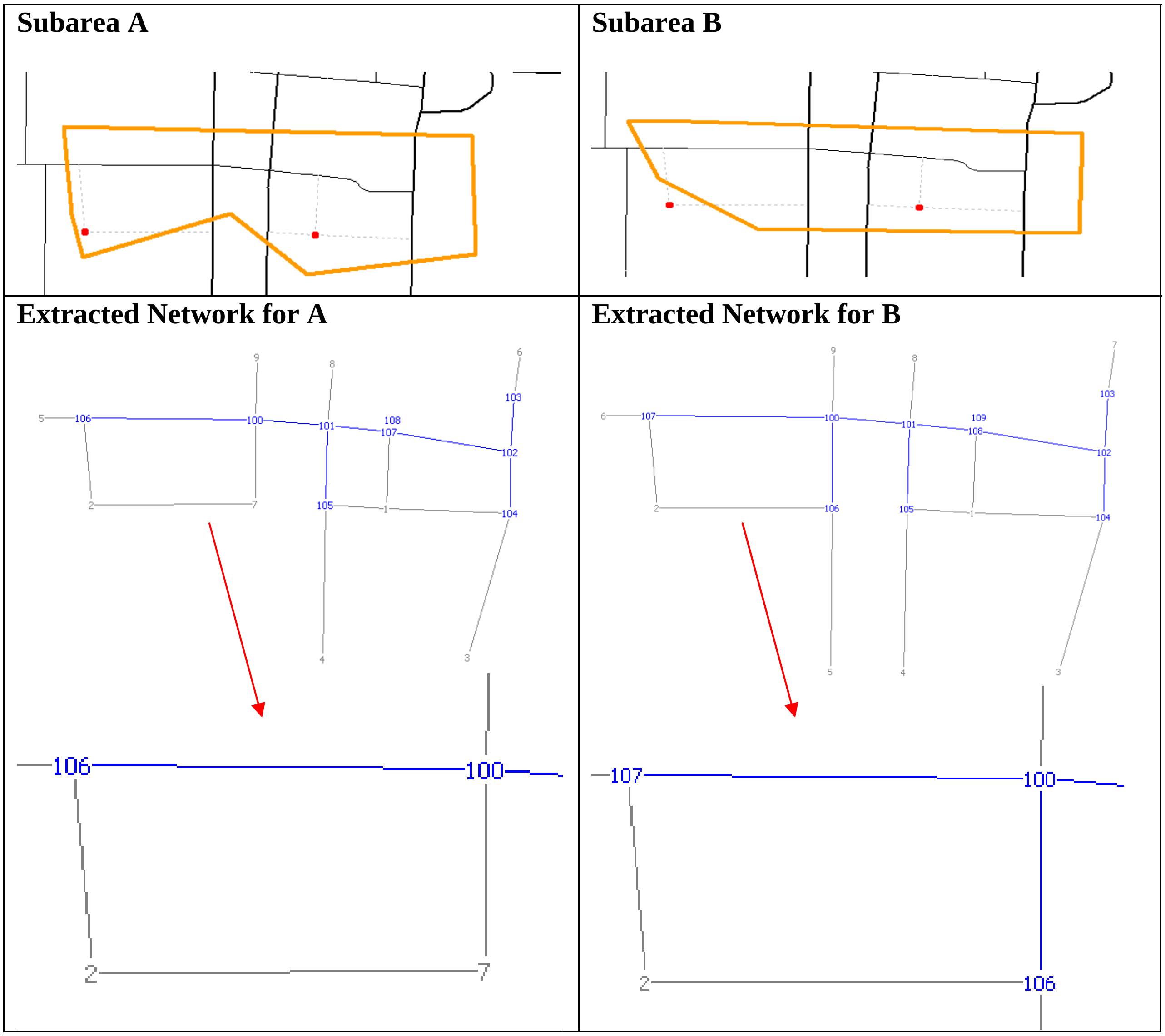


Figure 3-2 Potential problems in network extraction

3.1.2 Network Geometry Update

The modeled network geometry needs to be updated to better represent the existing real-world network since the details and accuracy of modeling the network in demand forecasting models are not sufficient for DTA applications. The attributes of each link were adjusted in this study based on the Google Earth map. The network geometry update was performed following the procedure presented in Figure 3-3 Network geometry and distance update procedure flow charts

3-3. The subarea network was converted into the KML format for the Google Earth application and the SHP file format for the ArcGIS application.

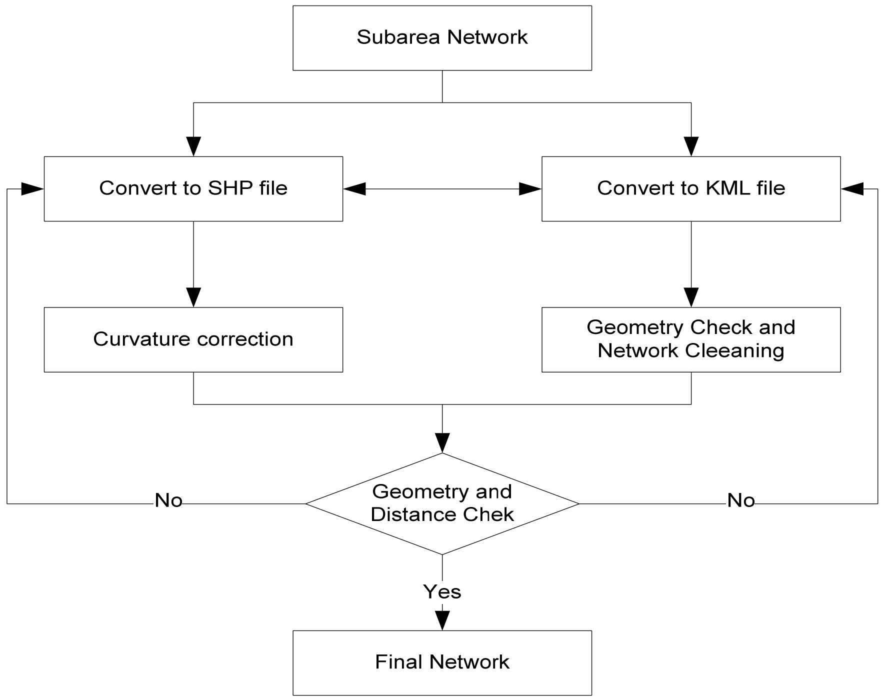


Figure 3-3 Network geometry and distance update procedure flow charts

By imposing the network on the Google Earth map, it was possible for the network curvature, connections, and other geometry attributes to be corrected. The link lengths were accordingly modified. The links in the original network file in the demand forecasting model were established based on direct node-to-node connections. Therefore, all of the links in the demand forecasting model are straight lines. In order to obtain the real-world curvature of the links, the network was converted into the SHP file format from the Cube network format, allowing the links curvature to be drawn based on the real curvature using the GIS modification tool. All of the links' lengths were updated based on the identified curvatures. Based on prior experience with the DTA tool used in this study (Cube Avenue), short links can produce unrealistic congestion. Therefore, it is very important to identify these links in the extracted subarea network and properly adjust their lengths to prevent the unrealistic congestion from occurring. It was found that in most cases, the issue of short links could be addressed by updating the links' lengths, considering the true curvature of the links, and moving the merge/diverge nodes based on their real-world location, based on Google Earth maps. Figure 3-4 shows an example of the network geometry adjustment conducted as part of this project.

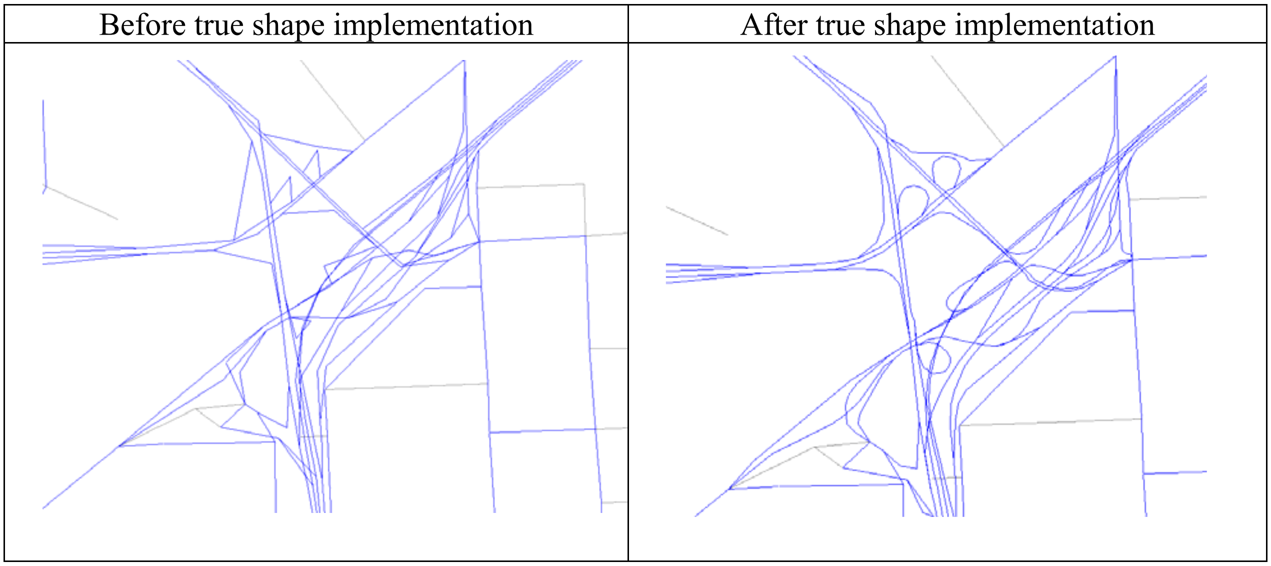


Figure 3-4 Network curvature correction

Further network cleaning and editing processes include adding details, moving and deleting nodes and links to avoid short links, as well as to reflect the current network. Modifying zones and connectors may also be necessary and should be considered. Additional examples of network cleaning processes are shown in Figure 3-5 and Figure 3-6. An important consideration in the cleaning process is also checking the consistency of the number of lanes between successive links, especially in the merge and diverge segments, and at intersections with exclusive left- and right-turning lanes. Network connectivity also needs to be checked to determine any issues with missing connections and link directionality errors.

Consideration should also be made at this stage of disaggregating trips from larger regional zones to the smaller ones and updating the zone connectors. The FHWA DTA Guidance report (Sloboden et al., 2012) discusses approaches to accomplish this disaggregation.

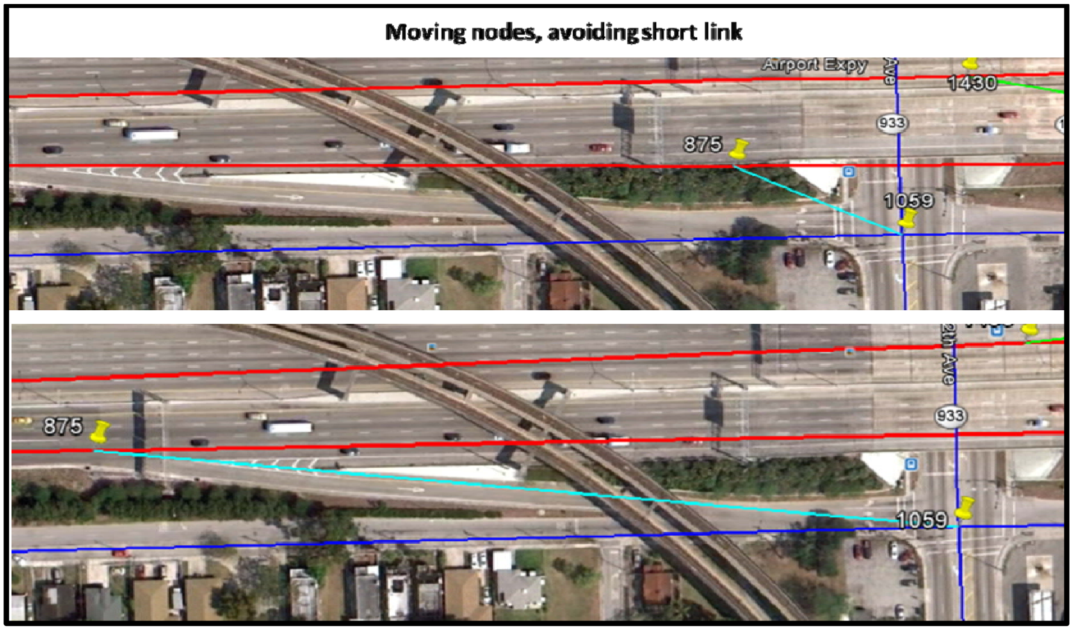


Figure 3-5 Moving nodes to avoid short links

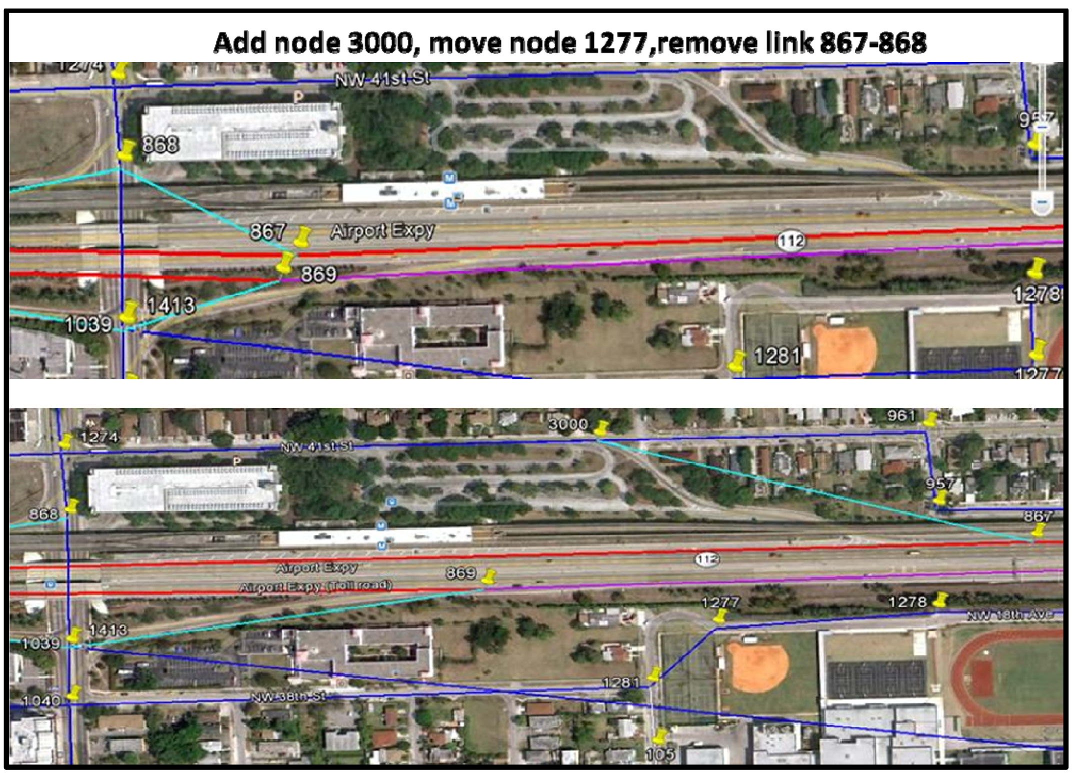


Figure 3-6 Adding and removing nodes

Most of the work conducted in this study was performed on a linear network, specifically the subarea shown in Figure 3-1 that represents the I-95 corridor facilities in the northbound direction, and which includes managed lane (ML), general purpose lanes (GPL), and associated on- and off-ramps. The main purpose of this study is to test the ability of the DTA platform proposed for use to model and develop methods to enable such testing. This allows for a more detailed validation of the utilized tools and methods, detailed exploration of the demand and supply calibration processes, and close observation of the assignment behavior. A linear network is further extracted that mainly contains the I-95 freeway, including general purpose lane (GPL) and parallel ML, and the associated on- and off-ramps. In most segments, the freeway includes four general purpose lanes and two managed lanes, which are separated from one another by a soft barrier. In this linear network, analysis of bottlenecks is more feasible, and demand estimation is more tractable. Further analysis of the full network is presented in Chapter 7. Some of the information about the linear network, full subarea, and the SERPM network is shown in Table 3-1.

Table 3-1 Network Information about Subarea and SERPM

Information	Subarea	Corridor
Zone	284	57
Node	1303	303
Link	3106	303
Drive Alone (DA)	234,067	79,182
Share Ride 2 (SR2)	48,417	16,316
Share Ride 3 Plus (SR3P)	23,178	11,368
Truck	27,855	10,675

3.1.3 Demand Data

In this study, initial trip matrices were extracted from the SERPM regional demand model, calibrated for the year 2005. Regional travel demand models represent an important source of origin-destination (OD) trip information since these trips are estimated through detailed and approved processes that ensure consistent behaviors of travelers in the demand generation, distribution, and mode choice steps. However, some issues with these models include lack of detailed model calibration at the subarea level and the potential changes in the network and demands since the model's last calibration. Even more critical to DTA modeling is that the regional demands are forecast for daily trips or three to four hours of time of day model period. These demands need to be distributed over better time intervals for DTA applications. The most common interval study for DTA modeling is 15 minutes.

3.2 Detector Data Acquisition and Preprocessing

Detector data collected from an existing intelligent transportation system (ITS) operated by the regional traffic management center was critical for the demand estimation, model calibration, and validation in this effort. The corridor of interest is instrumented every 0.3 to 0.5 mile with microwave detectors that report volume, speed, and density measurements in 20-second intervals for each lane. This data was supplemented by measurements from the Portable Traffic Monitoring Sites (PTMS) ramp counts from the Statistics Office of the Florida Department of Transportation (FDOT). The PTMS data include 15-minute ramp counts for two or three days per year. No speed or classification data are provided. Ramp counts obtained from the PTMS and ramp metering detectors represent the total origin and destination demand on the linear network and are very useful in the demand estimation process.

The ITS data were obtained from the Statewide Transportation Engineering Warehouse for Archived Regional Data (STEWARD). This system was developed as a proof-of-concept to centrally archive data from traffic management centers around Florida in a practical manner. The effort concentrated on archiving information from the SunGuide traffic sensor subsystem (TSS) and the travel time subsystem (TVT). The STEWARD database contains summaries of traffic volumes, speeds, occupancies, and travel times aggregated by 5-, 15-, and 60-minute periods, as requested by the user. Using a Web-based interface, the user can specify date and time ranges and detector locations for which the data are needed. The user can also download all generated reports in comma-delimited formats, which can be easily imported into database management tools. Table 3-2 lists the numbers of the available ITS and PTMS detectors in the subarea network and the selected detectors after removing redundant or erroneous detectors.

Table 3-2 Available and Selected Detectors

Detector Station	No. of Stations	
	Available	Selected
ITS	**109**	**87**
• General- Purpose Lane	78	56
• Express Lane	31	31
PTMS	**150**	**150**
• Mainline	10	10
• Ramps	99	99
• Arterial	41	41

After imposing the network and detector maps onto Google Earth's map, it was possible to manually associate the detectors in Table 3-2 with the network links, as depicted in Figure 3-5 Moving nodes to avoid short links

. Based on several criteria, if any link is associated with more than one detector, only the most reliable one was kept. Guidelines on removing redundant or erroneous detectors are provided later in this section.

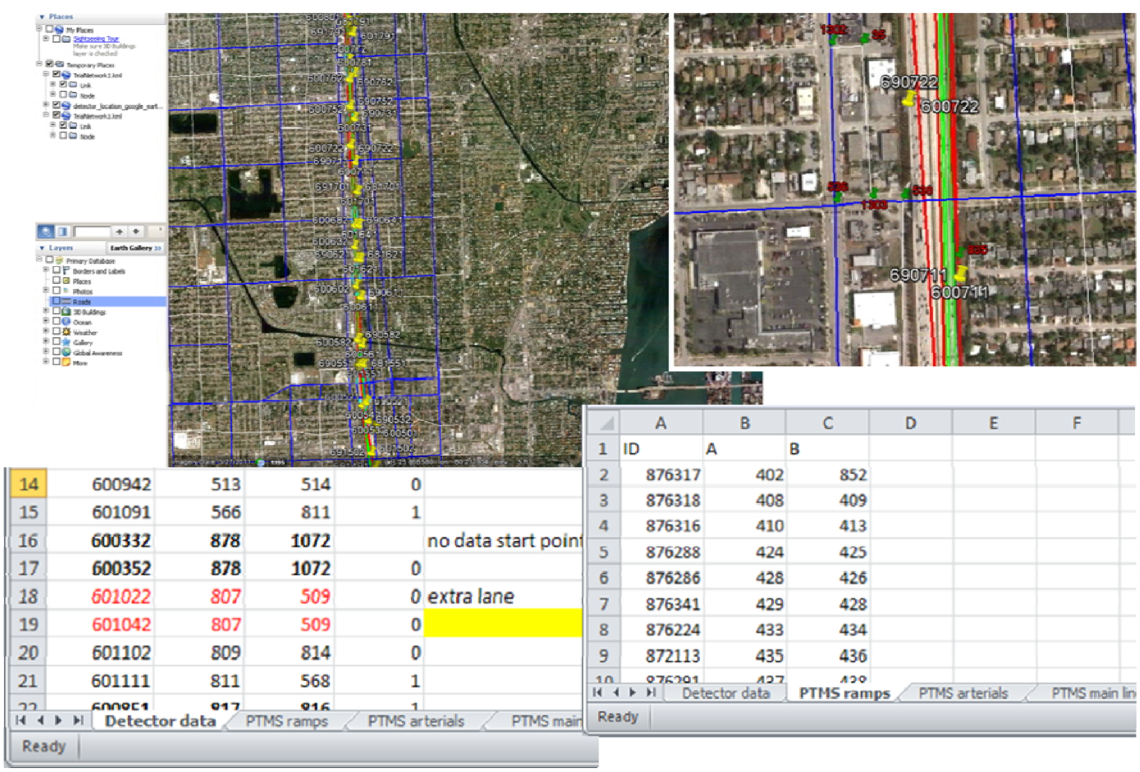


Figure 3-7 Associating detectors with links

Truck percentages are available for on-ramps from PTMS data. For the mainline, the truck percentage was obtained from nearby permanent Telemetered Traffic Monitoring Sites (TTMS), also operated by the FDOT Statistics Office. These percentages were confirmed based on the manual counting of recorded videos at selected corridor locations.

3.3 Intersection Control and Geometry

Signal information at each signalized intersection is obtained from the Miami-Dade County Traffic Management Center and is manually coded into the network. Information such as cycle length, green time, and split can be coded into the Cube assignment tool (Cube Avenue). However, due to problems with the use of the tool for managed lane operations, it was decided that the DTA implementation should be tested without consideration of this control so as to focus on resolving the abovementioned problem. It is realized that in real-world applications, the signal control should be an important component in the DTA modeling process.

Checking the intersection geometry to ensure that the coding and connectivity are correct and contain a suitable level of detail can be a time-consuming effort, particularly for large networks. Turn restrictions at each intersection should also be carefully checked.

An example of signal control files obtained from Miami-Dade County is presented in Figure 3-5 Moving nodes to avoid short links

. Figure 3-9 shows how this information is modeled and coded in the Intersection Data Editor of the Cube software. Depending on the density of signalized intersections in the network, collecting and coding their signal timing can be quite time-consuming. If time or budgetary limitations exist, it is possible to model all of the intersections as actuated signals in a number of DTA tools, including Cube Avenue, which allows the model to estimate the timing based on demands. In this case, the user needs to provide only the cycle length, and minimum and maximum green times. The signal timings are calculated by the assignment tool, based on the assigned demands.

Current TOD Schedule	Plan	Cycle	Green Time 1 NBL	2 SBT	3 EBL	4 WBT	5 SBL	6 NBT	7 WBL	8 EBT	Ring Offset	Offset
	1	110	7	42	20	25	7	42	20	25	0	63
	2	90	7	38	7	22	7	38	7	22	0	45
	3	90	6	40	6	22	6	40	6	22	0	61
	4	60	5	20	5	14	5	20	5	14	0	21
	5	75	6	27	6	20	6	27	6	20	0	14
	6	85	6	36	6	21	6	36	6	21	0	7
	7	85	6	36	6	21	6	36	6	21	0	14
	8	110	7	55	7	25	7	55	7	25	0	26
	9	110	7	55	7	25	7	55	7	25	0	26
	10	90	6	42	6	20	6	42	6	20	0	7
	11	60	5	19	5	15	5	19	5	15	0	4
	16	120	6	67	6	25	6	67	6	25	0	105
	18	85	6	33	19	11	6	33	19	11	0	15

Local TOD Schedule

Time	Plan	DOW
0000	4	Su M T W Th F S
0030	Free	M T W Th F
0100	Free	Su S
0130	Free	M T W Th F
0230	Free	Su S
0600	4	Su S
0600	5	M T W Th F
0700	2	M T W Th F
0730	1	M T W Th F
0900	6	M T W Th F
1000	6	Su S
1330	18	M T W Th F
1430	6	W
1530	3	M T W Th F
1800	6	M T W Th F
1840	7	M T W Th F
2030	4	Su M T W Th F S

Splits

PH 1	PH 2	PH 3	PH 4	PH 5	PH 6	PH 7	PH 8
NBL	SBT	EBL	WBT	SBL	NBT	WBL	EBT
0	0	0	0	0	0	0	0

Active Phase Bank: Phase Bank 1

Phase		Walk (Phase Bank 1 2 3)	Don't Walk (1 2 3)	Min Initial (1 2 3)	Veh Ext (1 2 3)	Max Limit (1 2 3)	Max 2 (1 2 3)	Yellow	Red
1	NBL	0 - 0 - 0	0 - 0 - 0	5 - 5 - 5	2 - 2 - 2	7 - 7 - 7	10 - 10 - 10	3	0
2	SBT	7 - 0 - 7	11 - 0 - 11	7 - 16 - 7	1 - 1 - 1	30 - 30 - 30	0 - 40 - 40	4	0.8
3	EBL	0 - 0 - 0	0 - 0 - 0	5 - 5 - 5	2 - 2 - 2	7 - 7 - 7	10 - 10 - 10	3	0
4	WBT	7 - 0 - 7	12 - 0 - 12	7 - 7 - 7	2.5 - 2.5 - 2.5	20 - 20 - 20	35 - 35 - 35	4	0.9
5	SBL	0 - 0 - 0	0 - 0 - 0	5 - 5 - 5	2 - 2 - 2	7 - 7 - 7	10 - 10 - 10	3	0
6	NBT	7 - 0 - 7	11 - 0 - 11	7 - 16 - 7	1 - 1 - 1	30 - 30 - 30	0 - 40 - 40	4	0.8
7	WBL	0 - 0 - 0	0 - 0 - 0	5 - 5 - 5	2 - 2 - 2	7 - 7 - 7	10 - 10 - 10	3	0
8	EBT	7 - 0 - 7	12 - 0 - 12	7 - 7 - 7	2.5 - 2.5 - 2.5	20 - 20 - 20	35 - 35 - 35	4	0.9

Figure 3-8 Example of the signal control data obtained from Miami-Dade County

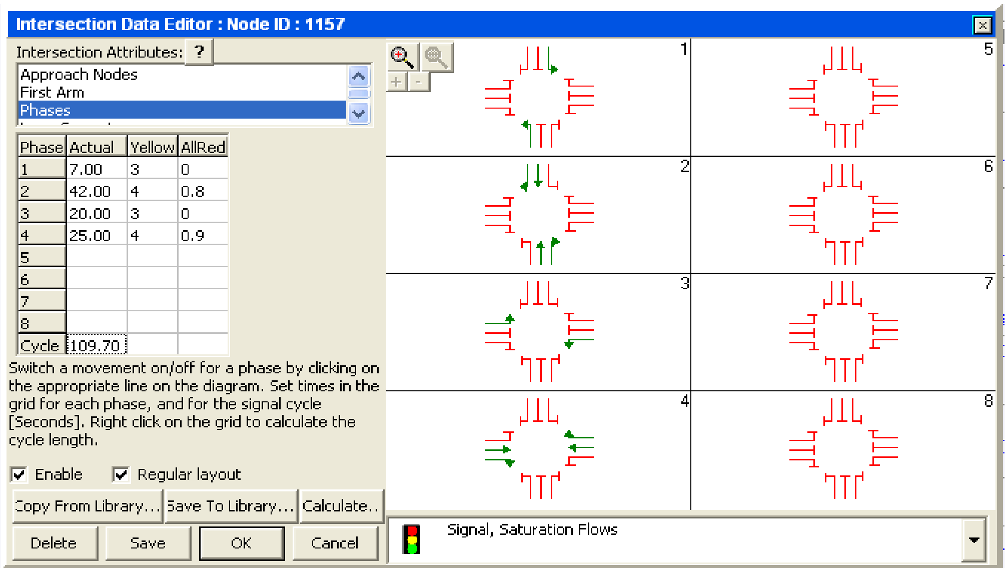


Figure 3-9 Coding signal control in the Cube tool

3.4 Other Data Sources

Available data from other sources were also obtained and used in this study, as listed below:

- A previously calibrated micro-simulation model of the study area that includes traffic demand estimates.
- Real-world toll values for each 15-minute interval from FDOT District 6 Traffic Management Center.
- Ramp metering data from FDOT D6 Traffic Management Center.
- URS calibrated logit model for willingness-to-pay prediction along I-95 corridor.
- Updated toll algorithm from FDOT D6 Traffic Management Center.

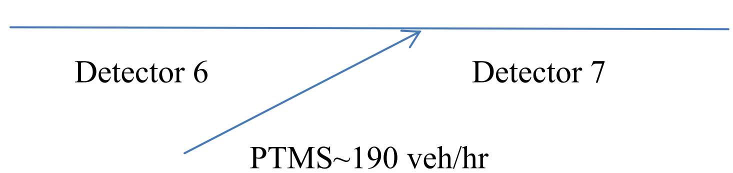
3.5 Data Preprocessing and Validation

This section provides additional details and lessons learned with regard to the data processing and validation processes.

Demand and congestion patterns vary greatly day by day. The representative days for modeling and calibration are considered weekdays (Tuesday through Thursday) without incidents or abnormal external conditions such as heavy rain. Non-representative days can be filtered out by different methods that exclude days with special events or conditions. Also, data mining methods can exclude days with significantly different volumes or speed patterns from normal days.

Between May 2010 and May 2011, 16 days were identified as ideal days to represent normal day traffic based on detector data. Among these days, the speeds vary with a coefficient of variance between 5% to 20% for different detector locations, and the volumes vary with a coefficient variance between 3% to 7% for different locations. For different purposes, a specific day or an average of all repetitive days may be used for calibration. Using the median day data may be better than using the averages, since the averages do not represent any of the real-world days. This issue will be further discussed in Chapter 4.

Inconsistencies between consecutive detector counts should be a major consideration. Sometimes it is not enough to compare just one pair of detectors, and there is a need to check several stations upstream and downstream of each location. The addition of on-ramps and subtraction of off-ramps to estimate the expected volume for the station can be used as a reference when assessing the accuracy of the measurements. In the presence of queue, this procedure becomes more complicated, and the capacity constraints should be considered. Figure 3-10 is an example of two successive detectors with an on-ramp between them, with approximately 190 vehicles per 15-minute intervals. The upstream and downstream detectors, however, show the exact number of counts. It should be noted that the reported counts are below capacity at all times, therefore, this issue is not caused by capacity restrictions. The comparison of detectors with additional upstream and downstream detectors disclosed that the detector located downstream (Detector 7) is not reliable.



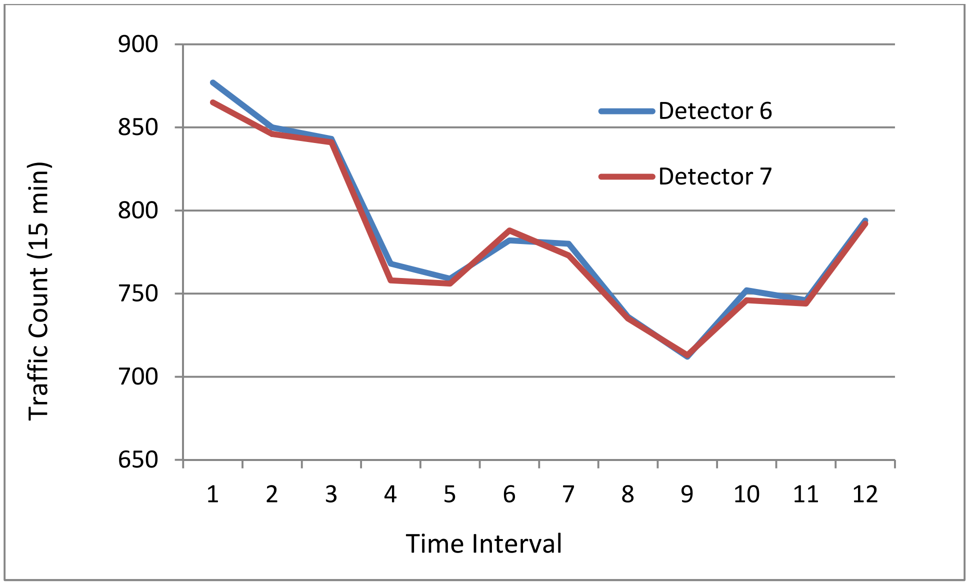


Figure 3-10 Volume inconsistency between successive detectors

Figure 3-11 shows another example of volume inconsistency between successive detectors. The detector numbers in this figure (e.g., 600891) are the original numbers from the STEWARD database. Selecting the right detector for each segment was only possible by having benchmark, reliable detectors upstream and downstream of the segment, and selecting the most reliable detectors by calculating the volumes from several upstream/downstream detectors, as previously discussed.

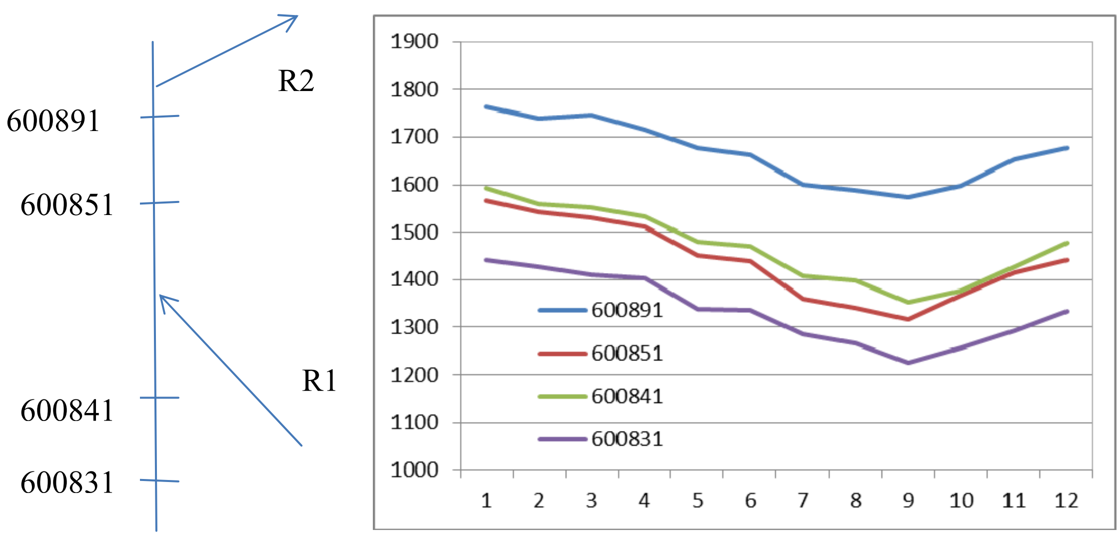


Figure 3-11 Volume inconsistency between successive detectors

In addition, the number of lanes that the detector covers, which is included as an attribute in the detector database, should be checked, because some detector counts include mainline and merge/diverge volume, which needs to be better understood when using ITS data for modeling. The consistency between detector counts and link capacity should also be checked to ensure that the reported count is below the segment capacity, due to detector errors.

Whenever data is available, comparing PTMS and ITS count data may improve the reliability of the data. In this study, it was found that there was an acceptable match between the used ITS and PTMS counts for the ramps. The PM peak on the mainline showed that the PTMS reported higher volumes, compared to ITS data. Manual counts of recorded videos were conducted to determine the accuracy of the data. It was found that the manual counts are closer to ITS data than the PTMS during the PM peak. Figures 3-12 and 3-13 show the comparison of PTMS and ITS volume data for three days: August 9, August 10 and August 11, 2011.

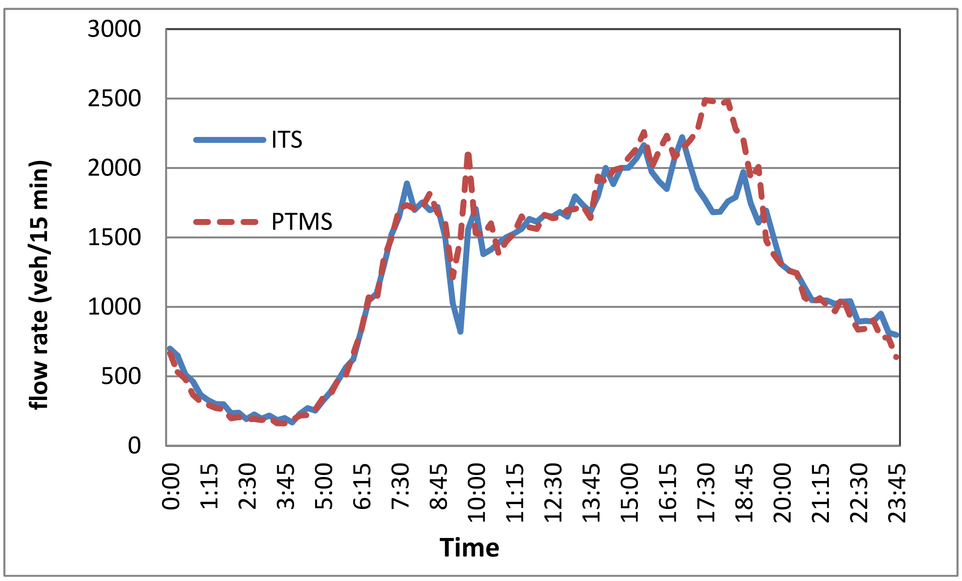


Figure 3-12 Comparison of PTMS vs ITS volume data (August 9, 2011)

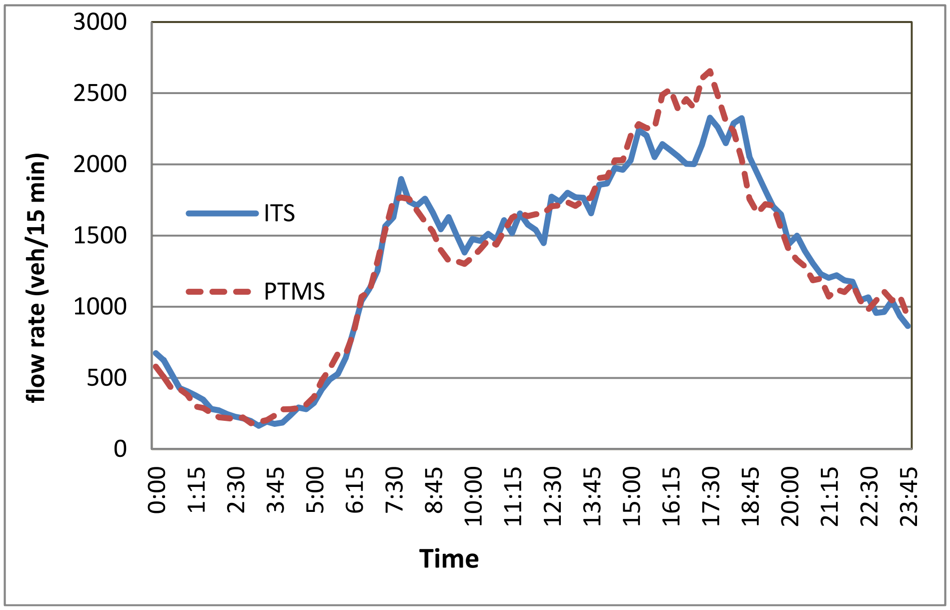


Figure 3-13 Comparison of PTMS vs ITS volume data (August 10, 2011)

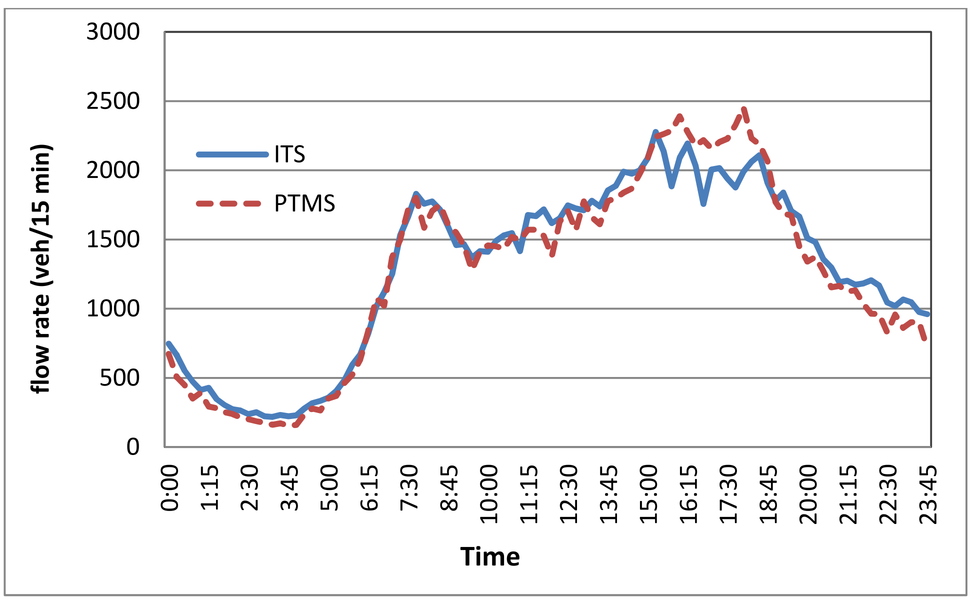


Figure 3-14 Comparison of PTMS vs ITS volume data (August 11, 2011)

ITS data normally do not include detectors for the on- and off-ramp locations, unless ramp metering exists. As previously stated, the study section includes ramp metering and thus, ramp detectors. In terms of ramp metering, there are three types of detectors: upstream (queue) detectors that measure the demand, and downstream (arrival and departure) detectors before and after ramp signals. When ramps are equipped with ramp metering, the modeler should decide which information is used: either the upstream demand, or the volume that passes through the ramp metering. For OD estimation purposes, the former should be used.

Detailed examination of the ITS data may help identify the reason for the congestion, so as to assist in the calibration process. Figure 3-15 and Figure 3-16 show lane by lane data of speed and occupancy for one detector at a congested location. This detector location was initially defined as an active bottleneck for potential capacity measurement. Lane by lane data of speed and occupancy, however, revealed that the congestion at this location is caused by a spillback from a downstream off-ramp. Therefore, the two left lanes have considerably lower speeds and higher occupancy than the other lanes, indicating that this location is not a candidate for use in estimating capacity.

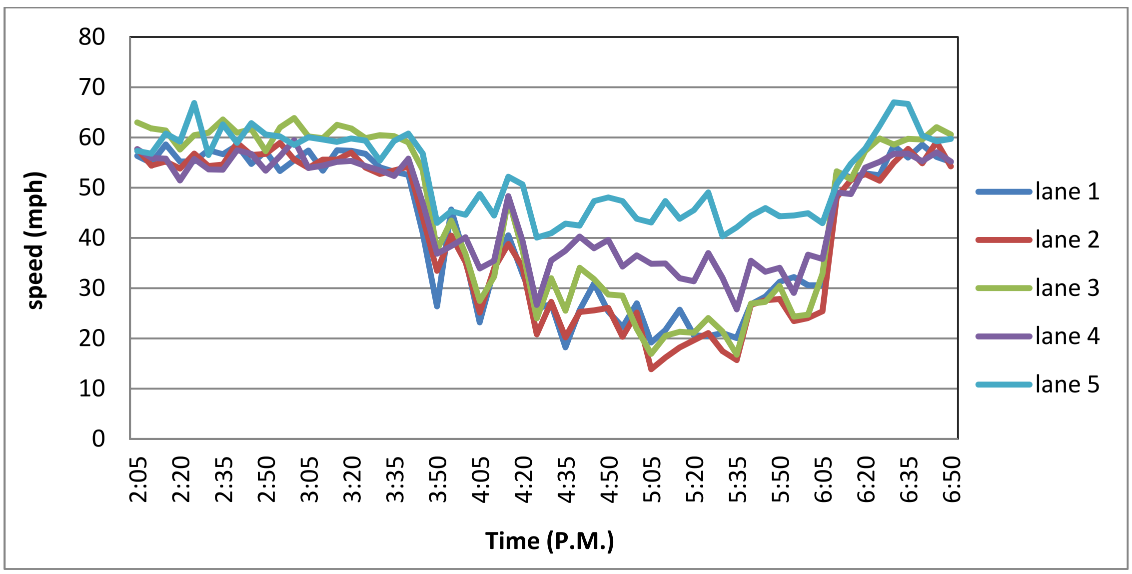


Figure 3-15 Lane by lane speed data

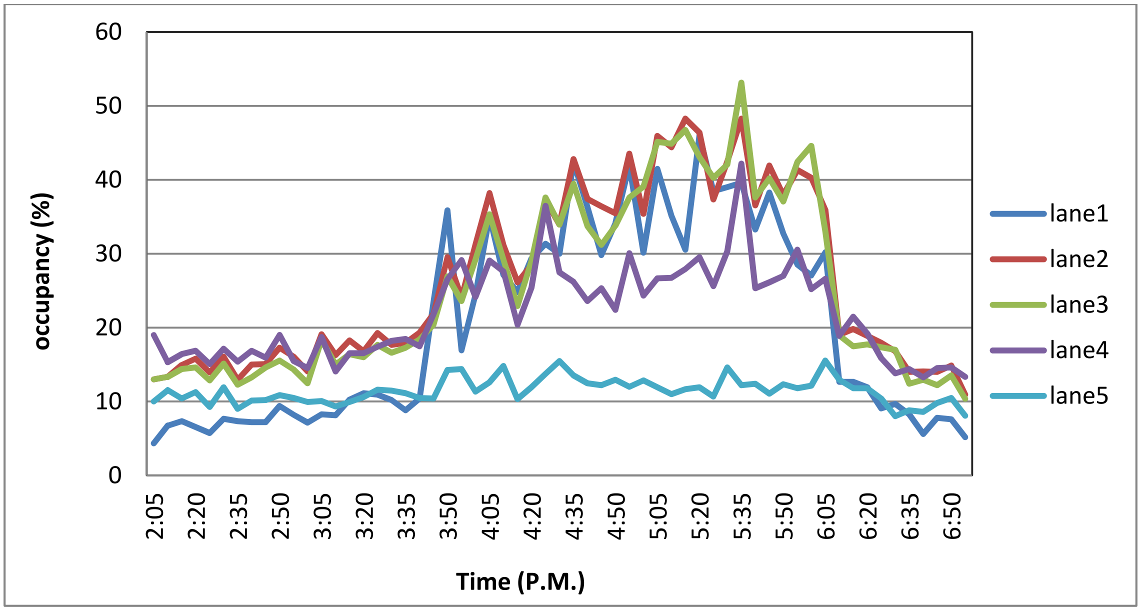


Figure 3-16 Lane by lane occupancy data

During the OD estimation process (estimating trip tables in a way that when assigned to the network produce volumes that are close to the observed link volume), it was found that the detector for one of the screenlines does not produce the correct volumes. This link was associated with a high volume close to capacity, with a speed close to free-flow speed, without any flow breakdown. A simulation model of the study area was also under development at the time and did not support the reported volumes by this detector. It should be noted that this value passed through all abovementioned filtering processes. Utilizing this screen data significantly affected the OD estimation process. This example shows that comparing the data from multiple sources of information should be a continuous and iterative process throughout the modeling and calibration tasks.

Depending on the network under consideration, there may be a need to disaggregate the zones from larger zones used in the regional model to smaller zones. There may also be a need to modify the zone connector setting. Careful examination is needed of how the zones and their connection setup affect the results of the modeling.

4. Modeling Managed Lane in Cube Avenue

4.1 Introduction

The managed lane (ML) modeling process was implemented and evaluated in this study using two different approaches: 1) Managed lane costs in the objective function, which is an approach traditionally applied in toll modeling and that has been utilized with dynamic traffic assignment (DTA) modeling of toll facilities and managed lanes and 2) Utilizing a willingness-to-pay curve in conjunction with the DTA, which is the approach recommended in the Florida Department of Transportation (FDOT) Phase 1 managed lane modeling process based on static traffic assignment. The research team worked with Citilabs, a research team member and the developer of the Cube Avenue, which is the tool used in this study to implement a prototype that utilizes the second approach mentioned above. The involvement of Citilabs in developing and testing this prototype was significant since this implementation was found to go beyond the existing capabilities of Cube Avenue and required many modifications, updates, and corrections in the Cube Avenue modules.

Citilabs developed a prototype managed lane model based on the DTA model that was adopted from the FDOT Phase 1 managed lane modeling process based on static traffic assignment (see aforementioned second approach). The prototype model contains a toll diversion process, as well as the congestion-based (dynamic) tolling process, so that it estimates the toll trips and the toll costs for each time segment in the managed lanes. The I-95 managed lane facility in Miami, Florida, was implemented, not only to test the model, but also to review the results estimated by the model. This chapter includes the following information, which is pertinent to understanding this research effort:

- General information on the Cube Avenue program
- Procedure of proposed prototype model
- Toll diversion process
- Dynamic tolling process
- Scripting in the Cube Avenue model

- Major input and output files
- Required network attributes and model settings
- Output results
- Dynamic origin-destination matrix estimation (DODME)

4.2 General Information of Cube Avenue Program

The prototype Cube Avenue model was implemented based on the DTA approach used in Cube Avenue. Cube Avenue loads and tracks the movement of vehicle packets throughout the roadway network. Vehicle packets can be of any size, from an individual vehicle up to platoons of 20 or more vehicles. However, only one vehicle per packet is recommended in the prototype managed lane model so that each vehicle has an option to choose either a free road or a toll road.

Through an iterative process, Cube Avenue aims to achieve dynamic user equilibrium network demands. Vehicle packets move, stop, and queue to upstream roads and intersections. Based on vehicle departure time, Cube Avenue computes the most inexpensive path for each vehicle unit, and computes interactions among vehicle units as they travel through the network.

Initially, the link and junction times are the same for all time segments. However, once the modeling process begins, the segment-by-segment times are recalculated independently. Hence, on the second and subsequent iterations, there will be different estimates of travel time for vehicles arriving on a link (or at a stop-line) for each time segment.

4.3 Procedure of Proposed Prototype Model

The prototype toll managed lane model was developed using the Cube Avenue program, which can perform the DTA for each time segment. Figure 4-1 shows the overall procedure in the Cube catalog, including the three major required processes: the highway network (supply) process, the origin-destination (OD) table (demand) process, and the toll managed lane modeling process.

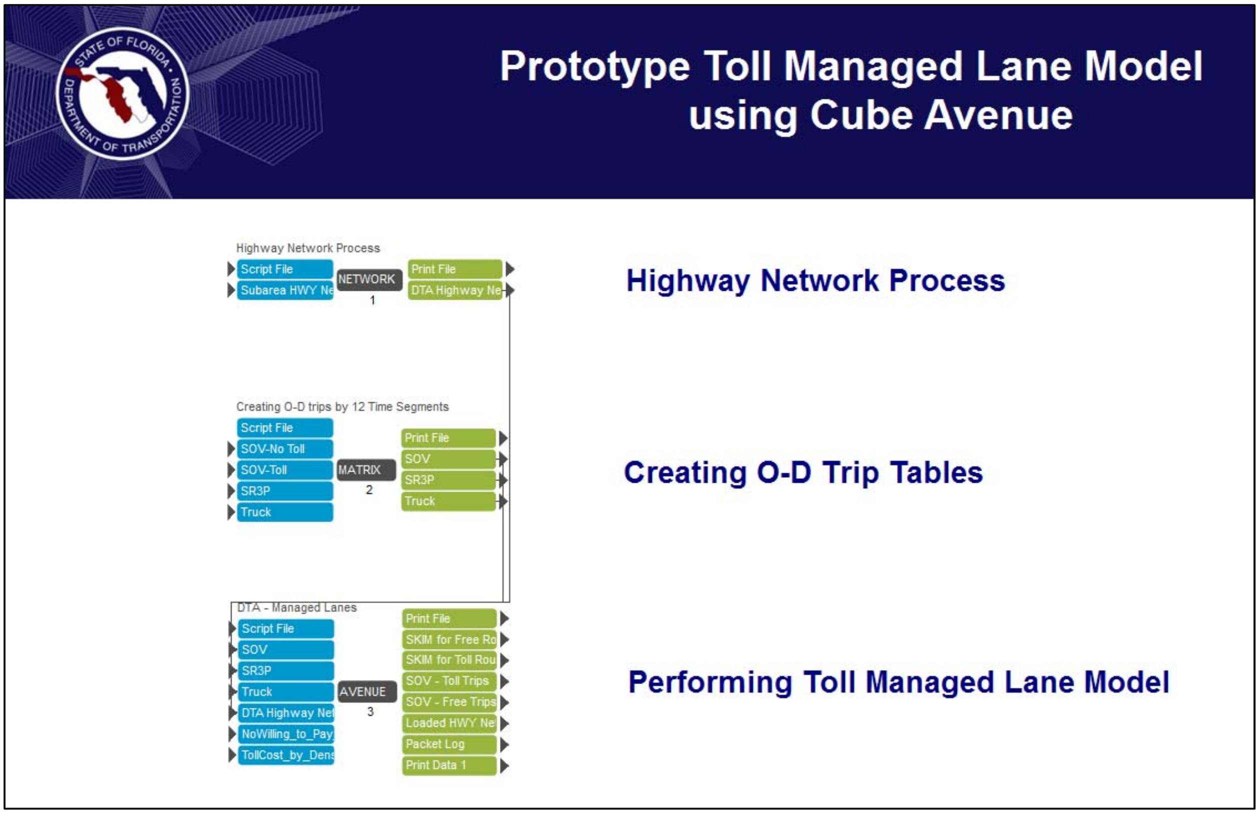


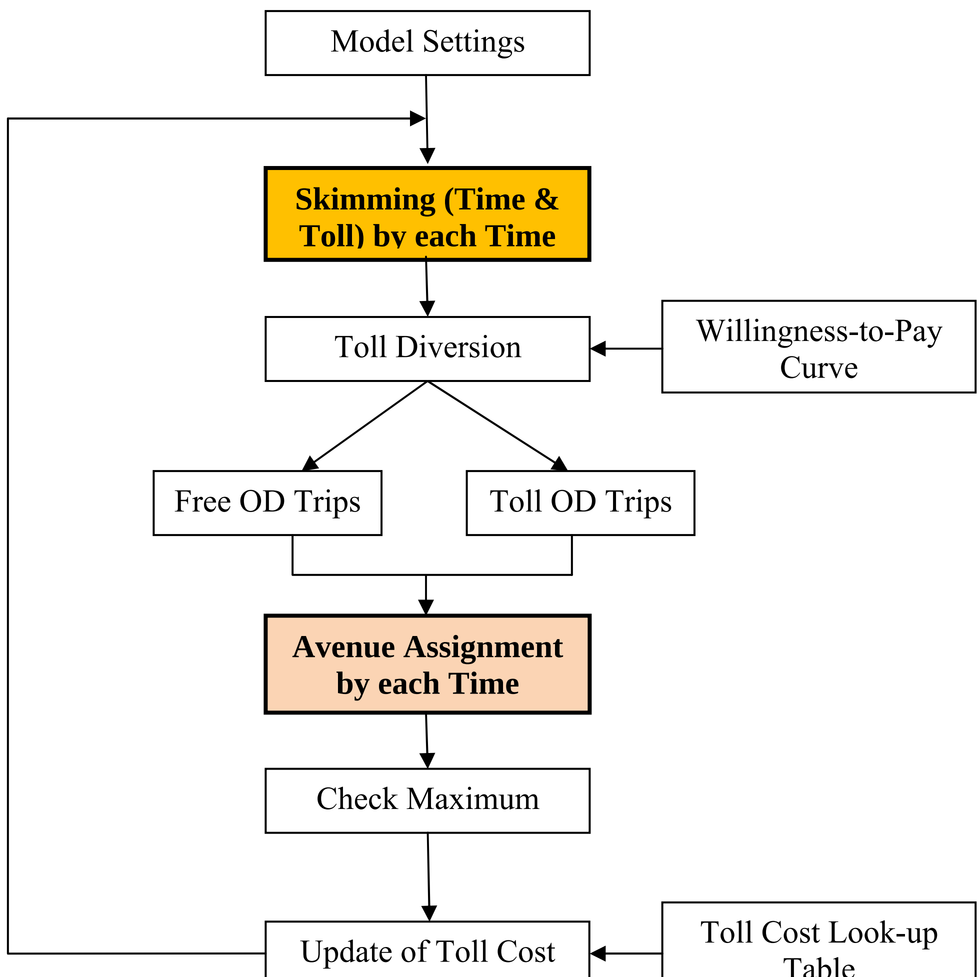
Figure 4-1 Process of Cube-based managed lane model

First, the highway network process sets up the required link attributes, such as speed (mph), free-flow travel time (min), capacity for the modeling period, storage, parameters for volume-delay functions (VDF), toll links, high occupancy toll (HOT) links, and initial toll costs in the HOT entry gates.

Next, the OD trip process reconstructs the OD trip matrices by adding two-person high occupancy vehicle (HOV2P) trips into the drive-alone, or single occupancy vehicle (SOV) trips because the managed lane implementation in Miami charges the same toll for SOV plus HOV2P trips for the use of managed lanes. Note that it is assumed that both SOV and HOV2P vehicle trips can use the HOT lanes by paying the tolls in the I-95 corridor if it improves generalized costs (disutility in the assignment). Three-person or more high occupancy vehicles (HOV3+) are allowed to use the managed lanes for free, as is the case with the I-95 express lanes in Miami. The input OD trips by each trip mode (e.g., SOV, HOV2P, HOV3P+, and truck) are prepared for

12 time segments, with a duration of 15 minutes per time segment. More details about the estimation of demand trips are discussed in Chapter 6.

The last step of the prototype managed lane model is to estimate the vehicle trips that use the managed lanes for each time segment considering the tolls, in addition to updating the tolls based on the congestion level. The estimated toll, toll-free vehicle trips, and tolled trips are output for each OD pair to be reviewed during the model validation process. One of the major outputs, the loaded network, contains useful, varied estimation results at the link level that will be described later on this report. The travel activity for each packet is generated in the packet log file, which can also be animated using the loaded highway network in the Cube Base program.



The detailed procedure for the prototype toll managed lane is illustrated in Figure 4-2.

Figure 4-2 Prototype toll managed lane model based on willingness-to-pay curve

First, as suggested in Figure 4-2, all of the model parameter values are defined, in addition to setting up the highway network attributes to run the Cube Avenue program. Next, the path-building process is performed to obtain the impedance (skimming) values, such as the travel time (min) and cost ($) for each origin-destination pair. These skimming values are used to compute the toll in cents for the time saved between free and toll routes from origin to destination. Then, the toll shares are obtained from the willingness-to-pay curve lookup table. Once the toll diversion process is implemented to estimate the free vs. toll trips, the dynamic traffic assignment loads each packet into the highway network by each traffic mode (e.g., free trips, toll trips, HOV3P+, and truck trips) for each time segment. Finally, the model determines the maximum density for each directional corridor in the managed lane toll facility. The toll cost per time segment is updated based on the density lookup table. The estimated toll-trips based on the lookup table mentioned above are allowed to divert to general purpose lanes, if this improves their travel time (or generalized cost).

4.4 Toll Diversion Process

Initially, the prototype managed lane model uses the toll diversion model implemented by the willingness-to-pay curve, as shown in Table 4-1, as a test sample. This information is can be found in the FDOT report titled "Managed Lane Modeling Application for FSUTMS (Phase I)." The proportions listed in the table are percentage units (%) that use the free roads to avoid the toll payments. The demand category can be set by traffic modes or traffic patterns. These setting values were modified during the validation of the model in order to be suitable for the I-95 HOT facility in Miami, as further discussed in Chapter 7.

Table 4-1 Initial Not-Willing-to-Pay Proportion for Cost per Time Saved by Demand Category

Toll Cents per Minute saved	Demand Category							
	1	2	3	4	5	6	7	8
0.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
8.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0	50.0
10.0	60.0	60.0	60.0	60.0	60.0	60.0	60.0	60.0
16.3	75.0	75.0	75.0	75.0	75.0	75.0	75.0	75.0
20.0	81.7	81.7	81.7	81.7	81.7	81.7	81.7	81.7
23.7	85.0	85.0	85.0	85.0	85.0	85.0	85.0	85.0
31.4	90.5	90.5	90.5	90.5	90.5	90.5	90.5	90.5
41.7	95.0	95.0	95.0	95.0	95.0	95.0	95.0	95.0
51.8	96.0	96.0	96.0	96.0	96.0	96.0	96.0	96.0
58.3	98.0	98.0	98.0	98.0	98.0	98.0	98.0	98.0
66.7	98.8	98.8	98.8	98.8	98.8	98.8	98.8	98.8

The toll in cents per minute saved can be computed by dividing toll cost (in cents) by the saved travel time (in min) for each OD pair, as follows:

$$\text{Toll Cents per Minute Saved} = \frac{\text{Total toll cost (cents) for toll route}}{\text{Free route time (min)} - \text{Toll route time (min)}} \tag{4-1}$$

For certain OD pairs, a vehicle driver can travel to the destination using only free roads. The driver, however, has another option, which is to use the toll roads by paying the tolls in the managed lanes. In that case, the toll in cents per minute saved can be computed by comparing the route travel times for the two options (free vs. toll). Next, the toll trip share (%) can be obtained by looking up the willingness-to-pay curve table (see example in Table 4-1). For example, suppose that a driver can travel the free road in 25 minutes, while the driver also has an opportunity to use the toll road with a travel time of 20.78 minutes by paying $1 as a toll cost. In this case, the toll cents per minute saved is 23.7 cents per minute (=100 cents/(25-20.78)). Thus, the vehicle's probability of using the free road is 85%, based on the willingness-to-pay table, while the probability of using the toll road is estimated at 15% (=100%-85%).

The toll diversion logit model (example is shown below) is an alternative to estimating the toll trip share. It can also be utilized to estimate the toll trip proportion (%) for each OD pair. The coefficient values can be calibrated using the stated preference survey data.

$$P_{toll} = \frac{1.0}{1.0 + e^{[\alpha(T_{toll} - T_{free}) + \beta(C_{toll})]}} \times 100\% \tag{4-2}$$

Where,

P_{toll}	=	toll trip proportion (%) for toll route
T_{toll}	=	travel time (min) for toll route
T_{free}	=	travel time (min) for free route
C_{toll}	=	total toll cost ($) for toll route
α	=	coefficient for time
β	=	coefficient for toll cost

4.5 Dynamic Tolling Process

The toll cost per time segment (e.g., 15 minutes) is updated dynamically based on the largest density in each toll facility corridor. The managed lane model computes the link density (vehicles per mile per lane) for each time segment by dividing the hourly assigned volumes per lane by the link speed (mph) as follows:

$$Link\ Density = \frac{Houly\ Link\ Volumes\ \div Lanes}{Link\ Speed\ (mph)} \tag{4-3}$$

The model identifies the largest density by comparing the link densities for all of the links in each directional corridor of the toll road. Once the largest density is found for each directional corridor at the end of each time segment, the toll cost ($) is obtained from the input lookup table listed in Table 4-2. Next, these updated toll costs are stored using the link work variable (e.g., LW.TOLLCOST1) into the entry gate links in order to be used in the next iteration.

Table 4-2 Dynamic Toll Cost ($) by Road Density (vehicle per mile per lane)

LOS	Road Density		Toll Cost ($)	
	Minimum	Maximum	Minimum	Maximum
A	0	11	$0.25	$0.25
B	12	18	$0.50	$1.25
C	19	26	$1.50	$2.75
D	27	35	$3.00	$3.75
E	36	45	$3.75	$6.00
F	>45		$6.00	$7.00

The dynamic tolling process can be controlled in various ways, based on two different Cube Avenue program sets. One setting can run the model for all time segments (for each iteration), while the other setting can run the model iteratively for each time segment. The first method is similar to an iterative process in the static highway assignment that searches the optimal solution at the end of iterations, but the second method searches the optimal solution consequently by each time segment. For example, suppose that the model containing three time segments is performed for two iterations. Figure 4-3 shows that all time segments are run for each iteration and that the toll cost for each time segment is also updated after running the first iteration. The following second iteration uses these updated toll costs for all time segments.

ITERLOADIN C = 0

< Iteration: 1 >

Time Segment	Initial Toll	Road Density	Updated Toll
1	$0.25	22	$2.04
2	$0.25	18	$1.25
3	$0.25	26	$2.75

< Iteration: 2 >

Time Segment	Toll from Iter# 1	Road Density	Updated Toll
1	$2.04	20	$1.68
2	$1.25	27	$3.00
3	$2.75	24	$2.39

Figure 4-3 Update of toll costs by an iteration-by-iteration basis

The alternative is to run the model and check the performance on a time segment-by-time segment basis using the ITERLOADINC keyword in the Cube Avenue program. This keyword allows a partial convergence of the dynamic loading assignment for each time segment. The user can specify the number of iterations in ITERLOADINC so that the Cube Avenue program can perform the assignment, consequently, for the time segments. That is, it performs the dynamic loading process iteratively for one time segment before the subsequent time segments are loaded. In Figure 4-4, if ITERLOADINC is 2, every time segment runs the assignment for 2 iterations, but one additional iteration would be run because the maximum number of iterations is 3. Note that the number of iterations in each time segment should not exceed the maximum number of iterations specified by the MAXITERS keyword.

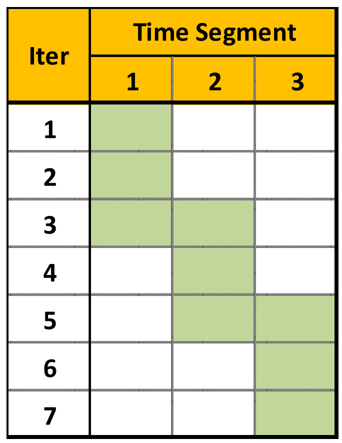


Figure 4-4 Incremental time segment loading (ITERLOADINC=2 & MAXITERS=3)

The total number of iterations is computed by (number of time segments – 1) * ITERLOADINC + MAXITERS. At the very minimum, (number of time segments – 1) * ITERLOADINC + 1 iterations are performed to allow every time segment to be loaded and simulated at least once. After the maximum number of iterations has been performed for each time segment, the simulation results for that time segment are saved.

Figure 4-5 shows how the toll costs are updated when specifying the ITERLOADINC keyword. In the first time segment, the loading simulation is performed up to 3 iterations unless the convergence is reached. The toll costs are updated every iteration after the initial toll cost in the first iteration. Once the first time segment is complete, the final toll cost is saved. Then, the second time segment is simulated over the loading volumes assigned in the first time segment to find the optimal solution (user equilibrium) in the second time segment. Hence, the toll costs are dynamically updated on a time segment-by-time segment basis.

< Time Segment: 1 >

Iteration	Initial/current Toll	Road Density	Updated Toll
1	$0.25	20	$1.68
2	$1.68	16	$1.00
3	$1.00	19	$1.50

ITERLOADIN
C = 2

< Time Segment: 2 >

Iteration	Initial/current Toll	Road Density	Updated Toll
3	$0.25	28	$3.10
4	$3.10	23	$2.00
5	$2.00	25	$2.57

< Time Segment: 3 >

Iteration	Initial/current Toll	Road Density	Updated Toll
5	$0.25	22	$2.04
6	$2.04	24	$2.39
7	$2.39	20	$1.68

Figure 4-5 Update of toll costs by a time segment-by-time segment basis

4.6 Scripting in Cube Avenue Model

The prototype managed lane model was developed using the Cube Avenue program that can be scripted by using the Cube Avenue-based programming language. The Cube Avenue program normally includes five phases, as follows:

- SETUP phase: to initialize certain variables and arrays
- LINKREAD phase: to obtain required values for each link
- ILOOP phase: to create packets and to determine routes for each time segment

- ADJUST phase: to simulate the movement of packets through the network
- CONVERGENCE phase: to check the convergence of the model run

The parameter values, as shown in Figure 4-6, are set at the beginning of the model run before starting the Cube Avenue phases. The user should specify these parameters for proper model performance. The prototype managed lane model uses the average loading method (e.g., COMBINE=AVE), along with the packet allocation mode (e.g., PACKETS=PA), but the packet splitting option (e.g., PACKETS=PS) can also be tested in the model validation step. If the incremental loading process by time segment is more suitable to the model, the incremental loading keyword (e.g., ITERLOADINC) can be used in addition to setting the maximum iterations (e.g., MAXITERS). The length of the model period, the time segments, and the number of queuing (storage) vehicles are specified using the MODELPERIOD, SEGMENTS, and VEHPERDIST keywords. As a new keyword (developed as part of this study as discussed later in this chapter), GENPKTBYITER should be set as true (default is false) in order to generate the packets for each iteration. Otherwise, Cube Avenue will only generate the packets one time in the first iteration. This keyword should be declared immediately after PACKETS=PA. The PRESERVEMW keyword allows Cube Avenue to preserve the matrix cell values for the specified intermediate (MW) matrices for every iteration. Otherwise, Cube Avenue discards the matrix values at the end of each iteration.

```
;set total number of zones
PARAMETERS ZONES={ZONES}
;set dynamic traffic assignment methodology
PARAMETERS COMBINE=AVE, PACKETS=PA, GENPKTBYITER=T
PARAMETERS ITERLOADINC={ITERLOADINC}, MAXITERS={MAXITER}
;set model period and time segment list
PARAMETERS MODELPERIOD=180, SEGMENTS=12*15 ; fifteen-minute time interval
;set assumptions for default storage (vehicles per lane per mile)
PARAMETERS VEHPERDIST={VEHPERDIST}
PARAMETERS PRESERVEMW=201-224,261-284,321-332,341-352
REPORT SPEED=YES CAPACITY=YES       ;report speed/capacity tables in network
```

Figure 4-6 Setting model parameter values

In the first SETUP phase, three array variables are defined, as shown in Figure 4-7, to indicate the maximum density in each directional toll road and to store the updated toll costs using the maximum density by each time segment, for each iteration. Note that the SETUP phase header can be omitted.

```
ARRAY _MAX_DENSITY=100
ARRAY TOLL_SEG1_SB=20, TOLL_SEG1_NB=20
```

Figure 4-7 Set of SETUP phase

In the LINKREAD phase, as in Figure 4-8, the required link attributes such as speed, capacity, storage, and free-flow travel time are set for the input network, and additional link attributes, such as link class and initial toll cost, are specified for use in other phases.

to set required link

```
PROCESS PHASE=LINKREAD
SPEED=LI.SPEED
T0=LI.TIME
C=LI.CAPACITY
STORAGE=LI.STORAGE

IF (STORAGE=0) STORAGE=9999999
IF (C=0)  C=9999999
```

to set VDF in FUNCTION

```
;--- link class
IF (LI.FT=51,52)                   ; zonal centroid connector
   LINKCLASS= 1
ELSE                               ; other roads
   LINKCLASS= 2
ENDIF
```

to exclude links in path-

```
;--- link activations
IF (LI.TOLL_LINK=1)                ; fixed toll lanes
   ADDTOGROUP=1
ELSEIF (LI.HOV_LINK=1)             ; HOV lanes (free)
   ADDTOGROUP=2
ELSEIF (LI.HOT_LINK=1)             ; HOT lanes
   ADDTOGROUP=3
ENDIF
```

to set the initial tolls

```
;--- toll costs
IF (ITERATION=0)
   T1=LI.TIME
   LW.TOLL1 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 1 )
   LW.TOLL2 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 2 )
   LW.TOLL3 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 3 )
   LW.TOLL4 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 4 )
   LW.TOLL5 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 5 )
   LW.TOLL6 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 6 )
   LW.TOLL7 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 7 )
   LW.TOLL8 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 8 )
   LW.TOLL9 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 9 )
   LW.TOLL10=LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 10)
   LW.TOLL11=LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 11)
   LW.TOLL12=LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 12)
ENDIF
ENDPROCESS
```

Figure 4-8 LINKREAD phase setting

The LINKCLASS values correspond to the index of link performance functions (e.g., TC[]) in the FUNCTION statement (a LINKCLASS index can be used, for example, to specify the parameters of the traffic flow model (TFM) by facility type). The initial toll costs in the toll road are set as $0.25 for the entry ramp (to the managed lane) facilities. The ADDTOGROUP keyword can be used to exclude certain links when searching the routes for a skimming or loading process.

In the ILOOP phase, Cube Avenue runs the skimming process for each iteration, as shown in Figure 4-9, to allow the toll diversion process to use these impedance values (e.g., travel times and toll costs), so as to estimate the toll trip proportions for the SOV and HOV2P trips. The skimming process is performed simultaneously for 12 time segments by specifying the departure time for each time segment.

```
;--- building paths
DYNAMICLOAD PATH=COST, EXCLUDEGRP=2-3,                               ; SOV free roads
            MW[201]=TRACE(0,TIME),             NOACCESS=0,    ; time segment 1
            MW[202]=TRACE(0,LW.TOLL1),         NOACCESS=0,
            MW[203]=TRACE(15,TIME),            NOACCESS=0,    ; time segment 2
            MW[204]=TRACE(15,LW.TOLL2),        NOACCESS=0,
            MW[205]=TRACE(30,TIME),            NOACCESS=0,    ; time segment 3
            MW[206]=TRACE(30,LW.TOLL3),        NOACCESS=0,
            MW[207]=TRACE(45,TIME),            NOACCESS=0,    ; time segment 4
            MW[208]=TRACE(45,LW.TOLL4),        NOACCESS=0,
            MW[209]=TRACE(60,TIME),            NOACCESS=0,    ; time segment 5
            MW[210]=TRACE(60,LW.TOLL5),        NOACCESS=0,
            MW[211]=TRACE(75,TIME),            NOACCESS=0,    ; time segment 6
            MW[212]=TRACE(75,LW.TOLL6),        NOACCESS=0,
            MW[213]=TRACE(90,TIME),            NOACCESS=0,    ; time segment 7
            MW[214]=TRACE(90,LW.TOLL7),        NOACCESS=0,
            MW[215]=TRACE(105,TIME),           NOACCESS=0,    ; time segment 8
            MW[216]=TRACE(105,LW.TOLL8),       NOACCESS=0,
            MW[217]=TRACE(120,TIME),           NOACCESS=0,    ; time segment 9
            MW[218]=TRACE(120,LW.TOLL9),       NOACCESS=0,
            MW[219]=TRACE(135,TIME),           NOACCESS=0,    ; time segment 10
            MW[220]=TRACE(135,LW.TOLL10),      NOACCESS=0,
            MW[221]=TRACE(150,TIME),           NOACCESS=0,    ; time segment 11
            MW[222]=TRACE(150,LW.TOLL11),      NOACCESS=0,
            MW[223]=TRACE(165,TIME),           NOACCESS=0,    ; time segment 12
            MW[224]=TRACE(165,LW.TOLL12),      NOACCESS=0
DYNAMICLOAD PATH=COST, EXCLUDEGRP=2,                                 ; SOV HOT-lane toll roads
            MW[261]=TRACE(0,TIME),             NOACCESS=0,
            MW[262]=TRACE(0,LW.TOLL1),         NOACCESS=0,
            MW[263]=TRACE(15,TIME),            NOACCESS=0,
            MW[264]=TRACE(15,LW.TOLL2),        NOACCESS=0,
            MW[265]=TRACE(30,TIME),            NOACCESS=0,
            MW[266]=TRACE(30,LW.TOLL3),        NOACCESS=0,
            MW[267]=TRACE(45,TIME),            NOACCESS=0,
            MW[268]=TRACE(45,LW.TOLL4),        NOACCESS=0,
            MW[269]=TRACE(60,TIME),            NOACCESS=0,
            MW[270]=TRACE(60,LW.TOLL5),        NOACCESS=0,
            MW[271]=TRACE(75,TIME),            NOACCESS=0,
            MW[272]=TRACE(75,LW.TOLL6),        NOACCESS=0,
            MW[273]=TRACE(90,TIME),            NOACCESS=0,
            MW[274]=TRACE(90,LW.TOLL7),        NOACCESS=0,
            MW[275]=TRACE(105,TIME),           NOACCESS=0,
            MW[276]=TRACE(105,LW.TOLL8),       NOACCESS=0,
            MW[277]=TRACE(120,TIME),           NOACCESS=0,
            MW[278]=TRACE(120,LW.TOLL9),       NOACCESS=0,
            MW[279]=TRACE(135,TIME),           NOACCESS=0,
            MW[280]=TRACE(135,LW.TOLL10),      NOACCESS=0,
            MW[281]=TRACE(150,TIME),           NOACCESS=0,
            MW[282]=TRACE(150,LW.TOLL11),      NOACCESS=0,
            MW[283]=TRACE(165,TIME),           NOACCESS=0,
            MW[284]=TRACE(165,LW.TOLL12),      NOACCESS=0
```

Figure 4-9 Skimming process in ILOOP phase

As shown in Figure 4-10, the toll values at the toll entry ramps for each corridor are updated at the beginning (e.g., TIMESEGMENT=0 and I=1) of every iteration (e.g., ITERATION>1) with the exception of the first iteration, which uses the initial toll values. These toll values are computed using the maximum link density in the ADJUST phase from the previous iteration and are stored in the array variables (e.g., TOLL_SG1_SB[] or TOLL_SG1_NB[]). The user can specify additional array variables if the model contains multiple toll roads. Note that the LINKLOOP keyword can be utilized to access the link attributes for each link in the ILOOP phase that is performed in the zonal basis.

```
;--- update LW toll variables in ILOOP
IF (ITERATION>1 & TIMESEGMENT=0 & I=1)   ; before performing skimming
   LINKLOOP
      IF (LI.HOT_SEG_R=11)          ; toll for each time segment in I-95 HOT SB ramp
         LW.TOLL1 =TOLL_SEG1_SB[1]
         LW.TOLL2 =TOLL_SEG1_SB[2]
         LW.TOLL3 =TOLL_SEG1_SB[3]
         LW.TOLL4 =TOLL_SEG1_SB[4]
         LW.TOLL5 =TOLL_SEG1_SB[5]
         LW.TOLL6 =TOLL_SEG1_SB[6]
         LW.TOLL7 =TOLL_SEG1_SB[7]
         LW.TOLL8 =TOLL_SEG1_SB[8]
         LW.TOLL9 =TOLL_SEG1_SB[9]
         LW.TOLL10=TOLL_SEG1_SB[10]
         LW.TOLL11=TOLL_SEG1_SB[11]
         LW.TOLL12=TOLL_SEG1_SB[12]
      ENDIF
      IF (LI.HOT_SEG_R=12)          ; toll for each time segment in I-95 HOT NB ramp
         LW.TOLL1 =TOLL_SEG1_NB[1]
         LW.TOLL2 =TOLL_SEG1_NB[2]
         LW.TOLL3 =TOLL_SEG1_NB[3]
         LW.TOLL4 =TOLL_SEG1_NB[4]
         LW.TOLL5 =TOLL_SEG1_NB[5]
         LW.TOLL6 =TOLL_SEG1_NB[6]
         LW.TOLL7 =TOLL_SEG1_NB[7]
         LW.TOLL8 =TOLL_SEG1_NB[8]
         LW.TOLL9 =TOLL_SEG1_NB[9]
         LW.TOLL10=TOLL_SEG1_NB[10]
         LW.TOLL11=TOLL_SEG1_NB[11]
         LW.TOLL12=TOLL_SEG1_NB[12]
      ENDIF
   ENDLINKLOOP
ENDIF
```

Figure 4-10 Update of toll costs in the ILOOP phase

The toll diversion process described in the previous section is constructed using the Cube Avenue programming scripts, as shown in Figure 4-11, in order to estimate the toll trip proportions (%) for the SOV and HOV2P trips. In this prototype model, the willingness-to-pay curve is applied to search for the toll traffic proportions based on the toll cents per minute saved between the toll and free routes. This process is run in every time segment in the ILOOP phase to split the input SOV and HOV2P trips into the free trips vs. the toll trips. These working matrices are output at the end of the Cube Avenue run to be reviewed during the calibration and validation processes of the model.

```
;--- applying not-willingness to pay proportions for free vs. toll
IF (TIMESEGMENT>0)
   _TS=TIMESEGMENT
   _IDX1=70+_TS                              ; input SOV total trip index
   _IDX2=560+_TS                             ; for SOV free trip index
   _IDX3=580+_TS                             ; for SOV toll trip index
   _T_IDX=200+_TS*2-1                        ; index for free time skim
   _C_IDX=260+_TS*2                          ; index for toll skim
   ;--- SOV diversion process for HOT-lanes
   JLOOP
      IF (MW[_IDX1]>0)                       ; if there exist SOV trips
         TSAVE=MW[_T_IDX]-MW[_T_IDX+60]      ; time saving (free time - toll time)
         IF (TSAVE>0 & MW[_C_IDX]>0)         ; if positive time saving & positive toll cost($)
            TCPS=(MW[_C_IDX]*100)/TSAVE      ; average toll cost (cents) per minute saved
            PWPT=100-DIVERT(1,TCPS)          ; proportion (%) willing to pay toll by average toll-cents
         ELSE
            PWPT=0
         ENDIF
         MW[_IDX3]=MW[_IDX1]*(PWPT/100)      ; SOV toll trips (Debug_8.MAT)
         MW[_IDX2]=MW[_IDX1]-MW[_IDX3]       ; SOV free trips (Debug_9.MAT)
      ELSE
         MW[_IDX3]=0
         MW[_IDX2]=0
      ENDIF
   ENDJLOOP
ENDIF
```

Figure 4-11 Toll diversion process in ILOOP phase

Finally, a dynamic loading process in the ILOOP phase is scripted, as in Figure 4-12, and will be performed for each traffic mode (e.g., free trips, toll trips, HOV3+ trips, and truck trips) and each time segment. Cube Avenue builds paths according to some attribute minimization criterion, as well as determines the number of trips that use each path, but it defers updating the volume fields until the simulation results are obtained from the ADJUST phase. The trips are placed into packets, where each packet contains a start time, a volume in one or more volume fields, and a route. However, the network's volume fields may only be updated after calculations are processed for each point on the route that will determine the actual time (that is, in which time segment) of the packet's arrival. It is suggested that the packet size is set as one in the managed lane model so that each packet has an option to select either the free route or toll route. The BUILDALLPATHS keyword (another new keyword added in this research) should be set as true in order to generate all possible OD paths for each iteration. Otherwise, Cube Avenue cannot process the managed lane model properly because the vehicle trips in the OD pairs vary between iterations.

```
;--- loading free and toll trips
;  SOV free trips
DYNAMICLOAD PATH=COST, BUILDALLPATHS=T VOL[1]=MW[(320+__TS__)], PACKETSIZE={PacketSize}, EXCLUDEGRP=2-3
;  SOV HOT-lane toll trips
DYNAMICLOAD PATH=COST, BUILDALLPATHS=T VOL[2]=MW[(340+__TS__)], PACKETSIZE={PacketSize}, EXCLUDEGRP=2
;  HOV free trips
DYNAMICLOAD PATH=COST, BUILDALLPATHS=T VOL[3]=MW[(401+__TS__)], PACKETSIZE={PacketSize}
;  Truck free trips
DYNAMICLOAD PATH=COST, BUILDALLPATHS=T VOL[4]=MW[(451+__TS__)], PACKETSIZE={PacketSize}, EXCLUDEGRP=2-3
```

Figure 4-12 Dynamic loading process in ILOOP phase

It should be mentioned that functions can be defined anywhere in Cube Avenue to compute certain values using expressions for each function, such as link performance, costs, and volumes. The prototype model specifies the standard volume-delay functions in the ADJUST phase, as shown in Figure 4-13, but the user can modify these functions during the validation process of the managed lane model.

```
PROCESS PHASE=ADJUST
FUNCTION {
   V=VOL[1]+VOL[2]+VOL[3]+VOL[4]
   TC[1]=T0
   TC[2]=T0*(1.0 + LI.COEF*((V/C)^LI.EXPO))
   COST=TIME}
```

Figure 4-13 Set of functions in ADJUST phase

In the ADJUST phase, as shown in Figure 4-14, the toll costs are computed using the road density for each directional corridor, each time segment, and each iteration. This script searches the maximum density for each directional road. Then, the toll costs are updated using the input lookup table, which is based on the real-world toll schedule (as a function of density) used for the I-95 managed lane in Miami.

```
;--- updating tolls based on maximum density for each directi... c
IF (TIMESEGMENT>0)                ; for each time segment
   _CTS=TIMESEGMENT               ; current time segment
   IF (LINKNO=1)
      LOOP _I=11,12               ; maximum density for each HOT direction (1=SB & 2=NB) for I-95 (segment 1)
         _MAX_DENSITY[_I]=0
      ENDLOOP
      TOLL_SEG1_SB[_CTS]=0        ; toll value for each time segment for each HOT direction in I-95 (segment 1)
      TOLL_SEG1_NB[_CTS]=0
   ENDIF

   ;--- computing link density (vpmpl=vehicle per mile per lane)
   ; density = (vehicle/hour)/(mile/hour) = hourly-vehicle/speed = vehicle per mile per lane
   IF (LI.DISTANCE>0 & TIME>0 & LI.LANES>0)
      _SPEED=(LI.DISTANCE/TIME)*60
      _LINK_DEN=((V*(60/{IntTimeSeg}))/_SPEED)/LI.LANES
   ELSE
      _SPEED=0
      _LINK_DEN=0
   ENDIF

   ;--- updating toll values based on maximum density for each direction in each HOT corridor
   IF (LI.HOT_SEG_C=11)           ; I-95 (segment 1) HOT SB corridor links (11=SB & 12=NB)
      IF (_MAX_DENSITY[11]<_LINK_DEN)
         _MAX_DENSITY[11]=_LINK_DEN
         TOLL_SEG1_SB[_CTS]=MIN(7,TOLL_HOT(1,_MAX_DENSITY[11]...   ;
      ENDIF
   ELSEIF (LI.HOT_SEG_C=12)       ; I-95 (segment 1) HOT NB corridor links (11=SB & 12=NB)
      IF (_MAX_DENSITY[12]<_LINK_DEN)
         _MAX_DENSITY[12]=_LINK_DEN
         TOLL_SEG1_NB[_CTS]=MIN(7,TOLL_HOT(1,_MAX_DENSITY[12]))   ; toll in I-95 HOT NB ramp
      ENDIF
   ENDIF
ENDIF
```

to initialize variables

to compute link density

to compute toll costs

Figure 4-14 Dynamic toll update process

4.7 Major Input and Output Files

The prototype managed lane model was implemented and tested based on the I-95 managed lane facility in Miami, Florida. Mainly, this model includes 6 input files and 2 main output files, as shown in Figure 4-15 and in Table 4-3, but the user can update the model by including more input/output files, as needed. Three OD trip tables for SOV plus HOV2, HOV3+, and truck are constructed as the input vehicle trips for every 15 minutes for the 3-hour modeling period. Thus, each OD input file contains 12 matrices for 12 time segments. Note that the toll diversion process in the model will estimate the toll proportions for vehicles traveling on the managed lane corridor so as to split the SOV plus HOV2 OD trips into willing to pay and non-willing to pay trips. As stated earlier, the willingness-to-pay trips will be allowed to select paths other than the ML, if using these paths will reduce travel disutility, as measured by the generalized cost function of the assignment. Highway network attributes are required to simulate the traffic packets as they move in the network. Thus, the highway network and associated link attributes should be provided as input data. The input highway network can be constructed in a Cube network format (e.g., *.NET), GIS shape format (*.SHP), or GIS format stored in a geodatabase file (e.g., *.MDB or *.GDB). Two lookup tables are set to estimate the toll proportions (%) and update the toll costs dynamically, but these input files are not required if the model has a script process that specifies these tables internally.

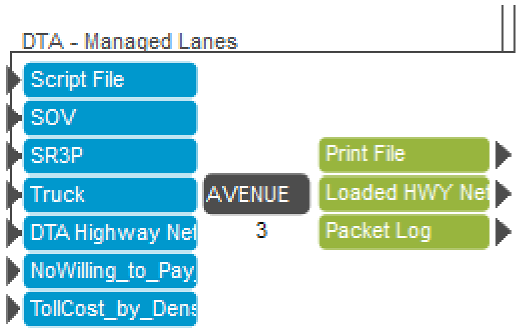


Figure 4-15 Cube application module for Avenue model

Table 4-3 Major Input and Output Files for the Avenue Model

File Type	Filename	Description
Input File	DTA_Highway.NET	Input Cube highway network (*.NET)
	SOV_TS_12.MAT	OD vehicle trips for drive alone (DA) & two persons (SR2) modes by 12 time segments
	SR3_TS_12.MAT	OD vehicle trips for three or more persons (SR3P) mode by 12 time segments
	Truck_TS_12.MAT	OD vehicle trips for truck mode by 12 time segments
	NoWiling_to_Pay_Proportions.dbf	Non-willing to pay proportions (%) for toll diversion process
	TollCost_by_Density.dbf	Toll rates ($) by road density (vehicles per mile per lane)
Output File	DTA_Managed_Lanes_Loaded.NET	Output loaded network (*.NET)
	DTA_Managed_Lanes_Loaded.LOG	Packet log from assignment

Once the prototype managed lane model is performed successfully and completely, three output files are generally created as a result. As the first major output file, the highway loaded network file (*.NET) contains output link attributes estimated for each individual time segment and for the whole model period. The packet log file (*.LOG) includes each individual movement of every packet for travel route, arrival time, and departure time. This packet log file can be animated, along with the highway network in Cube Base as a post-process to review any issues due to a bottleneck, as well as to confirm the vehicles' proper movements. Finally, the print file

(*.PRN) provides a summary of model performance for every iteration and inspects any warning messages generated during the model run. In addition, multiple matrix or text files can be generated, allowing user reviews of the model performance outputs during iterations or at the end of the model run.

4.8 Required Network Attributes and Model Settings

In order to run the model properly, several necessary values specified in the input highway network and Cube key variables for each scenario are required. Any improper values may cause an error or generate warning messages during the model operation.

4.8.1 Highway Network

The input highway network contains various link attributes to be utilized during the model run. As listed in Table 4-4, the prototype Cube Avenue-based model uses link data, such as length (mile), speed (mph), free-flow time (minutes), capacity, storage, facility type, and area type. Note that the input capacity is for the whole model period, not for an hour or for a time segment. It means that if the model period is three hours, and time segments are 15 minutes, capacity should be provided for 3 hours, not for one hour or 15 minutes. It is also suggested that the storage for each link can be computed in a pre-process step rather than computing the link storage in the LINKREAD phase of the Cube Avenue program. This will avoid confusion in setting up the correct link storages.

Link parameter values (such as COEF and EXPO for the volume-delay functions) can be directly input for each link in the highway network, but these values are not required if the model has already specified volume-delay functions. Cube Avenue allows the specification of any type of customized volume-delay functions. It also allows specifying junction data input.

Two types of toll systems are allowed to be specified at the toll facility links of the input highway network: fixed tolls and dynamic tolls. The fixed tolls do not change with changing traffic congestion, while the dynamic tolls are updated based on the traffic congestion level, either on the ML facility by itself, or a combination of the conditions of ML, general purpose lanes (GPL), and possibly other facilities. For example, the I-95 ML implementation selects the toll costs based on the vehicle density on the ML. Other measures, such as the volume-to-capacity ratio can also be used to update the toll costs for each time segment. The fixed toll costs are specified in TOLL_FIXED as the dollar unit, and any positive values in TOLL_LINK indicate that it is a toll facility with the fixed toll costs.

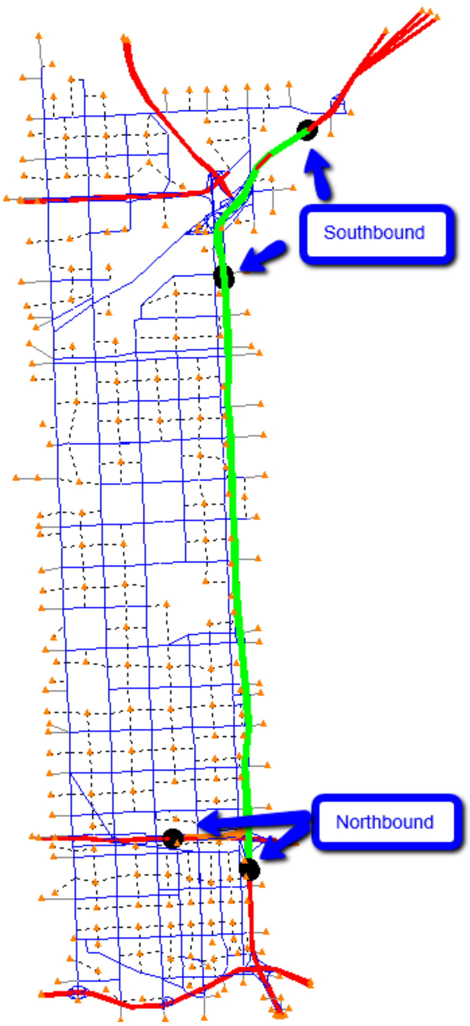
The dynamic toll costs are specified as a dollar unit in HOT_SEG_T that would be used as the initial input tolls to begin the Cube Avenue-based model. Hence, the model updates the toll costs based on traffic conditions for each time segment. The HOT_LINK indicates that this is a ML link if it is set as 1, while HOV_LINK indicates a HOV facility where only HOV trips can utilize the HOV lanes without paying any tolls.

The HOT_SEG_C indicates the directional information for the HOT facilities. For example, these values for the I-95 ML facility are set as 11 for the southbound corridor, and 12 for the northbound corridor. The prototype model uses these values to search the maximum density for each direction of the I-95 ML corridor. If the input network includes multiple managed lane facilities in the project area, a user can specify additional values for the HOT_SEG_C attribute.

Table 4-4 Required Link Attributes in Input Highway Network

Link Attribute	Description
DISTANCE	Link distance (mile)
SPEED	Link speed (mph)
TIME	Link free-flow time (min)
CAPACITY	Link capacity (vehicles) for 3 hours period
FT	Link facility type (=FTC2)
AT	Area type
LANES	Number of lanes
STORAGE	Number of vehicles that can fill the link
COEF	Coefficient value for BPR link performance function
EXPO	Exponent value for BPR link performance function
TOLL_FIXED	Fixed toll rate ($) for other toll facilities
TOLL_LINK	Toll links with fixed tolls (FT=91 or 95)
HOV_LINK	HOV links (FT=81) for no-toll
HOT_LINK	HOT links (FT=88), on-ramps (FT=87), and off-ramps (FT=89)
HOT_SEG_C	HOT lanes for each directional bound (e.g. 11=SB lane & 12=NB lane)
HOT_SEG_R	HOT entry ramps (e.g. 11=SB ramp gate & 12=NB ramp gate)
HOT_SEG_T	Initial toll cost ($) in each HOT ramp facility

The HOT_SEG_R link attribute shows the specific locations of the ML entry ramps, which are the points of charging the tolls for SOV or HOV2 vehicles that are taking the ML in the model prototype. As shown in Figure 4-16, the highway network has two ML entry (gate) ramps (HOT_SEG_R=11) for the southbound direction, and two other entry gate ramps (HOT_SEG_R=12) for the northbound direction. If the project area includes multiple ML facilities, the user can specify different values in this link attribute.



Highway Links

AX/BX	915758.69	915807.06
AY/BY	574964.69	574583.31
A	1428	
B	1427	
DISTANCE	0.4723	
SPEED	30	
TIME	0.9445	
CAPACITY	2030	
FT	87	
AT	4	
LANES	1	
STORAGE	103.8994	
COEF	0.32	
EXPO	7	
TOLL_FIXED	0	
TOLL_LINK	0	
HOV_LINK	0	
HOT_LINK	1	
HOT_SEG_C	0	
HOT_SEG_R	11	
HOT_SEG_T	0.25	

Figure 4-16 Locations of HOT entry ramp facilities

4.8.2 *Cube Catalog Keys*

The prototype model uses various Cube catalog keys listed in Figure 4-17 and Table 4-5 that can be utilized during the model script runs. The Cube Avenue script can use these key variables to run the model for each scenario. These keys in the script are automatically set as the modeling values prior to the run. A user can modify these key values in the scenario panel of the Cube Base tool, if necessary. For example, if the model is based on 12 time segments for a 3-hour modeling period with 15 minutes for each time segment, both {TimeSegment} and {IntTimeSeg} keys should be specified as 12 and 15, respectively. The number of iterations for the model run is controlled by specifying the maximum iteration ({MAXITER}) and the incremental time segment loading ({ITERLOADINC}).

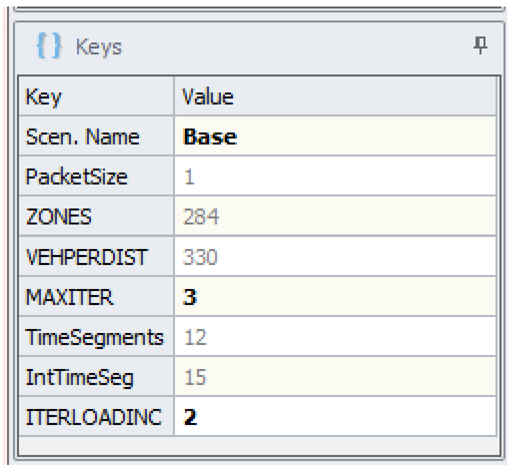


Figure 4-17 Setting of Keys in Cube Base

Table 4-5 Description of Key Variables

Key Name	Description
{ZONES}	Number of zones
{PacketSize}	Packet size in Avenue run (e.g. 1 as default)
{VEHPERDIST}	Number of vehicles that can physically fit on a link (e.g. 330 vehicles/mile/lane)
{MAXITER}	Maximum iteration in running Avenue program (e.g. 5 as default)
{TimeSegment}	Number of time intervals (e.g. 12 as default)
{IntTimeSeg}	Time length (min) per each time segment (e.g. 15 min as default)
{ITERLOADINC}	Number of iterations for loading in each time segment (e.g. 0 as default)

4.9 Output Results

One of the major outputs of the model is the loaded highway network, which contains various link attributes estimated by the Cube Avenue model, as shown in Table 4-6. Unlike static traffic assignment, in which only one model period is used, the DTA simulates each individual vehicle (or packet) for each time segment period. Hence, it can estimate the assigned volumes (e.g., VS1_1, VS2_1, VS3_1, etc.) for each time segment, in addition to the total volumes (e.g., VSMP_1) in the model period.

The congested travel time (e.g., TIMES1_1, TIMES2_1, TIMES3_1, etc.) and speed (e.g., SPEEDS1_1, SPEEDS2_1, SPEEDS3_1, etc.) are also estimated for each link in each time segment. The congested travel time and speed are averaged for the model period, as represented by TIME_1 and CSPD_1. The loaded network also includes additional vehicle estimated measures, such as the number of vehicles (e.g., VITS1_1, VITS2_1, VITS3_1, etc.) on each downstream link at the end of the time segment, the average number of vehicles queuing (e.g., QUEUEVS1_1, QUEUEVS2_1, QUEUEVS3_1, etc.) on the link during the time segment, and the number of vehicles (e.g., BLOCKVS1_1, BLOCKVS2_1, BLOCKVS3_1, etc.) in the queue at the end of the simulation. The final toll costs in the HOT entry ramps are referred to as LW_TOLL1_1 for the first time segment, LW_TOLL 2_1 for the second time segment, LW_TOLL 3_1 for the third time segment, and so on.

Table 4-6 Link Attributes in Loaded Highway Network

Link Attribute	Description
TIME_1	Averaged congested time (min)
TIMES#_1	Congested time (min) for each time segment (#)
CSPD_1	Averaged congested speed (mph)
SPEEDS#_1	Congested speed (mph) for each time segment (#)
VSMP_1	Total assigned vehicles
VS#_1	Assigned vehicles for each time segment (#)
V*S#_1	Assigned vehicles for each mode (*) and each time segment (#)
VITS#_1	Number of vehicles on each downstream link at the end of each time segment (#)
QUEUEVS#_1	Average number of vehicles queuing on the link during time segment (#)
BLOCKVS#_1	Number of vehicles in the queue that will remain in the queue at the end of the simulation
LW_TOLL#_1	Toll values in HOT entry on-ramps for each time segment (#)

Figure 4-18 and Figure 4-19 show the assignment link results, including tolls, volumes, queues, vehicles-in-transit, and blocks estimated for each time segment. As shown in Figure 4-20, the packets can be animated into the highway network using the packet log file (*.LOG) that is generated at the end of the model run.

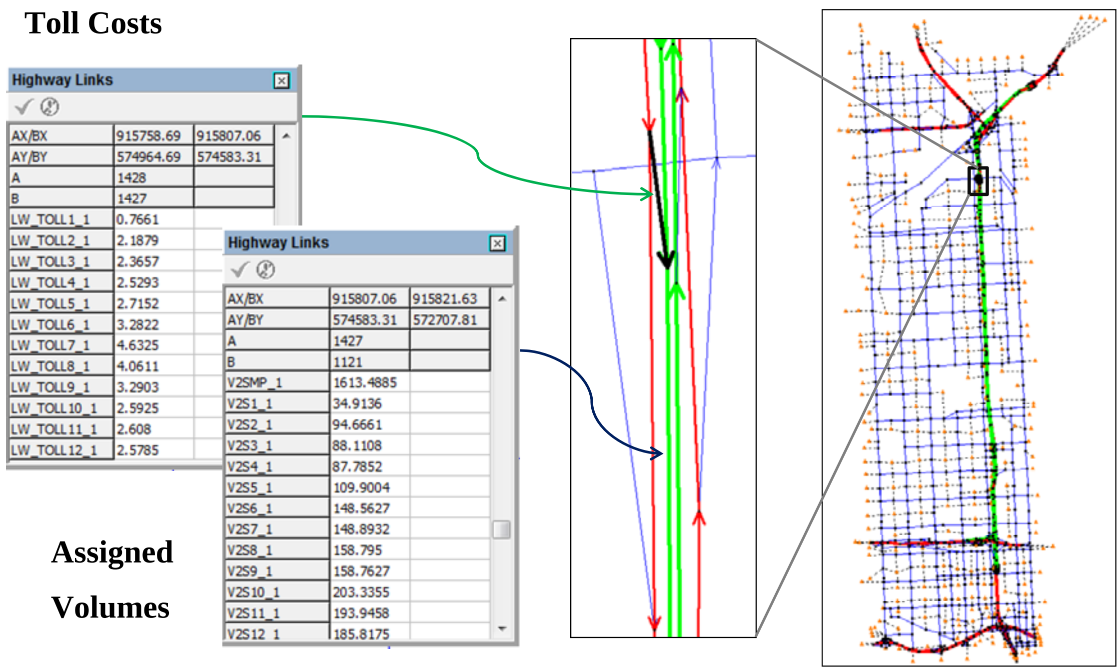


Figure 4-18 Estimated tolls and link volumes

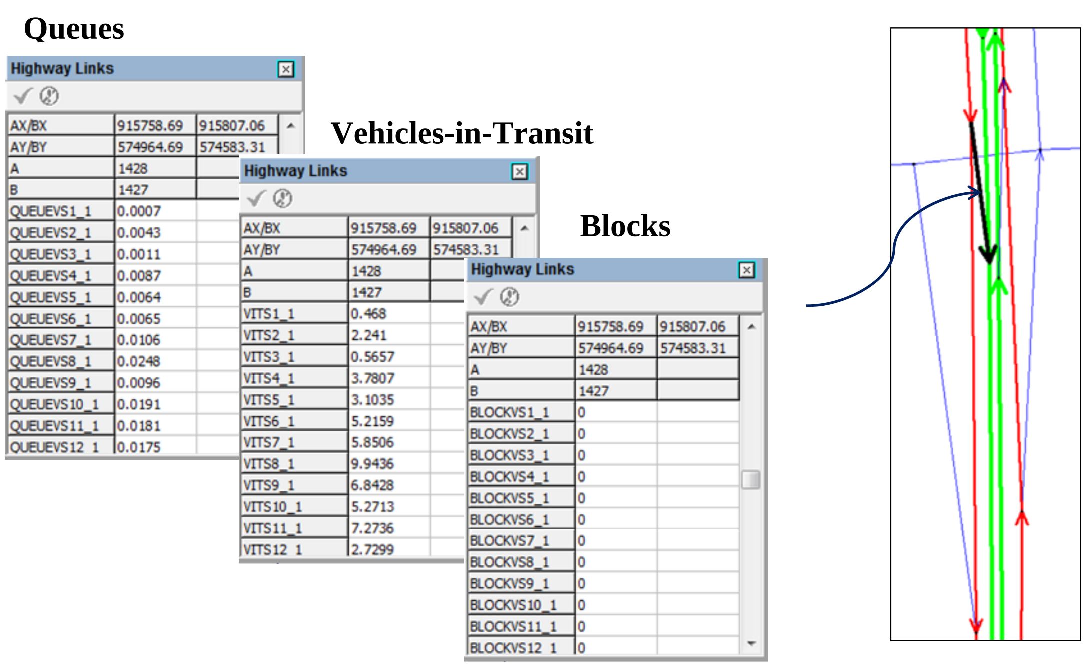


Figure 4-19 Estimated queue, vehicle in transit, and block

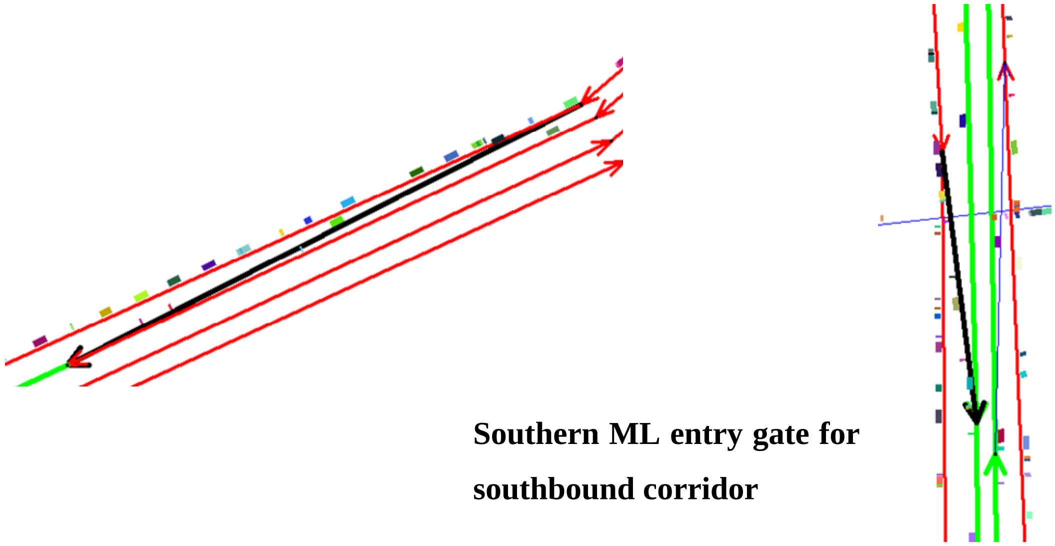


Figure 4-20 Animation of vehicle packets in ML entry gates

4.10 Cube Avenue Enhancements

During the course of this research, several issues were identified that limited Cube Avenue's modeling abilities, in particular, as it relates to managed lane (ML) as implemented in the abovementioned prototype. Citilabs, which is a member of this research, addressed the identified issues, allowing the final development and test of the prototype. Following is a list of the most significant enhancements made to Cube Avenue during this research:

- Cube Avenue used packets generated only in the first iteration of the user equilibrium process. Implementing the willingness-to-pay curve in combination with the assignment requires changing the packets willingness-to-pay in the following iterations based on the toll diversion process. The GENPKTBYITER keyword is a new keyword that when used, allows the generation of new packets in each iteration.
- The BUILDALLPATHS keyword is another new keyword that was introduced. Without this keyword, if a certain OD pair has no packet in an iteration, Cube Avenue will not generate the route in the iteration. But this becomes a problem if the same OD pair has some positive packets in other iterations causing an error under this condition. Thus, the new GENPKTBYITER keyword should be used to allow the managed lane model to

generate the path for each OD pair and every iteration using the BUILDALLPATHS keyword. The two new keywords GENPKTBYITER and BUILDALLPATHS are needed to avoid incorrect results and errors.

- Cube Avenue could not perform the skimming process at the beginning of each time segment because it was implemented after the last time segment. This is significant since the skimming at the beginning of the time period is required to allow the updating of the willingness-to-pay share based on the skimming results. This process is now updated to perform matrix skimming at the beginning of the time period.
- The output matrices for skimming and toll trips included incorrect values. This is now resolved.
- When the INTELOADINC keyword was defined, which allows user equilibrium iterations for each time period, the resulting assigned volumes were unreasonable. This is now resolved.
- In some cases, the assigned volumes exceeded the link capacities. This is now resolved.
- Increase in the packet size may result in incorrect results. The PACKETSIZE=1 is recommended for managed lane applications.
- Cube Avenue did not recognize the LOOKUP command. This is now resolved.

5. Supply Calibration

Supply or network calibration estimates the network parameters that define network performance in producing travel time, forming queues, and queue spillback. As previously mentioned, a systematic multilevel approach of network calibration is adopted in this study, with an increasing calibration scope in each level. The process starts at the level of separated bottlenecks where capacity is estimated by various methods based on field data. The network is gradually extended to connect the bottlenecks and then to the whole corridor and subarea coverage. The parameters from previous steps are fine-tuned, and the supply-demand calibration runs iteratively until a desirable convergence is achieved.

The advantage of this approach is twofold: First, critical spots of the network can be better identified, analyzed and replicated. Second, a more reliable demand can be estimated for the smaller networks that are the focus of this study, which is very important in the iterative process of demand-network calibration. Focusing on isolated bottleneck locations and the freeway corridor for managed lane assessment enables the capturing of the interactions between supply and demand in addressing the causes for congestion. This is not feasible or tractable in more complicated networks. Once the supply and demand for the focus corridor is well-calibrated, then the network can be extended to other corridors, possibly with less detail in the calibration process.

Speed time-space contours are used extensively as part of the methodology of this study to identify traffic and bottleneck conditions, and their impacts. Figure 5-1 displays speed contours for representative days of low, medium, and high congestion, and an average for all selected days. As can be seen in this figure, the traffic patterns and the reason for the congestion can vary from day-to-day, even after removing non-representative days.

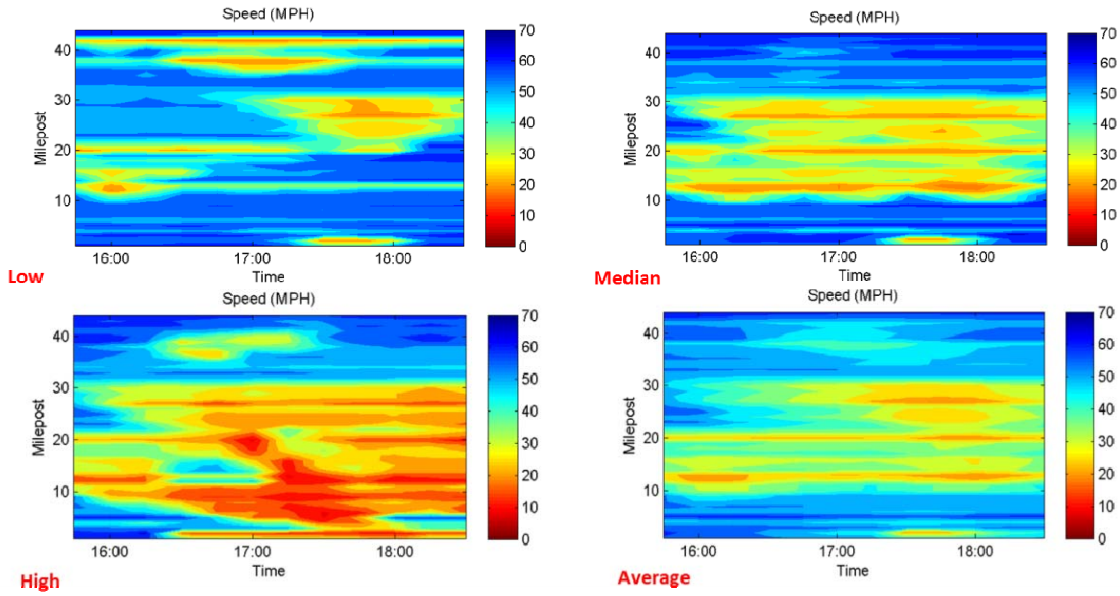


Figure 5-1 Speed contour for different classes of demand

In this research, initial demand matrices were obtained utilizing the network subtraction process from the regional travel demand forecasting model. Regional travel demand models represent a very important source of origin-destination (OD) information that is consistent with the behavior of travelers, as modeled in demand generation, distribution, and mode choice steps.

The regional matrix covers the entire study period of three hours. Acquiring time-dependent trip tables at 15-minute intervals that reflect the current demand and traffic situations required the use of a sequential scheme that iterates between the supply and demand calibrations until convergence. The details of the demand estimation process are presented in Chapter 6.

It should be mentioned here that before the start of the calibration process, checking for mistakes in coding was conducted to omit any errors. In addition, the model "validity" was checked according to FHWA guidance (Sloboden et al., 2012) including conducting a series of stress tests and diagnostic testing steps.

5.1 Bottleneck Identification

In this study, visualization techniques, in combination with comparisons between upstream and downstream measures, were used to identify congested areas and bottleneck locations. Based on the speed contours presented in Figure 5-1, Stations 12, 20, and 28 were identified as bottleneck locations in the PM peak period. Stations 12 and 20 are located in the on-ramp merging areas after the acceleration lane drops. Lane-by-lane data analysis of the intelligent transportation systems (ITS) detector located at Station 28, however, showed that the congestion in this location is definitely caused by a backup from an off-ramp exit to a major freeway (the Florida Turnpike), causing low speeds and high occupancy in the two left lanes, while the three right lanes have light congestion. Thus, the only bottleneck locations that can be used to estimate capacity are those at Stations 12 and 20.

5.2 Free-Flow Speed

In the network under study, based on the Highway Capacity Manual (HCM) 2000, assuming a 6-foot lateral clearance, for a lane width of 11 feet and interchange density of 1.16, the Free-Flow Speed (FFS) is estimated to be 63 mph for segments with three lanes, and 64 mph for segments with four lanes. Based on the HCM 2010 analysis, the FFS is estimated to be around 66.9 mph for most segments (FFS is not depending on the number of lanes in HCM 2010). Based on a combined criterion of volume less than 1000 pc/hr/ln, and occupancy below 10 percent, the FFS values were derived from detector data. Estimating FFS as the 85th percentile of speed over several days as suggested in literature showed very similar results. This value greatly varies between stations (from 54 mph to 64 mph), with an average of 59 mph, which is significantly lower than the HCM 2000 and particularly, the HCM 2010 estimates, as shown in Figure 5-2. It is worth mentioning that the posted speed on all I-95 corridor segment studied is 55 mph.

A previous study on an adjacent corridor (Florida State Road 826) with the same speed limit shows similar differences between the values estimated by the HCM 2000 and HCM 2010; however, it shows a higher measured FFS compared to the present study (28). It is expected that the selected I-95 segment operates differently from an average corridor since it passes through a dense urban environment with frequent interchanges, has vertical and horizontal alignments that may affect capacity, and includes parallel managed lanes that are separated from the general use lanes by soft barriers.

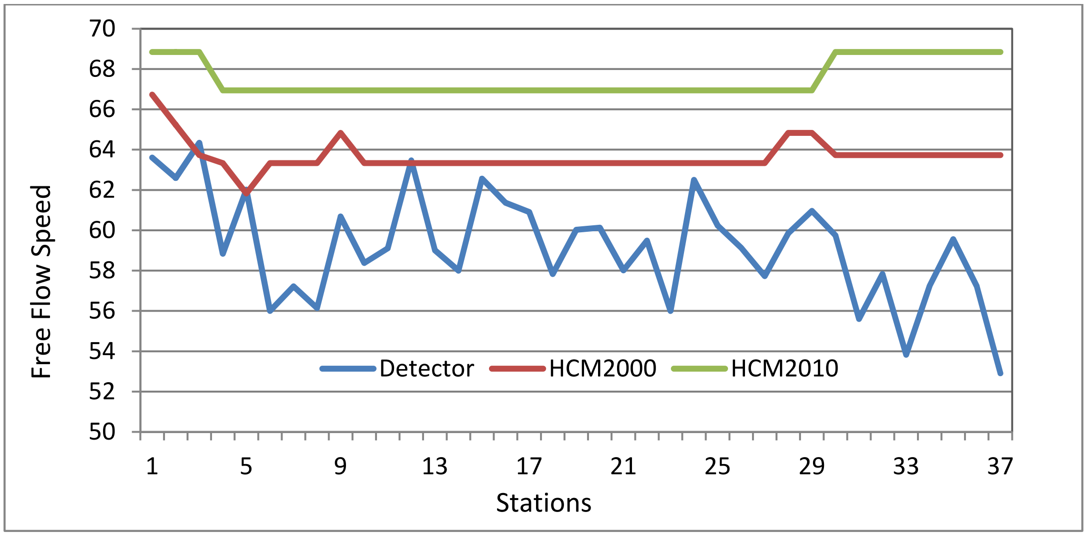


Figure 5-2 Variation of the FFS along the corridor (I-95 NB)

5.3 Capacity Estimation

This section presents a comparison between the capacity values estimated based on different sources and utilizing different methods. HCM is the primary source for estimating highway capacity for planning and operation applications. The HCM capacity values are expressed in personal car per lane per hour and should be converted to vehicle per lane per hour by considering heavy vehicle percentage for comparison with real-world measurements. The heavy vehicle percentage was estimated to be around 5%, based on recorded video observations. The HCM provides adjustment factors for different weather conditions and the degree of familiarity of the drivers with the road (driving population factor). The selected representative days of this study included normal weather conditions, and no necessary adjustments. The driver population factor has a significant effect on adjusting capacity, but is very difficult to obtain, and there is no guideline in the HCM on how to estimate it. The HCM mentions that this value usually varies between 0.85 and 1, and recommends using 1, unless there is sufficient evidence to reduce it, though a default value of 0.95 is mentioned for urban freeways. As is shown in Table 5-1, the values coded in the Southeast Regional Planning Model (SERPM) and estimated based on the Florida LOS/QS manual (5,6) corresponds to those values estimated by HCM for the 5% for percentage of trucks, and 95% for the familiar driver population.

To reflect site specifications, the capacity was also estimated based on detector volume data, aggregated at 15-minute intervals according to the HCM definition of capacity. In order to ensure that the only data utilized in estimating capacity at the bottlenecks are for intervals not affected by downstream congestion, an examination of speed contours was made so as to identify and exclude intervals in which the capacities of the bottlenecks are affected by a spillback from downstream. The difference that resulted from the removal of the data from these intervals in capacity measurements for some methods is presented in Table 5-1. The results clearly show the need for this step. For example, the capacity measurement based on the Rakha method is 1,710 vph without removing the spillback intervals and 1,800 vph when the data from these intervals are removed. Table 5-1 indicates, based on different methods including the pre-breakdown flow method, the Rakha model-based method, and the maximum occupancy method, that the capacity before breakdown is about 1,850 vph. The queue discharge rate appears to be lower than this value based on the results in Table 5-1.

Table 5-1 Estimated Capacity at Active Bottleneck Locations (VPH)

Different methods	St. 600561	St. 600711	Reference
HCM (5% truck, f_{hp}=0.975, f_p=0.98)	2,210	2,210	HCM, 2010
HCM (5% truck, f_{hp}=0.975,f_p =0.95)	2,140	2,140	HCM, 2010
Rakha	1,730	1,700	Rakha & Arafeh, 2010
Rakha (Removed spillbacks)	1,800	1,725	
SERPM coded	2,142	2,142	Cambridge Systematics,
Breakdown flow (15 minutes average before breakdown happens)	1,840	1,810	Elefteriadou and Lertworawanich, 2003
Queue discharge	1,625	1,630	Elefteriadou and Lertworawanich, 2003
Queue discharge (Removed spillbacks)	1,710	1,680	
Maximum 5 minute interval observed (averaged over selected days	1,930	1,925	Dervisoglu, 2009

Different methods	St. 600561	St. 600711	Reference
Maximum 15 minute interval observed (averaged over selected days	1,845	1,820	
Maximum hourly averaged over selected days	1,745	1,745	Chao et al, 2005
Top 1% of hourly volume over all selected days	1,775	1,880	Jia et al., 2010
Volume associated with maximum occupancy in fundamental diagram	1,825	1,810	Van Arem & Van Der Vlist, 1992

Figures 5-3 illustrates how the Rakha model fits the observed data. The parameters that can be estimated based on this model are the capacity (the apex of the fitted model), jam density, free-flow speed, and speed at capacity.

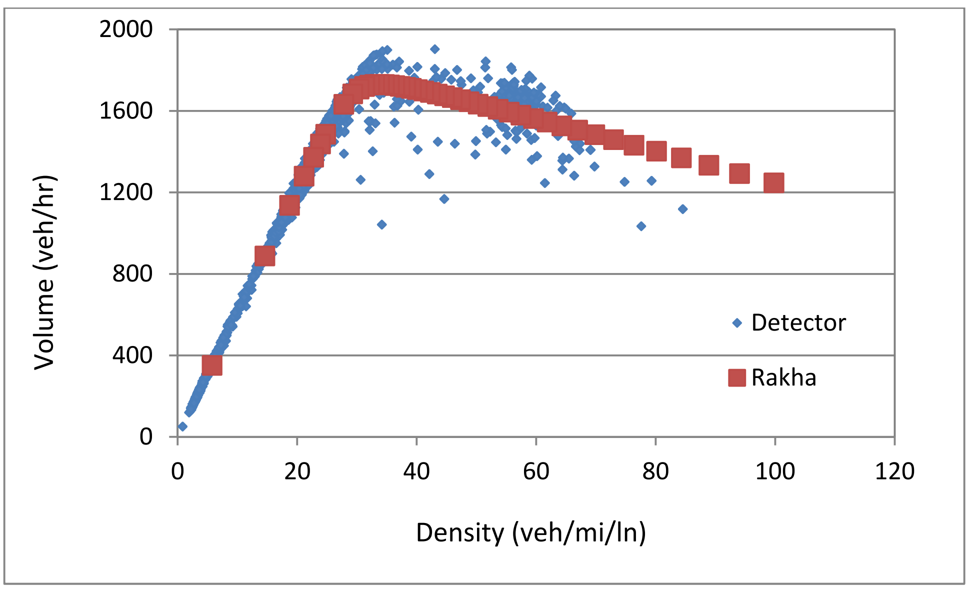


Figure 5-3 Rakha model fitting for capacity and TFM parameters estimation

Figure 5-4 shows how the pre-breakdown flow and queue discharge rates were identified. In this study, the average of flow rates in three intervals before the speed drops due to breakdown is considered as the pre-breakdown flow.

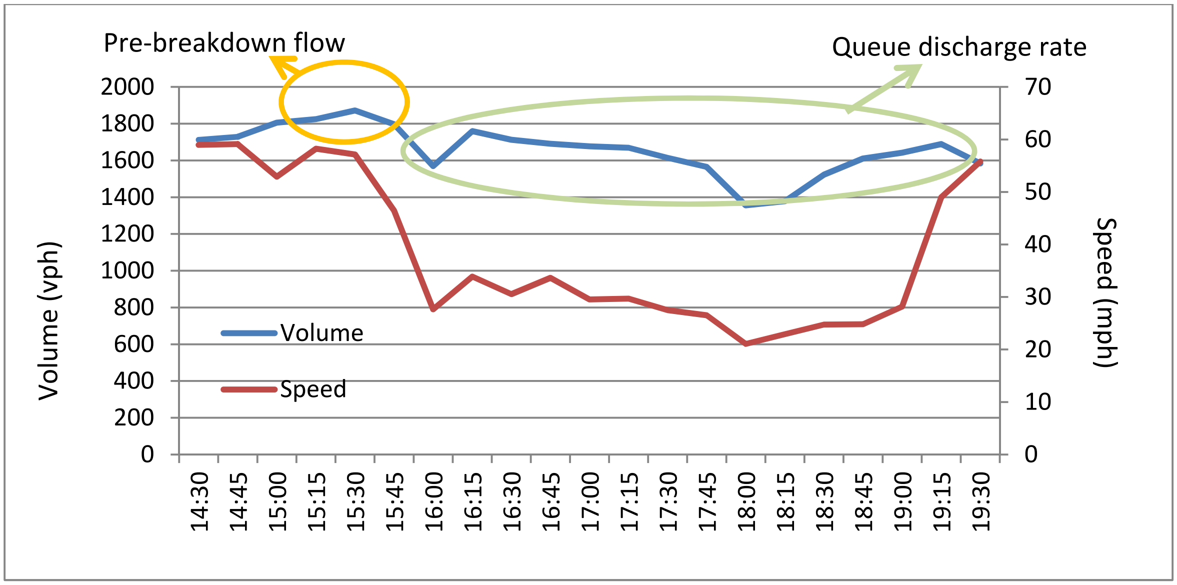


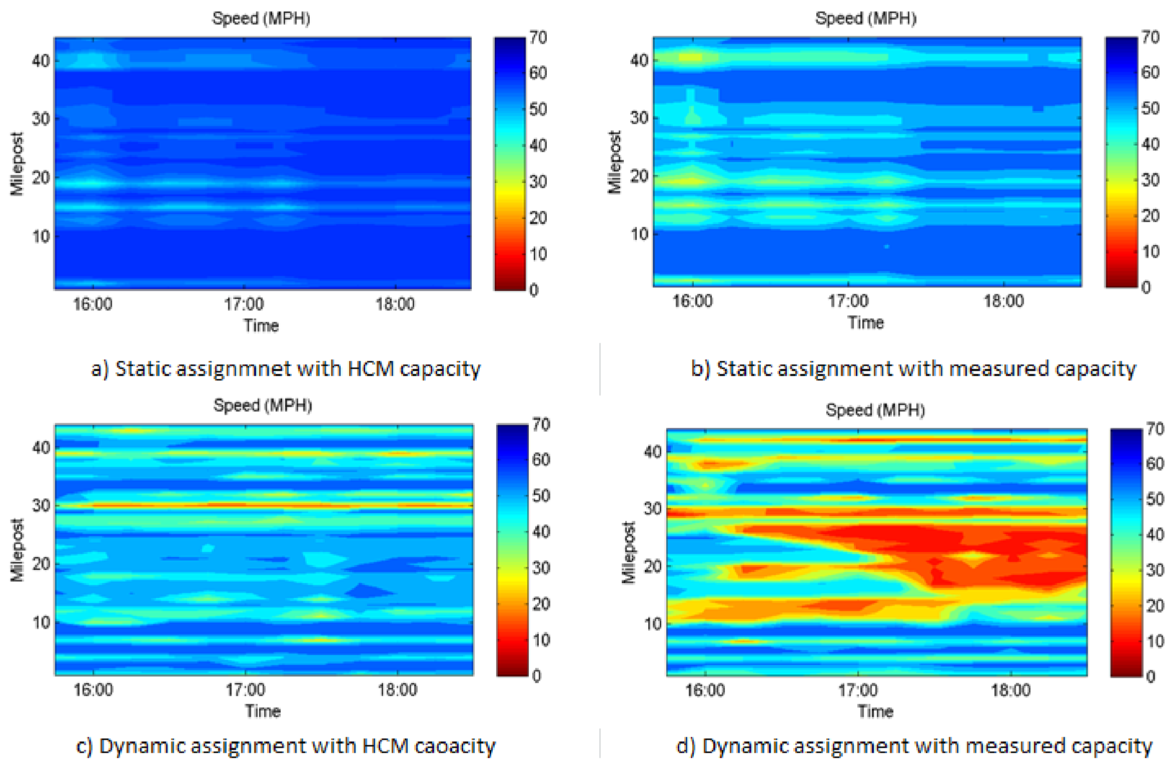
Figure 5-4 Demonstrating breakdown flow and queue discharge rate

The capacity values discussed above are for the general purpose lanes of the corridor's cross-section. As stated earlier, I-95 also includes managed lanes that are separated from the general purpose lanes by soft barriers. Since congestion is avoided in the managed lane (ML) by toll value, there are not enough observations to estimate capacity from the real world. Based on literature, 99.5% of observed volume can be used as capacity. In this study, this value is almost 1,700 vph. Washburn et al. (2010) mentions capacity values ranging from 1,600 vph to 2,100 vph for existing managed lane facilities across the country.

5.4 Coded Capacity Impacts

The purpose of the discussion in this section is to illustrate the importance of coding capacity values estimates based on field measurements as input into dynamic traffic assignment (DTA) tools, particularly when there is evidence that the modeled corridor capacity is lower than the HCM-based estimates. It also demonstrates the shortcomings of utilizing static assignment for assessing managed lane utilization, even when the correct capacity values are coded, and subsequently illustrates the need to utilize DTA modeling for such assessments.

To illustrate the difference in the performance of different traffic modeling approaches, the volumes on the general purpose lanes and managed lanes were forced, in all modeling approaches, to resemble as much as possible real-world measurements based on detector data. For these fixed volumes, this study compared the travel times estimated based on the traffic flow models in static assignment with HCM-based capacity, static assignment with measured capacity, DTA with HCM-based capacity, and DTA with measured capacity. Figure 5-5 shows the speed contour maps of the modeling results. This figure clearly shows that the only model that was able to replicate the real-world bottlenecks at Stations 12 and 20 was the DTA with measured capacity.



a) Static assignmnet with HCM capacity

b) Static assignment with measured capacity

c) Dynamic assignment with HCM caoacity

d) Dynamic assignment with measured capacity

Figure 5-5 Speed contour maps for static and dynamic traffic assignment with different capacity values

Figure 5-6 shows the difference in travel time between general purpose lanes and the managed lanes for the four modeling approaches. This figure confirms that the only model that could show the congestion observed in real-world conditions is the DTA model with the measured capacities. In static assignment, no queuing is assessed and the travel time is calculated based on a simple BPR curve. The change in the value of the capacity in static assignment does not have a significant effect on the modeling results. It is also important to point out that in the DTA tool, when using the regional network capacity, no queue is formed; therefore, the results are similar to the static assignment tool.

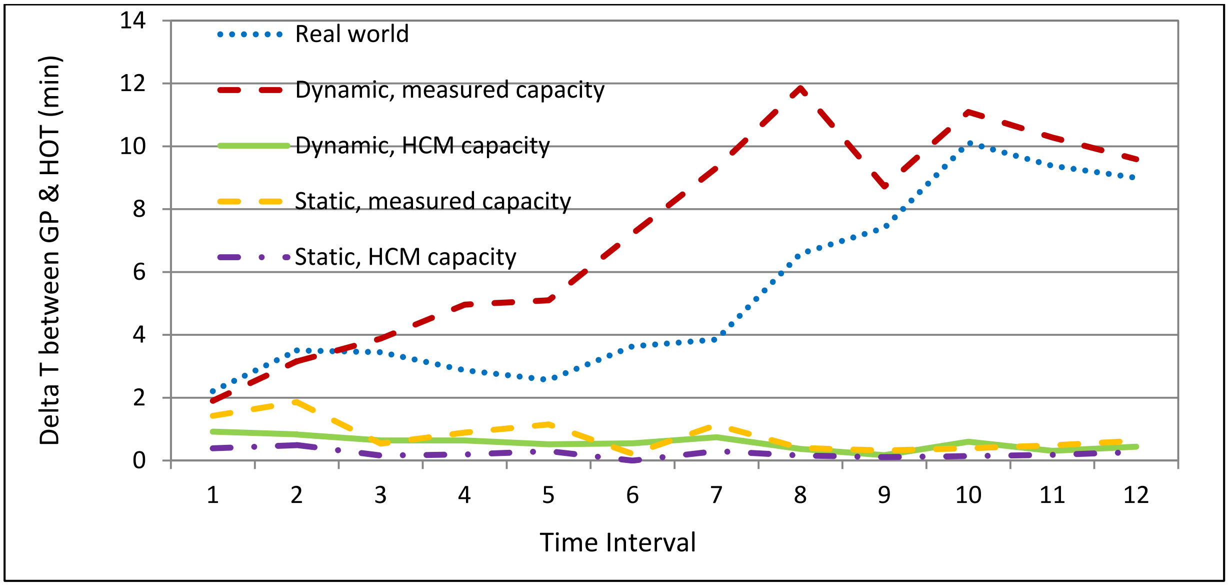


Figure 5-6 Travel time difference between general purpose and managed lanes

The findings above are important because the difference in travel time between general purpose and managed lanes is used in the modeling process to assess the proportions of traffic that utilize the managed lanes, either based on user equilibrium assignment, a willingness-to-pay table, or a logit model combined with the assignment. This importance is further illustrated by feeding the difference in travel time results from Figure 5-6 to a willingness-to-pay table derived in a previous study (29), so as to determine the change in the estimated percentages of traffic willing to use the managed lane. Assuming a $1 toll for this segment, the percentage of drivers who are willing to pay the toll is calculated based on the willingness-to-pay curve. This calculation is based on toll value (in cents) divided by the saved travel time (difference between general purpose and managed lane travel times). As is shown in Figure 5-7, the only model that was able to produce the expected results is the DTA model with the measured capacity.

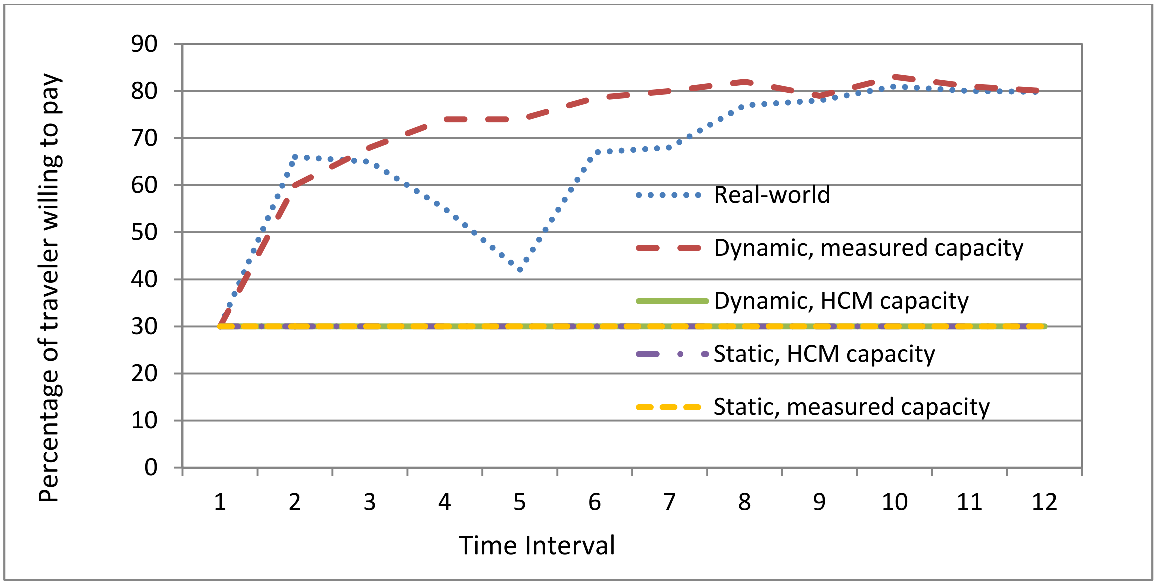


Figure 5-7 Percentage of travelers diverting to managed lane

It is worth noticing again that one of the congestion spots in this network is caused by a spillback from an off-ramp that causes low speeds in the two left lanes (the I-95 Northbound off-ramp to the Turnpike). Since the utilized DTA tool (Cube Avenue) does not support lane-by-lane modeling, it is not possible to correctly replicate that location, because the queue in the model first fills up the whole segment (including 5 lanes) before backing up to the upstream link. In the real-world, only the two left lanes are blocked. If replicating the congestion at such locations is important to a study, a tool that better handle this situation or multi-resolution analysis should be considered.

5.5 Other Traffic Flow Model Parameter Estimation

The Cube static assignment utilizes the widely used BPR traffic flow model to estimate the travel times during assignment for the whole analysis period, normally a peak period in case of time-of-day demand forecasting. On the other hand, Cube Avenue utilizes a mesoscopic simulation model to estimate the system performance at short time intervals during the simulation. The model generates individual vehicles and models and their interactions based on a TFM, with the performance further assessed using queuing analysis. Although the default traffic flow model is the BPR, the Cube script provides the flexibility to implement any desirable TFM. It should be emphasized, however, that in Cube Avenue, the TFM only affects travel time calculation when demand is below capacity. After queue formation, the delay values are calculated based on queuing analyses and can only be affected by adjusting the link capacity and storage parameters by the user. In other words, travel time is divided to two parts of moving on the link, and waiting at the link entrance gate due to capacity or storage restrictions. TFM affects the moving time, but the waiting delay is calculated internally.

Figure 5-8 shows the effect of implementing different TFMs on travel speed at the bottleneck location. Akcelik, Van Aerde, Greenshields and BPR curves are compared in the figure. It can be seen that during congestion the travel time for all TFMs is almost the same, but before and after breakdown, the travel time is slightly different, with BPR producing the lowest value.

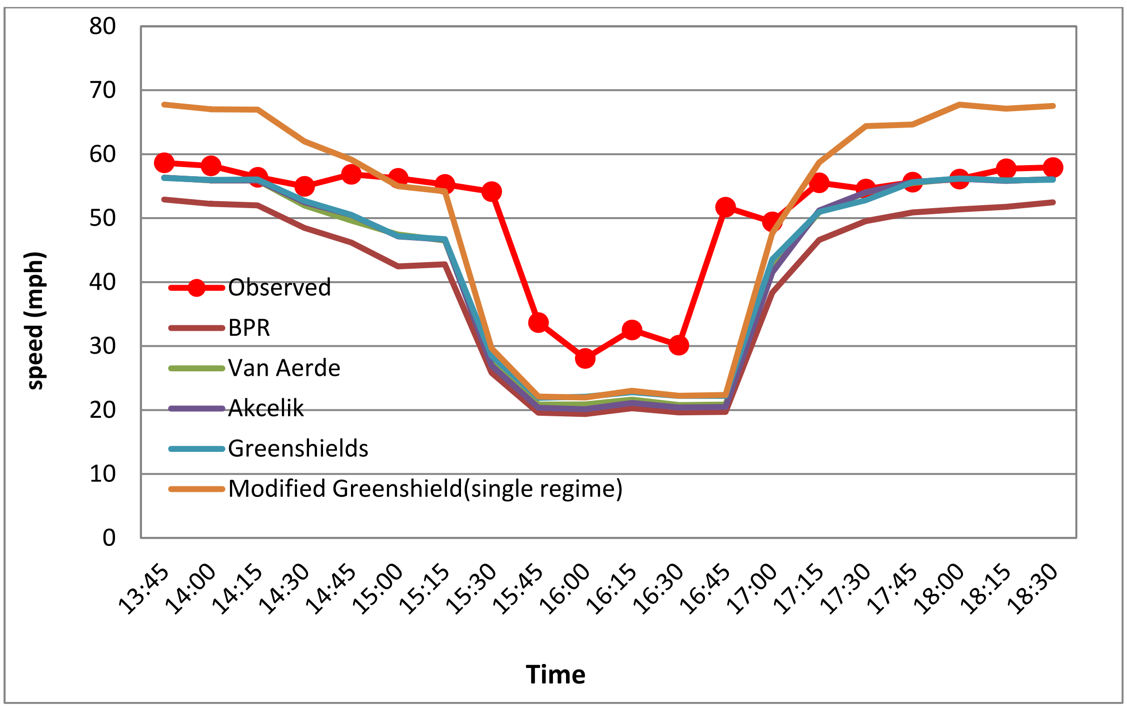


Figure 5-8 Effect of implanting different TFMs on travel speed

In Cube Avenue, it is documented that storage, along with capacity, are two constraints that limit the number of vehicles entering a link. The default value used in Cube Avenue is 190 veh/h/ln. This value is in the range of jam density rather than queuing density. Jam density is different from queuing density. Jam density is the density when all vehicles are stopped, while queuing density is the density of a moving queue. Coding the storage as jam density produces congestion spots with very low speeds (2 to 3 mph). ITS data however, shows a higher minimum speed in congested areas. In other words, cars move within queue, with a speed of 12 to 15 mph. This suggests that the storage should not be considered as jam density (completely stopped vehicles in a very congested network), but the queuing density should be used. This density can be calculated by dividing the volume by speed at the congested segment. This value is almost 3 times smaller than the jam density. By applying this value, the minimum speed increases and more closely resembles the observed speed. Queue length also more closely resembles the real world.

The quality of the supply calibration is evaluated based on performance measures. Primary performance measures that evaluate how well the network replicates a real-world situation are link volume versus observed counts, and link speed versus measured speed. Several goodness-of-fit tests were suggested to measure the distance between simulated and observed volume. Table 5-2 represents the most common goodness-of-fit measures that are used to assess network calibration.

Table 5-2 Goodness-of-Fit Measures

GOF	Formula
Root Mean Square Error (RMSE)	$\sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{N}}$
Root Mean Square Normalized (RMSN)	$\sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i y_i^2}}$
Percent Root Mean Square Error (% RMSE)	$\sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n}} * \frac{100 * n}{\sum \hat{y}_i}$
Mean Absolute Error (MAE)	$\frac{\sum_i^n \lvert y_i - \hat{y}_i \rvert}{n}$
Scale	$\sqrt{\frac{\sum y^2}{\sum \hat{y}^2}}$
GEH	$\sqrt{\frac{2(y - \hat{y})^2}{y + \hat{y}}}$

In formulas above, y is the simulated/estimated volume, and $\hat{y}$ is the observed volume. Except for GEH, the above goodness-of-fit measures can be used to calculate the distance between estimated and observed values of other traffic measures such as speed, density, and queue length. GEH is an empirical formula that has been proven useful for a variety of traffic analysis purposes, mainly for volume comparison purposes. A GEH of less than 5.0 is considered a good match between the modeled and observed hourly volumes (flows of longer or shorter durations should be converted to hourly equivalents to use these thresholds). According to the Federal Highway Administration (FHWA), 85% of the volumes in a traffic model should have a GEH less than 5.0 (FHWA, 2007). GEHs in the range of 5.0 to 10.0 may warrant investigation. If the GEH is greater than 10.0, there is a high probability that there is a problem with either the travel demand model or the data.

In congested networks, volume replication must be coupled with speed or density replication. Considering only volume as calibration assessment criteria in such conditions can lead to a network that does not reflect the congestion patterns in the real world. It should be mentioned that detectors can only measure the served volumes, not the actual demands. Once demand exceeds capacity, the served volume starts decreasing with an increasing level of congestion and increasing density. This phenomenon cannot be captured by solely considering the detector volume. Congestion patterns in the model should reflect real-world conditions, spatially and temporally. Speed contour is a strong visual inspection tool for comparing congestion patterns of modeled and observed situations.

It is important that speed-volume profiles (temporal speed and volume for each link) at bottleneck locations should also be replicated. A comparison between modeled and observed speed/volume profiles (similar to Figure 5-4) can be used to evaluate how well the model can replicate the following: starting and ending time of breakdown, speed and volume before breakdown, duration of breakdown, average volume and speed during breakdown, and covered speed and volume when the breakdown period is over.

Overall traffic measures such as VMT, VHT, and VMT/VHT can also be used for general evaluation of the calibration. It should be noted that abovementioned measures should be assessed in a calibrated network with fixed demand (calibrated demand).

6. Origin-Destination Matrix Estimation

6.1 Methodology Overview

Dynamic traffic assignment (DTA) requires trip matrices specified for short time intervals (e.g., 15 minutes or 30 minutes). These matrices are sometimes referred to as time-variant or dynamic. The derivation of these matrices is one of the most challenging aspects of dynamic traffic assignments. These matrices must be derived based on demand matrices that were estimated for longer periods of time by demand forecasting models.

The initial source of demand in this research is a trip table for a peak period, extracted from a regional demand forecasting model (the Southeast Florida Regional Planning Model (SERPM) model in the case of the I-95 managed lane case study). The demand forecasting modeling process is a mature and well established process that produces behaviorally consistent results among different demand forecasting steps, including the trip generation, mode choice, trip distribution, and trip assignment steps. These models are well calibrated based on real-world data and surveys. Therefore, they constitute a rich source of origin-destination (OD) information with inherent consistency among trip generation, distribution, and assignment. These trip tables should be considered an important source of demands. However, the trip tables need to be updated for operational purposes due to the necessity for shorter time intervals for demand and the need for more focused validation of the demands for the subarea under consideration. The demand calibration or estimation step in this project aims to estimate the OD table for short intervals (15-minute intervals in this case) based on an initial matrix obtained from the demand forecasting model. The resulting matrices, when loaded onto the calibrated network, will be able to replicate the measured link volume.

As previously mentioned, the demand estimation procedure can be significantly affected by the utilized network parameters, as well as by route choice (assignment) parameters. On the other hand, calibrating network and assignment parameters requires correct demands. Thus, an iterative approach is needed for estimating the demands and network (supply) parameters.

The first step is to extract an initial OD matrix for the whole peak period from the regional model (in the case study under consideration, this is the three-hour PM peak period from the SERPM model). First, the boundary of the subarea network of interest must be specified. The subarea boundary can be specified using the Cube Polygon feature or a GIS tool. Cube Voyage can then be used to extract the subarea network from the statewide model network using this predefined subarea boundary. The result of this extraction is a subarea network with new node and zone numbers, which are different from the original numbers. Cube stores the association between the old numbers (in the whole network) and the numbers in the new network (in the subtracted network) in two new node features in the subtracted network.

The next step is to convert the three-hour PM peak-period matrices obtained from the SERPM to 15-minute matrices using distribution factors that reflect the proportion of the trip tables for each 15 minutes of the day obtained, based on trip counts. This distribution was conducted to be consistent with the variations in observed volumes at uncongested locations at the beginning of the corridor, where detector volumes can represent actual demands (and not the capacity-restrained served volumes). The availability of these initial 15-minute interval matrices (referred to as factorized matrices in this study) made it possible to start an initial network (supply) calibration, as described in Chapter 5, based on the 15-minute volume and speed data.

The next step is to adjust these matrices using the Cube Analyst static matrix estimation program. This matrix estimation process performs the estimation by considering a number of input parameters based on the static assignment of Cube Voyager. This process applies a maximum likelihood approach to optimize the trip tables based on the deviation between traffic counts and assignment results and initial (seed OD) matrices. Since the STA runs over a single model period, each 15-minute interval must be run separately to estimate the OD matrix for the associated interval. The most important issue with STA in this process is its inability to capture queue spillback and make the resulting connection between consecutive intervals. This problem in the current study is minimized utilizing heuristics to account for queue presence.

The best approach to overcome static traffic assignment (STA) limitations is to use the DTA instead of the STA as part of the least-square optimization to better account for traffic dynamics and travelers' behaviors. Thus, the next step was to use the Cube Analyst Drive procedure, which includes an OD estimation procedure that derives the time-variant trip matrices based on minimizing the differences between the measured volumes and the volumes produced by the DTA, with consideration of initial trip tables resulting from the Cube Analyst estimation based on the STA. However, limitations were identified with the existing tool developed for this purpose, and modifications are proposed to improve the performance of this approach.

During the matrix estimation process, several manual adjustments and iterations were required. As demand changes, the network calibration may need to be slightly changed. The route choice behavior may also need to be adjusted, as described in Chapter 7, as better OD estimates are obtained. Adjustments and fine-tunings are also needed to avoid unrealistic deviation from the initial matrix derived from the SERPM OD matrix estimation. These adjustments are iteratively and continuously performed during the matrix estimation process.

6.2 Static OD Estimation

The factorized 15-minute matrices derived based on 15-minute traffic counts are used in some studies as input into DTA models. However, these matrices can be further refined by utilizing a matrix estimation procedure based on traffic counts. Such a procedure would consider the deviations of the link volumes assigned by the model from traffic count measurements. As stated in the previous section, in this study, the update is performed in two steps. The first step is based on the static assignment of Cube Voyage, and the second is based on the Cube Avenue DTA assignment. This section describes the first step.

The static OD matrix estimation process is implemented using the Cube Analyst program, which is provided as an optional tool within the Cube modeling environment. Cube Analyst is a tool that estimates trip matrices based on the maximum likelihood technique, coupled with an optimization procedure. The tool utilizes data from different sources and considers the different levels of confidence or reliability inputted by the user from these different sources. Not only can the data include vehicle traffic or passenger flow counts and prior (old) matrices, but also partially observed matrices, zonal trip end (generation and attraction) data, vehicle routing, travel cost matrices, and even previously calibrated trip cost distribution functions.

Different sequences of processes for OD estimation were investigated in this study to determine how they impact the model's ability to replicate different performance measures of real-world traffic conditions, required memory and time, and deviation of estimated OD from different sources of data. It was found that the best practice is starting with a factorized matrix, calibrating the network (supply side), followed by static OD estimation, fine-tuning the network calibration, and then fine-tuning the ODs by performing dynamic OD estimation. Static matrix estimation was found to be the most essential step that could not be skipped. Running the dynamic matrix estimation (matrix estimation based on DTA) directly after the factorization step did not produce good results, possibly due to the immaturity of the dynamic OD estimation procedure in Cube Analyst.

Since Cube Analyst is based on static assignment, it deals with only one matrix at a time. Thus, it had to be run twelve times to obtain the twelve 15-minute matrices in the three-hour period. Cube Analyst performs a set of iterative calculations that will automatically determine the statistically, most likely matrix for the set of input data values provided. The input data to Analyst can include the following:

- Screenline counts: These are observed link traffic counts at screenline locations. In other cases where multiple user class matrices are estimated, the aggregated link counts should be split accordingly (i.e., each matrix class should be associated with a class of observed counts). Each screenline can also be associated with a confidence factor. This feature enables the user to define the links with counts that are more important to be replicated, or are associated with more reliable traffic counts. In this study, traffic counts for each 15-minute interval were obtained from ITS detectors and PTMS locations.
- Initial trip tables: One trip table is required for each user class. Each matrix can be associated with a confidence matrix, which contains different confidence level values for each OD pair versus screenline traffic counts. In this study, the initial 15-minute trip tables were obtained from the factorization process described earlier. The screenline counts versus initial O-D trip values can be set to reflect the relative importance or

reliability of these two variables. If there is a high confidence in or desire to replicate the screenline counts and a low confidence on the initial matrix, the relative screenline count confidence should be set to a higher value. The appropriate confidence values can be identified as part of the iterative process of the supply/demand calibration.

- Zonal trip ends: These are the total number of trips originating and terminating in each zone. Each zone can be associated with a confidence factor, based on the level of reliability or importance of preserving the total number of trips. Trip ends are calculated in this study based on summation of the rows and columns in the trip matrices.
- Partial trip table: This optional input enables the user to incorporate any partial OD trips that are available from other sources such as Bluetooth readers, Electronic Toll Collection System, or OD surveys.
- Routing information: This information is provided by the assignment module. This input contains information of ODs that have passed each link.
- Optimization parameters: These are parameters provided to set convergence criteria for optimization, and to set weight that shows users' relative confidence on initial matrix versus screenlines (higher weight shows that the user prefers not to deviate significantly from the initial matrix, even if the screenlines cannot be completely replicated). These parameters are in a "control file" input to Analyst as a text file and the required and optional parameters can be easily edited by the user.

For more information about the required inputs and associated parameters, please refer to the Citilabs Analyst Manual (Citilabs, 2013).

In this study, four groups of matrices are available from the regional demand forecasting model: Drive Alone (DA), Shared Ride of 2 occupants (SRP2), Shared Ride of three or more occupants (SRP3), and Truck. The I-95 ML policy does not differentiate between DA and SRP2 (e.g., both groups should pay the same toll to access ML). Therefore, the DA and SRP2 matrices are grouped together and are referred to as SOV in the assignment module.

As described in Chapter 4, when using the willingness-to-pay approach to modeling ML, travelers are divided into two groups: toll payers and non-toll payers, based on the ratio of toll cost divided by saved travel time. It is assumed that SRP3 can use the ML without any cost or restriction, and trucks are not allowed to use ML. The SOV matrix (summation of DA and SRP2) is split, based on the willingness-to-pay curve, into two groups: SOV_wo_Toll (non-toll payers) and SOV_w_Toll (toll payers). The SOV_wo_Toll and trucks are not allowed to use ML, but the other user classes choose between GPL and ML, based on the generalized cost function. The routing information is saved in binary "intercept" files associated with each user class. The routing information, matrix, and screenline for each user class matrix are used in Analyst. Thus, the aggregated link counts acquired from detectors were split accordingly into four user classes of SOV_w_Toll, SOV_wo_Toll, SRP3, and Trucks.

In the specific case of ML, which is the main interest of this study, the route choice behavior is highly complicated and has more parameters to estimate. Before running Analyst, the assignment process should be checked to confirm that it is able to roughly estimate the portion of travelers that divert to the ML. If traffic assignment parameters, such as the willingness-to-pay curve, are not calibrated in this stage, the results negatively affect the OD estimation process. This creates another challenge, since a good assignment calibration requires a good demand estimation and vice versa. The network or supply calibration also affects the results. Thus, an iterative process is needed, but there is a need to ensure that the final OD matrix estimation is based on relatively well-calibrated network parameters and well-calibrated assignment parameters.

Another major consideration is capacity-constrained demands on congested corridors. Analyst is a robust optimization module that aims at replicating screenline volumes. However, in congested locations/periods, these volumes are the capacity-constrained served volumes. Thus, replicating these volumes based on counts will underestimate the demands. For example, Northbound I-95 in the PM peak is a congested corridor, and as a result, the static OD estimation failed to produce the correct demands during the congested periods. This problem may be solved by incorporating traffic measures that account for congestion presence, such as speed, density, or queue in the optimization tool. Due to the absence of these features in the current version, a method was developed to calculate the queue length, which was added to the traffic counts in the screenline file.

The queue length on each link was estimated based on the level of congestion identified from the detector data, which is in turn converted to the number of queued vehicles. These values were added to the screenline volume count, and the static OD estimation was run again. The resulting demand was the input to the Cube Avenue module, and it was confirmed that it could better replicate real-world congestion patterns.

A new tool called Analyst Drive was recently developed by Citilabs. Analyst Drive can be used for estimating OD matrices based on static and dynamic assignment (There is a keyword in the control file as "OD TYPE". Setting this value to 0 runs static estimation, and setting a value of 1 runs dynamic assignment). In this study, Analyst and Analyst Drive were both run for static OD estimation with the same input and with the default parameters. Figure 6-1 displays the demands for a specific OD pair over 12 intervals. Figures 6-2 to 6-4 present the real-world replication of mainline volumes when utilizing factorization, Analyst and Analyst Drive, respectively. Figure 6-5 compares the flows of one specific origin to all destinations, in the initial OD matrix and the estimated ones by factorization, Analyst and Analyst Drive. These figures show that Analyst tends to focus on replicating screenline counts, sometimes at the expense of deviating significantly from the initial matrix. The figures also show that the utilization of an OD estimation process of the types used in this study improves the results compared to using simple factorization.

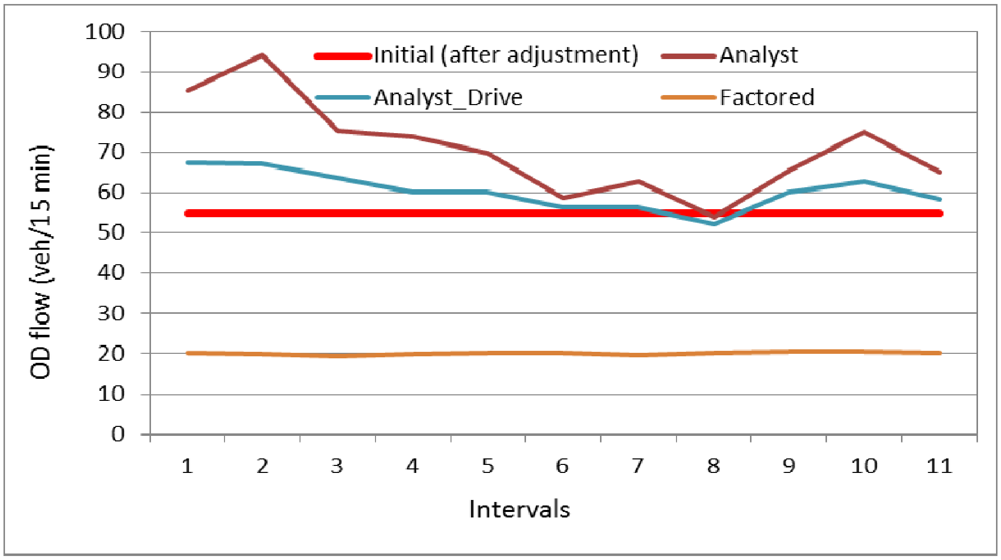


Figure 6-1 Temporal profile of initial, Analyst, and Analyst Drive OD

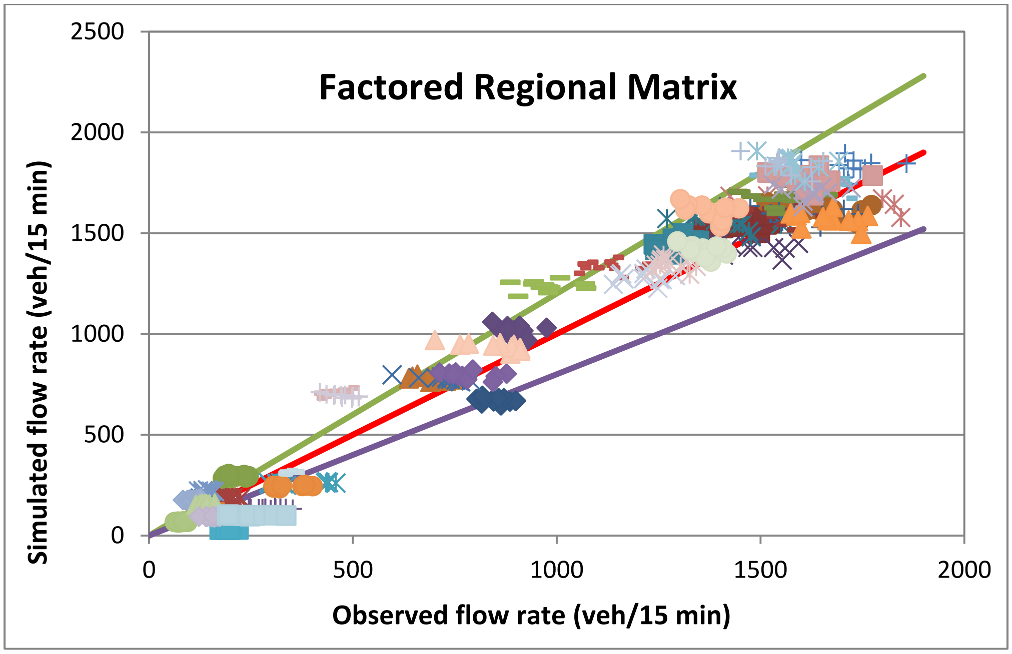


Figure 6-2 Screenline volume replication by factored regional matrix

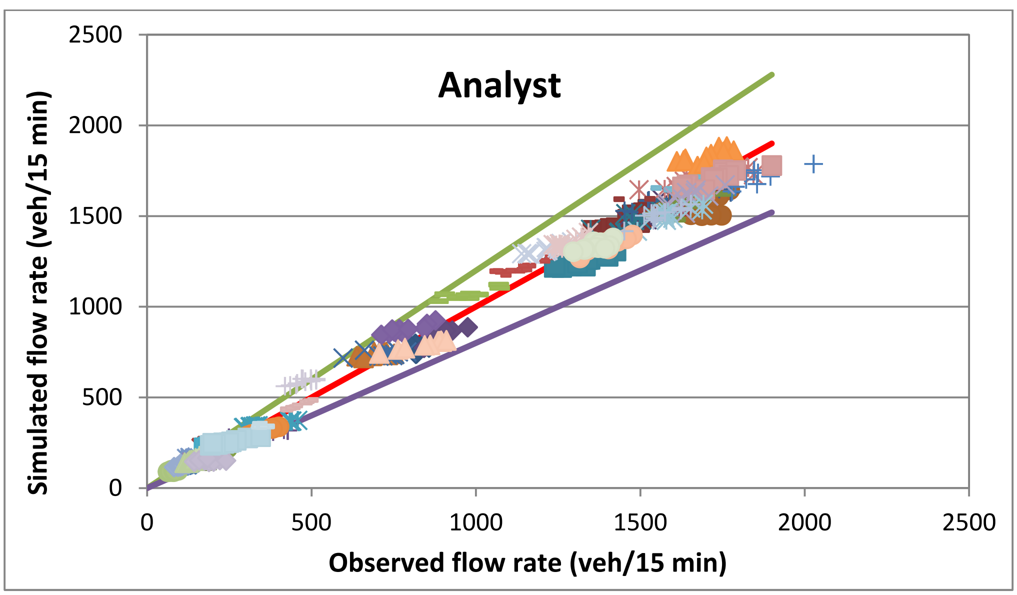


Figure 6-3 Screenline volume replication by Analyst

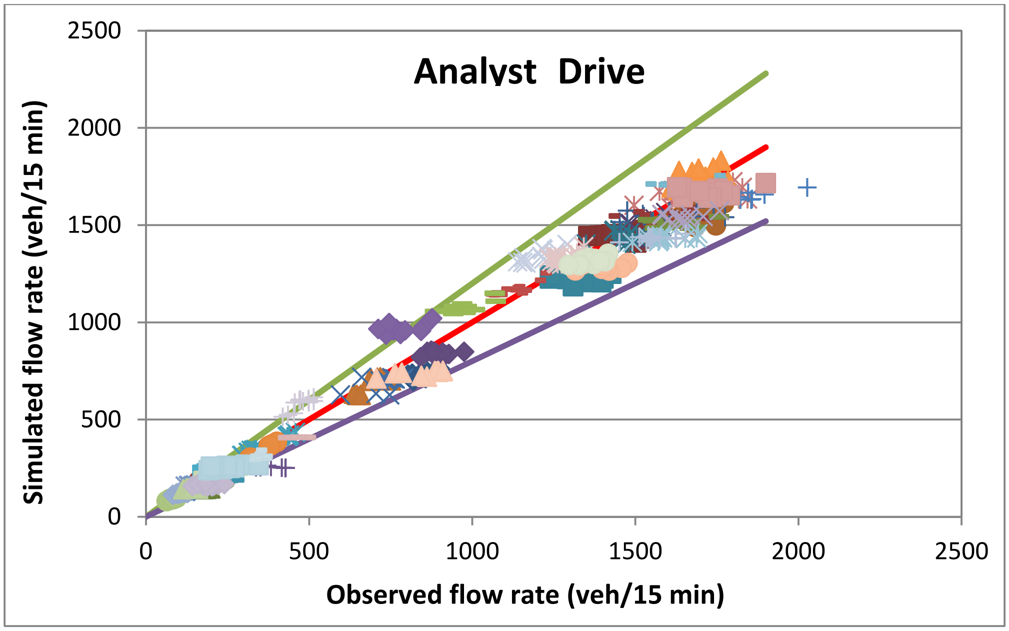


Figure 6-4 Screenline volume replication by Analyst Drive

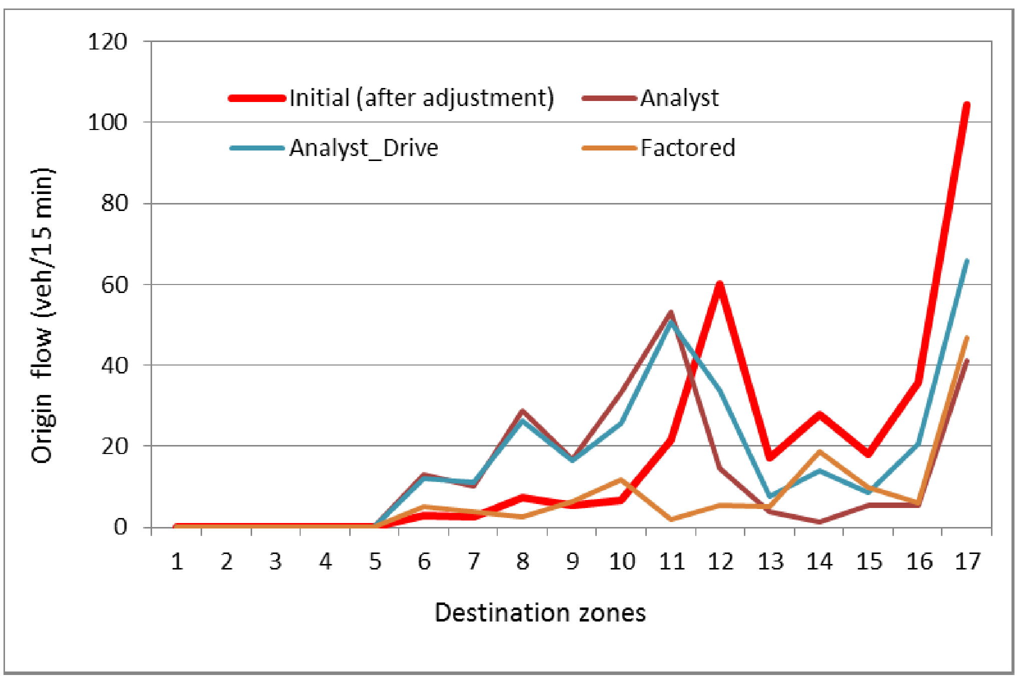


Figure 6-5 Comparing initial and estimated OD for one pair

6.3 Dynamic Matrix Estimation

The 15-minute matrix estimation that uses Cube Analyst is expected to represent significantly better demand estimations than the ones produced from the factorized matrices that were used as inputs to the Cube Analyst (in the estimation process). However, the Cube Analyst process utilizes demands from the static assignment during the optimization process. Thus, this study also investigates the use of the Analyst dynamic OD estimation process that utilizes traffic volumes from Cube Avenue in the optimization process.

The dynamic OD estimation follows a very similar process, as described in the previous section. Instead of running Analyst with Highway assignment module for 12 consecutive periods of 15-minute intervals, Analyst Drive (the dynamic OD estimation tool) runs during the whole model period, coupled with the Cube Avenue assignment module. In the single Cube Avenue run, the model period is divided into 15-minute intervals. This procedure is supposed to be superior to static OD estimation, because Cube Avenue models the queues and queue spillbacks, and thus can capture the effects of congestion on subsequent time intervals. However, without incorporating density or speed, dynamic OD estimation may also underestimate the demands under congested conditions. The dynamic OD estimation module in Cube package is not as mature as the static matrix estimation module, and has the following limitations:

- Analyst Drive for dynamic OD estimation does not incorporate zonal trip ends. Zonal trip ends are usually available from ramp count data and provide valuable, reliable data about origin and destination.
- Partial trips cannot be incorporated into the current version of Analyst Drive for dynamic OD estimation.
- The confidence matrix associated with each input matrix cannot be incorporated into Analyst Drive for dynamic OD estimation.

Due to the aforementioned limitations and the results of running dynamic OD estimation in this study, it is suggested that the user should use caution when utilizing the dynamic OD estimation module in Cube Analyst.

6.4 Performance Measures and Matrix Adjustment

When calibrating simulation, demand, and assignment parameters, a distance function between simulation outputs and field measurements is minimized. This function can include different measures, such as link volumes, OD demands, link speeds and/or densities, etc. Limiting the function to replicating link volumes, as is the case in many studies, can be misleading and fail to produce the correct demands or congestion patterns. Most OD matrix estimation methods are based on link traffic volumes and initial OD matrices. If enough data on speeds, densities, queue lengths, OD routes, or zonal trip end rates are available, they should be incorporated into the calibration process to better replicate real-world traffic conditions. There are different ways to incorporate this information into the calibration process. They can be included in the objective function of the optimization or be a part of a manual adjustment or a heuristic procedure outside the optimization tool.

As mentioned in Chapter 2, demand estimation is an underspecified problem. It means that the number of equations (the link counts) is usually much lower than the number of unknowns (OD pairs). Hence, different OD estimates may produce the same link volumes. It is important, therefore, to manage the estimation process to ensure the reasonableness and the correctness of the estimated demands. Regarding the dependency between demand estimation and traffic assignment, a wrong estimation of OD pair demands can sequentially propagate during the calibration process. In order to limit the systematic errors in OD estimation, the consistency and reliability of the adjusted OD pairs should be checked against different sources of data, such as trip end rates or specific route volume information. Following is a list of criteria that were identified in this study to justify manual adjustments of the estimated demands:

- Deviation from the initial matrix: It might be helpful to preserve certain structure or information that the initial matrix (subtracted from regional forecasting model) contains, such as the proportion between the total trips of the DA, SRP2, SRP3 and Truck user classes. Another example is to not modify the split between two major destinations in the network, such as I-95 and the Florida Turnpike northbound in the test network.
- Route information: There might be reliable information about specific route trips, which are necessary to replicate.

- Zonal trip end: On-ramp and off-ramp counts, in the absence of queues, can be reliable sources for origin and destination demand estimation, particularly in the case of linear corridor modeling. Thus, replicating these counts justifies the manual adjustment of the OD volumes.
- General temporal uniformity: There are no expectations of seeing unrealistically high rises or drops in the volumes of OD pairs in sequential intervals. In the Analyst optimization tool that is based on static assignment, the temporal variation cannot be controlled. The optimization process can achieve totally different local optimal solutions for sequential intervals, since the optimization does not guarantee achieving global optimal. To minimize the OD matrix variations between sequential intervals, the static OD estimations for different intervals were run with an identical initial matrix. After several OD estimation trials and matrix adjustments, one matrix was selected as a good initial matrix.

Since the demand estimation is underspecified and may result in a local minimum, it may be helpful to force the optimization to start the search from a certain point, more specifically, to restrict some of OD pairs from varying during the optimization. Manually adjusted values should be inserted in the process again for a new run of the OD matrix estimation. Different approaches can be used to combine the estimated and adjusted values to control the deviation from the general structure of the initial matrix, such as Kalman filtering, Bayesian inference, and MSA. The adjusted and combined values will then be fed back into the estimation process. Modifications to the existing OD estimation process are recommended so as to allow the user to have the flexibility required to incorporate additional factors as limiting criteria in the objective function (based on the analyst's knowledge), to avoid the need for manual adjustments.

6.5 Recommendations for OD Estimation Improvement

Additional recommended improvements to the Analyst OD estimation process are listed as follows:

- Incorporating speed, density, and/or queue length in the objective function of Analyst.
- Allowing the user to specify lower and upper bounds for each OD pair cell (there is already a global parameter that is applied to all cells yet cannot be varied by cells).

- Allowing the user to better control the temporal variability of the results.
- Allowing the user to keep the proportionality between specific OD pairs (e.g., from all of the trips originated from I-95, with 30% directed to SR 836 and 30% destined to the Florida Turnpike).
- Incorporating zonal trip end, partial matrix, partial trips, and confidence matrix in Analyst Drive for dynamic demand estimation.

7. Traffic Assignment

Traffic assignment is the last step in conventional demand forecasting models. In this step, the most common approach is to allocate the vehicles into different paths so that all of the vehicles departing from the same origin and arriving at the same destination experience the same travel time. This approach of assigning trips to available routes is referred to as "user equilibrium" or UE, as discussed in Chapter 2. In static assignment, as the name implies, the demand, network variables, the resulting travel times, and route choice behaviors are static and not time-variant. In congested networks, however, it is clear that capturing the time-dependent characteristics of the demands and network parameters are necessary for accurate modeling and calibration. This is accomplished in dynamic traffic assignment (DTA), in which the user equilibrium is sought for each time interval based on the dynamically changing traffic conditions and operation strategy.

As mentioned in Chapter 2, unlike static traffic assignment (STA), which defines the shortest paths and allocates all of the traffic to these paths all at once for the whole peak period, DTA conducts the traffic assignment with the goal of reaching equilibrium for each time interval that is specifically far shorter than the model period. DTA can model time variant demands, time variant operational strategies, such as those applied for managed lanes (ML), associated travelers' responses, dynamic variations in network performance, and dynamic events such as lane blockage incidents. In addition, simulation-based DTA can model queue building and dissipation and the queue spillback to upstream links by accounting for demands exceeding link capacity and queues that exceed downstream link queuing capacity, as happens in real-world conditions. Therefore, DTA provides a more realistic representation of travelers' behaviors and traffic conditions, and provides a better approach for assigning traffic and estimating travel cost and time, resulting in better demand and performance measure forecasting.

The tests and comparisons presented in this chapter were obtained based on the I-95 linear network, which was used as a case study in this report, and calibrated as described in previous chapters of this document. The origin-destination (OD) tables for 15-minute intervals were obtained based on the OD estimation procedure presented in detail in Chapter 6. For a better comparison between STA and DTA ability to model ML, the STA is run 12 times, one for each of the 15-minute trip table during the PM peak period. Therefore, the output file contains volumes and speeds for 12 time intervals that are used in the comparison with real-world data and DTA results. It should be mentioned, however, that these test runs are independent from each other, and the run for one interval is not affected by the results of the previous interval because STA is not capable of modeling these interactions between time intervals.

Two different approaches were investigated for ML assignment. The first approach involves adding the equivalent value of time of the toll cost value to the travel time function within the assignment, resulting in a generalized cost that considers the ML toll. In this approach, referred to as "Generalized Cost Function" approach in this study, vehicles use of ML is solely governed by the user equilibrium UE assignment procedure, based on the generalized costs of the competing paths. The second approach is referred to as the "willingness-to-pay curve" approach, described in Chapter 4 when discussing the Cube Avenue prototype. In this method, prior to the assignment, travelers are divided into two groups: a group that will not choose to pay the toll and is limited to using the general purpose lanes (GPL). The other group is eligible to use the ML lanes based on the willingness-to-pay curve, but the final decision to use either ML or GPL depends on its origin and destination points (if there is a managed lane in their paths) and on the difference in the generalized costs between ML and alternative routes according to the UE process.

In performing the ML assignment utilizing both of the abovementioned methods, the toll is updated for each interval, based on the maximum density of the ML so as to preserve the desired level of service in this facility, as is done in the real-world implementation of I-95. The schedule of the value of toll based on density was calibrated based on available charged toll data from the Florida Department of Transportation (FDOT) District 6, as well as based on intelligent transportation systems (ITS) volume and speed data. Another important parameter of the assignment that needs to be calibrated is the value of time versus the toll cost used in the generalized cost function and willingness-to-pay curve. The remainder of this chapter is structured as follows.

As stated above, the main two parameters calibrated in this study are the value of time versus cost and the toll schedule as a function of density. The calibration was performed based on sensitivity analyses results of these parameters.

7.1 Derivations from Observed Data

Implemented toll data, coupled with ITS data, can also be used to calibrate the toll-density curve (table). Table 7-1 includes the default toll values used in the prototype developed in Chapter 4. This table is a simplified version of the table that FDOT District 6 TMC uses to calculate the toll and it does not totally replicate the current I-95 toll table. It was found that this table overestimates the toll values. Figure 7-1 demonstrates the difference between the real-world charged toll and the calculated toll values based on the default toll-density curve for May 11, 2010. The calculated toll was obtained by estimating the density as the volume over speed according to the relationship between the three variables. The density was calculated at each ITS detector along the managed lane for each of the 15-minute modeling intervals. The maximum density value along the seven-mile length of the managed lane was then used to calculate the toll costs, based on the default toll-density table used in the prototype (Table 7-1).

Table 7-1 Default Toll Values Based on the ML Maximum Density

LOS	Road Density		Toll Cost ($)	
	Minimum	Maximum	Minimum	Maximum
A	0	11	$0.25	$0.25
B	12	18	$0.50	$1.25
C	19	26	$1.50	$2.75
D	27	35	$3.00	$3.75
E	36	45	$3.75	$6.00
F	>45		$6.00	$7.00

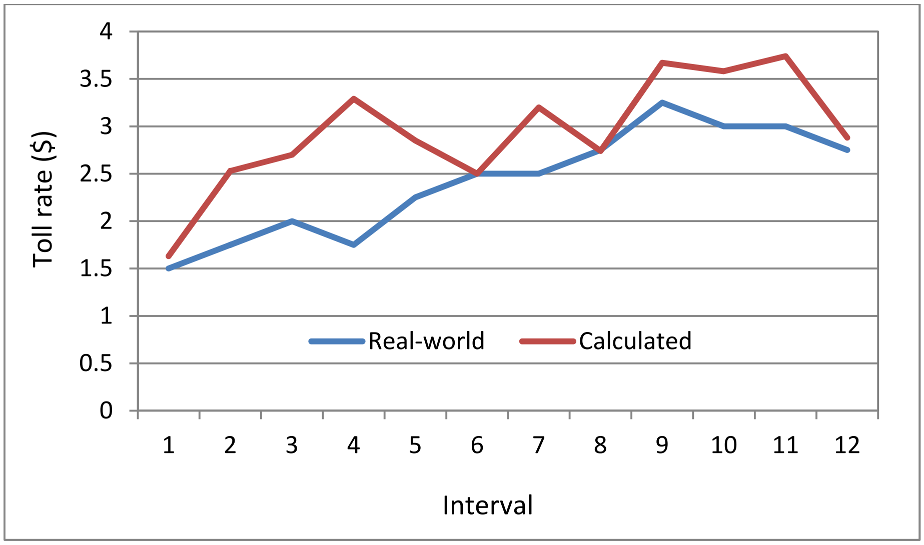


Figure 7-1 Comparison between implemented and calculated toll for a lightly-congested day

Figure 7-2 and Table 7-2 demonstrate the same comparison for severely congested intervals. The values for the calculated tolls are derived from Table 7-2. It is obvious that the utilized toll table overestimates the toll values.

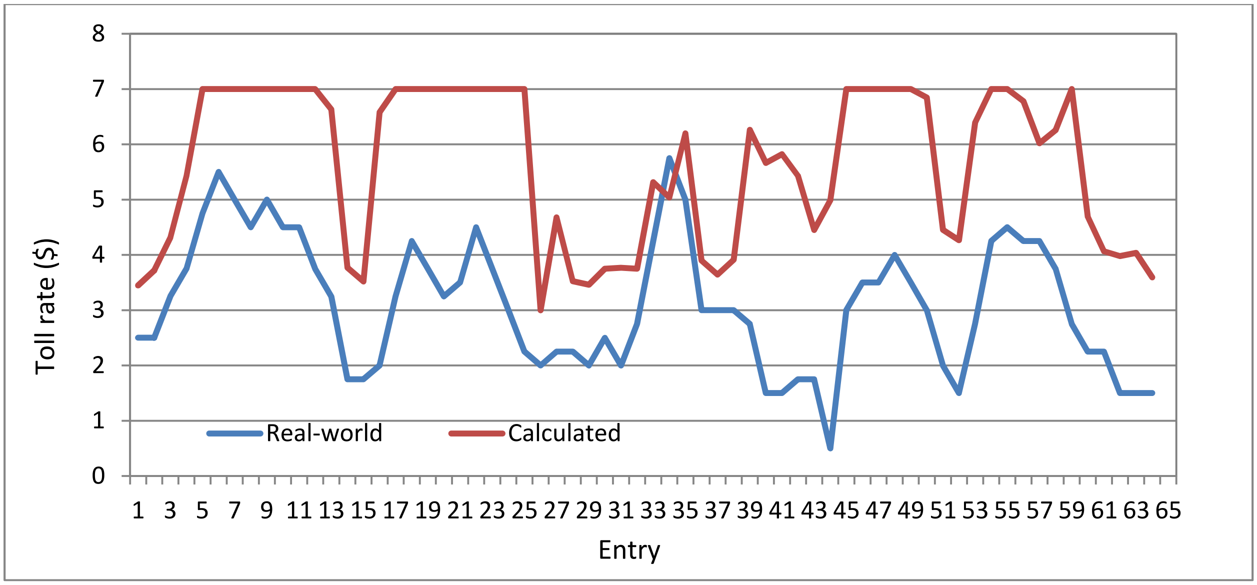


Figure 7-2 Comparison between implemented and calculated toll for highly congested intervals

Table 7-2 Implemented Toll Value for I-95 Northbound

Day	Time	Toll Rate ($)	Maximum Density from ITS Data	Calculated Toll
6/3/2010	3:41:00	2.5	31.46	3.45
6/3/2010	3:56:00	2.5	33.21	3.72
6/3/2010	4:11:00	3.25	37.01	4.31
6/3/2010	4:26:00	3.75	44.22	5.43
6/3/2010	4:41:00	4.75	60.84	7
6/3/2010	4:56:00	5.5	59.04	7
6/3/2010	5:11:00	5	66.1	7
6/3/2010	5:26:00	4.5	79.53	7
6/3/2010	5:41:00	5	77.43	7
6/3/2010	5:56:00	4.5	68.13	7
6/3/2010	6:11:00	4.5	67.29	7
6/3/2010	6:26:00	3.75	60.64	7
6/3/2010	6:41:00	3.25	51.99	6.63
6/8/2010	3:41:00	1.75	33.51	3.77
6/8/2010	3:56:00	1.75	31.91	3.52
6/8/2010	4:11:00	2	51.65	6.58

Based on the abovementioned data, a new toll-density relationship was developed as shown in Figure 7-3. It should be noted that the toll schedule in real world, is based on a more complex lookup table that is difficult to implement. To avoid complications, a simplified toll-density curve is developed in this study based on calibrating to real-world data. Figure 7-4 shows that the toll-density curve developed based on the observed data better replicates the real-world diversion to the ML. It should be noted that the results presented in Figure 7-4 are for a model with calibrated willingness-to-pay parameters, as described in Sections 7.2 and 7.3.

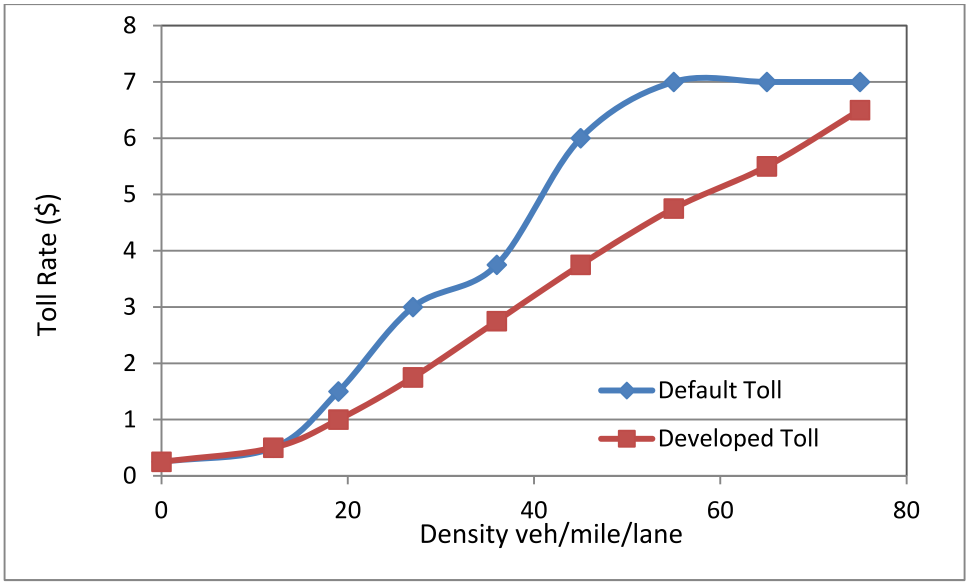


Figure 7-3 Default and developed toll density curve

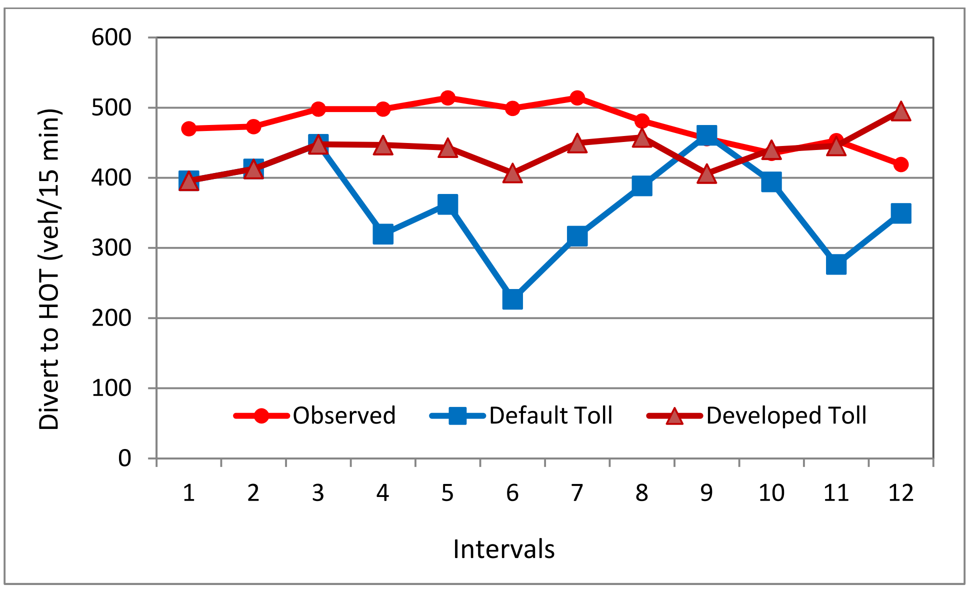


Figure 7-4 Comparison of diverted volume to managed lane for different toll curves

7.2 Calibrating the Value of Time

In this section, a discussion is presented of the calibration of the value of time in the generalized cost function, to better replicate the real-world observations. It was found that for the PM peak period, the average toll cost over several days in 2010 (excluding weekends) is between $2 and $3, with an average of $2.30, as presented in Table 7-3.

Table 7-3 Implemented Toll Value for I-95 Northbound

Time (PM)	day 1	day 2	day 3	day 4	day 5	day 6	day 7	day 8	day 9	day 10	Average
3:26	1	1.5	1.5	1.5	1.75	1.75	1.75	1.75	2	1.75	**1.63**
3:41	2	2.5	2	2.25	1.5	1.75	2	2.75	2.5	2.25	**2.15**
4:11	1.5	3.25	1.75	1.75	1.75	2	2	2	2	2	**2.00**
4:26	1.5	3.25	1.75	1.75	2	2	2.25	2	2	2	**2.05**
4:56	1.5	3	1.75	1.75	2	2	2.25	2.25	2.25	2	**2.07**
5:11	1.5	3	1.75	1.75	3	2	2.25	3.25	2.5	2	**2.30**
5:26	1.5	3	2	2.5	3	2.25	2.25	2.75	2.75	2.25	**2.42**
5:41	1.75	3.5	2	3	3.5	3.25	2.5	2.75	3	3	**2.82**
5:56	1.75	2.5	1.5	3	3.75	3	3	3.25	3.5	3	**2.82**
6:11	1.75	3.75	1.5	3.5	3.5	2.75	2.75	3.25	3	2.75	**2.85**
6:26	1.5	3.75	1.5	2.75	2.75	2	2.75	3	2.5	2	**2.45**
6:41	1.5	3	2.25	2.25	2.25	1.5	2.25	2.25	2.5	2	**2.17**

The time saved by motorists based on real-world detector data is 4-8 minutes, depending on the congestion level in the GPL for the day under consideration. By examining the Southeast Regional Planning Model (SERPM), it became apparent that it underestimates the value of time to motorists. This is confirmed by another study conducted by URS that involves the analysis of I-95 ML. The value of time in that study was estimated to be $42 compared to $12.6 used in the SERPM model. This difference can also be interpreted as the perceived benefits of using the ML beyond the absolute difference in travel time between ML and GPL. This means that if a value of time of $12.6 is used, a bias component toward using ML or a penalty for using GPL should be included to account for other perceived factors affecting travelers' choices, such as perceived reliability comfort and safety on ML.

In this study, the values of travel time used in sensitivity analysis are $12.6, $18.0, and $31.0, and $42. The results are displayed in Figure 7-5. From this figure it appears that values of time of $42.00 produce good results. It should be noted again that this value accounts for factors other than toll and saved time, such as travel time reliability, comfort, safety, and average of past days selection, which includes more congested days and incident days.

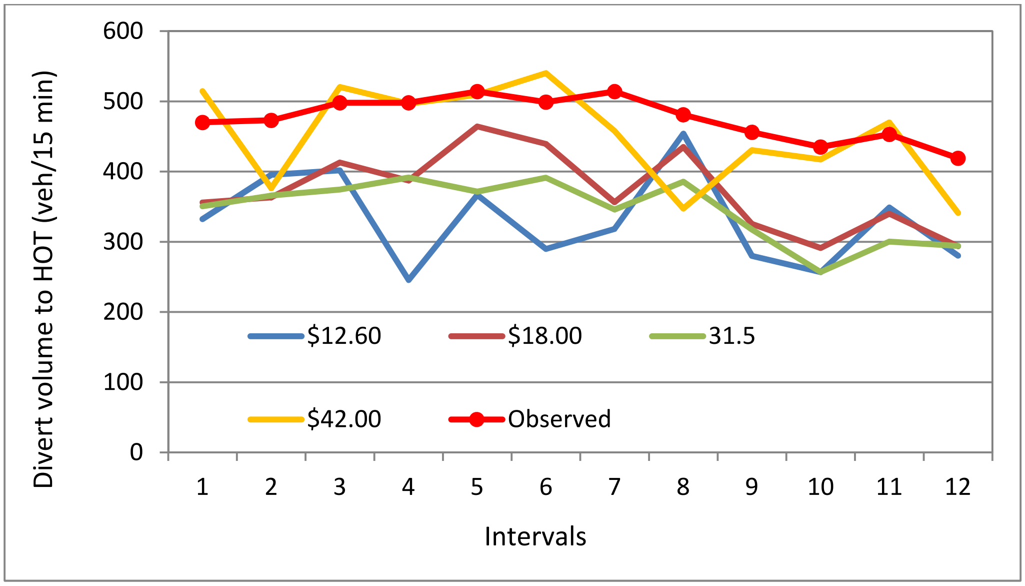


Figure 7-5 Comparison of diverted volume to ML for different value of times

7.3 Calibrating the Willingness-to-Pay Parameters

The parameters for the willingness-to-pay curve also need to be calibrated for use in the second ML modeling approach, as described in Chapter 4. The willingness-to-pay curve defines the proportion of people that are willing to divert to managed lanes, based on the ratio of the toll value (in cents) divided by the saved time (in minutes).

With regard to the willingness-to-pay curve, it should be clarified that this curve actually defines the percentages of people *not* willing to pay for different values of cents per minute. Figure 7-6 depicts different willingness-to-pay curves, and Figure 7-7 shows the resulting diverted volumes to the ML. As can be seen from the Figures 7-6 and 7-7, the shape of the willingness-to-pay curve should be convex, not concave. Curve I was selected as the willingness-to-pay curve that best reflect real-world conditions.

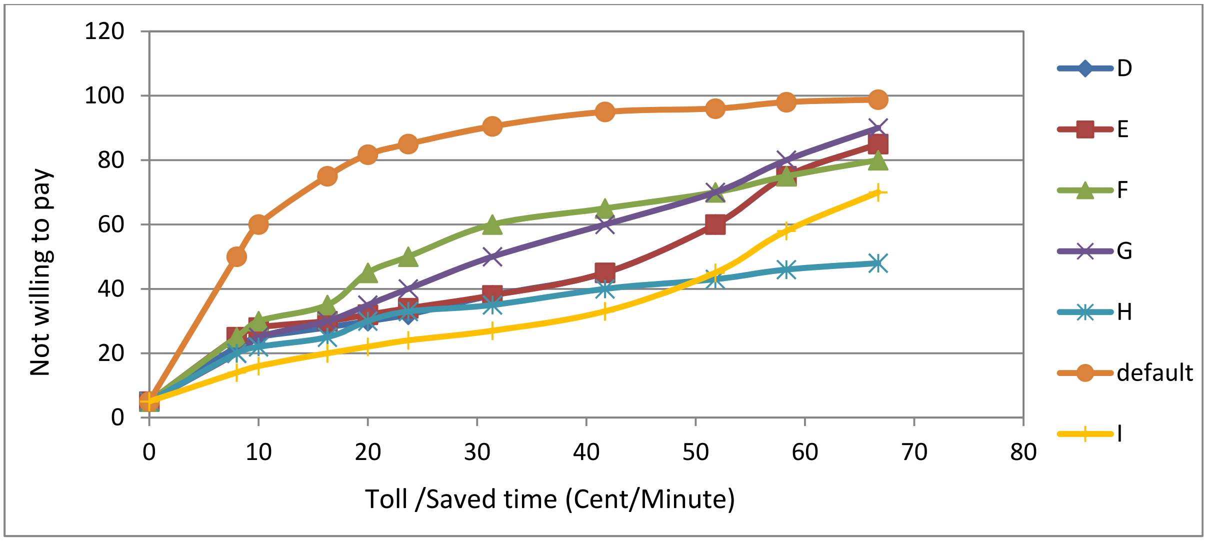


Figure 7-6 Different shapes of willingness-to-pay curve

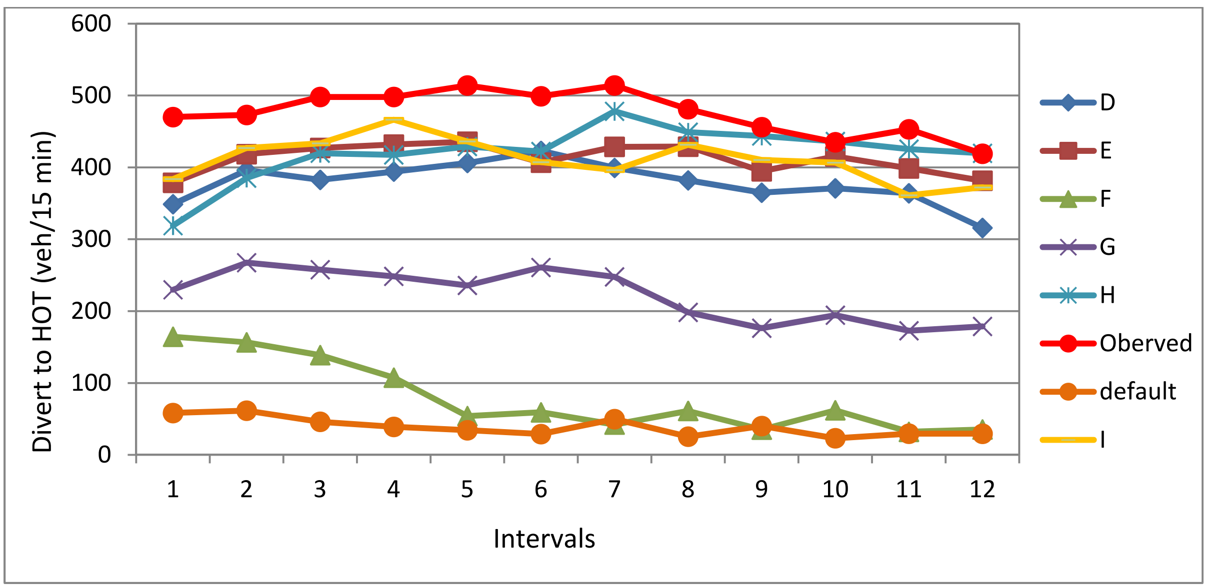


Figure 7-7 Diverted volume to ML associated with willingness-to-pay curves

Given that Curve I was selected as the best willingness-to-pay curve and a developed toll curve was calibrated, the next step was to assess if introducing a perceived benefit factor as a multiplier of the saved travel time can be used to improve the results by magnifying the saved travel time, originally obtained from skimming toll and toll-free routes. This is to account for other factors not accounted for such as reliability, comfort, and safety. The results are displayed in Figure 7-8.

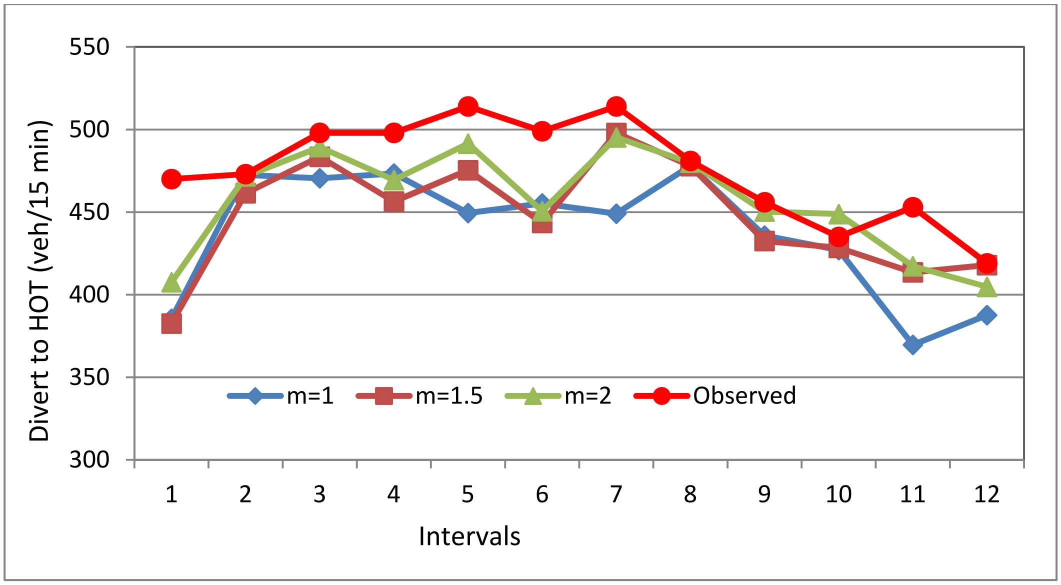


Figure 7-8 The effect of perceived benefit bias factor on predicting diverted volume to ML

In Figure 7-8, m=2 means that the saved travel time (due to paying the toll and using the ML lane instead of GPL) is doubled. As can be seen, the positive effects of increasing "m" dampen after a certain level (m=2 is not dramatically different from m=1.5).

An alternative method for dividing people into pay/not pay groups is the use of the logit probability model, as described in Chapter 4. Different parameters were suggested for calibrating the route choice logit model. There were two studies on the segment that were the main interest of this study: URS (2011) and Alvarez (2012). It is recommended that a comparison of the results of utilizing these logit models be made with the two approaches mentioned in this chapter. This will be done in the final version of this report.

7.4 Comparing Static and Dynamic Traffic Assignment

This section demonstrates the difference between how STA and DTA replicate the observed route choice behavior. Figures 7-9 and 7-10 demonstrate the difference between STA and DTA in predicting the divergence to the ML for the generalized cost function method and the willingness-to-pay curve method, respectively. The predicted divergence to the ML is also compared to the observed values derived from ITS data. As previously mentioned, the module in the Cube package for static assignment is called "Highway," and the module for dynamic assignment is called "Avenue." Please note that the Highway module was run for 12 intervals, one for each 15-minute interval, as described previously. Figures 7-9 and 7-10 show that both approaches of ML modeling produce results that are close to real-world results. However, the Highway module was not able to replicate real-world data. The two figures also show that the generalized cost approach and the willingness-to-pay approach produce comparable results in this case, although the generalized cost approach is much simpler to implement, calibrate, and converge.

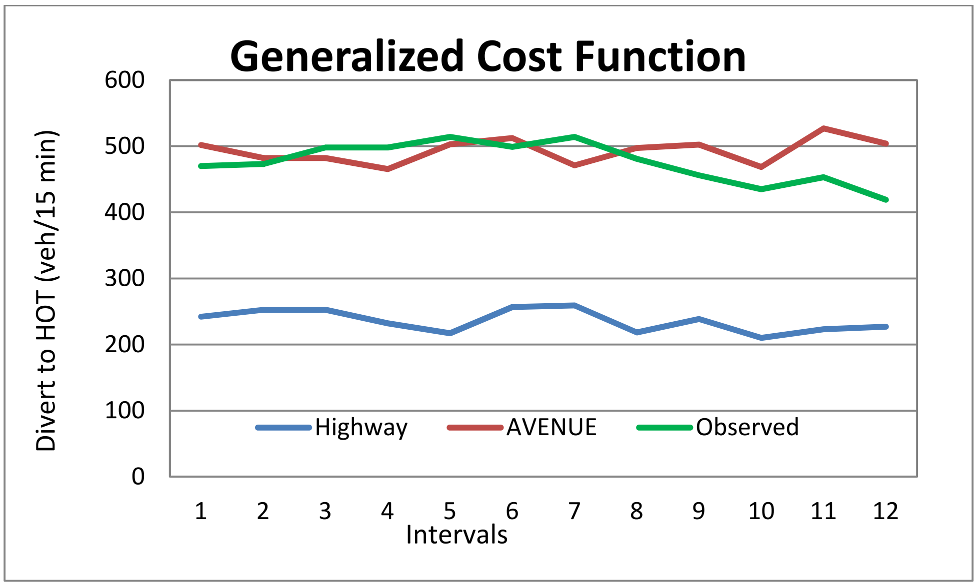


Figure 7-9 Comparison between modeled and observed ML volume for generalized cost function method

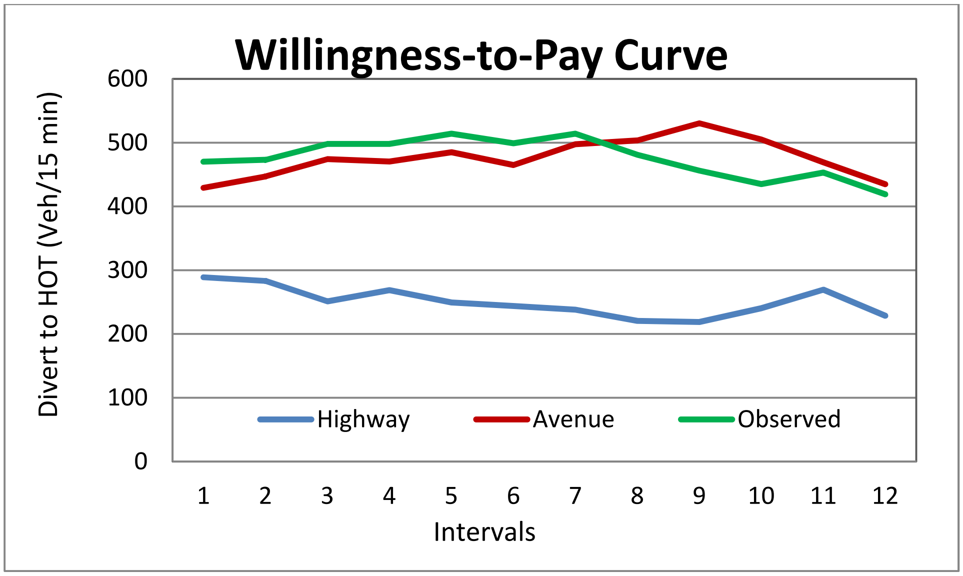


Figure 7-10 Comparison between modeled and observed ML volume for willingness-to-pay curve method

Figures 7-9 and 7-10 clearly show the superiority of DTA in predicting the route choice behavior. The main reason for the better prediction in DTA is its ability to model the dynamics of traffic breakdown (as discussed in Chapter 5), queues, spillbacks, and the associated delays. The difference in the travel time of using the GPL or the alternative ML, and the resulting number of travelers that decide to choose the ML, is considerably underestimated by static assignment.

7.5 Convergence and Stability

Figure 7-11 compares the relative gap for willingness-to-pay approach and the generalized cost function approach. The willingness-to-pay approach shows bad convergence. This issue is being discussed with Citilabs. Figure 7-12 shows the diverted volume to ML in each iteration for both assignment approaches. These figures clearly demonstrate the instability of the willingness-to-pay approach as it applied to DTA in this study.

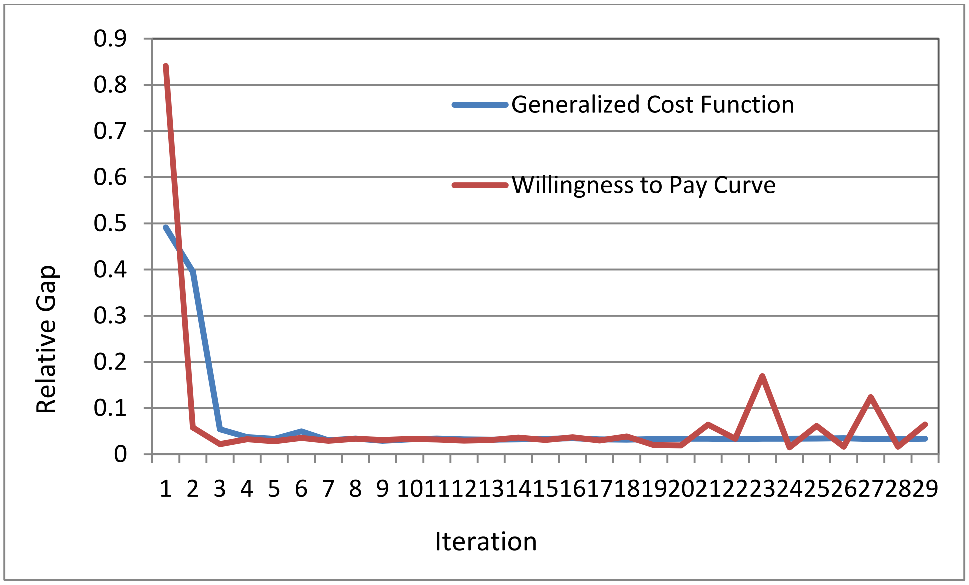


Figure 7-11 Relative Gap for different assignment approaches

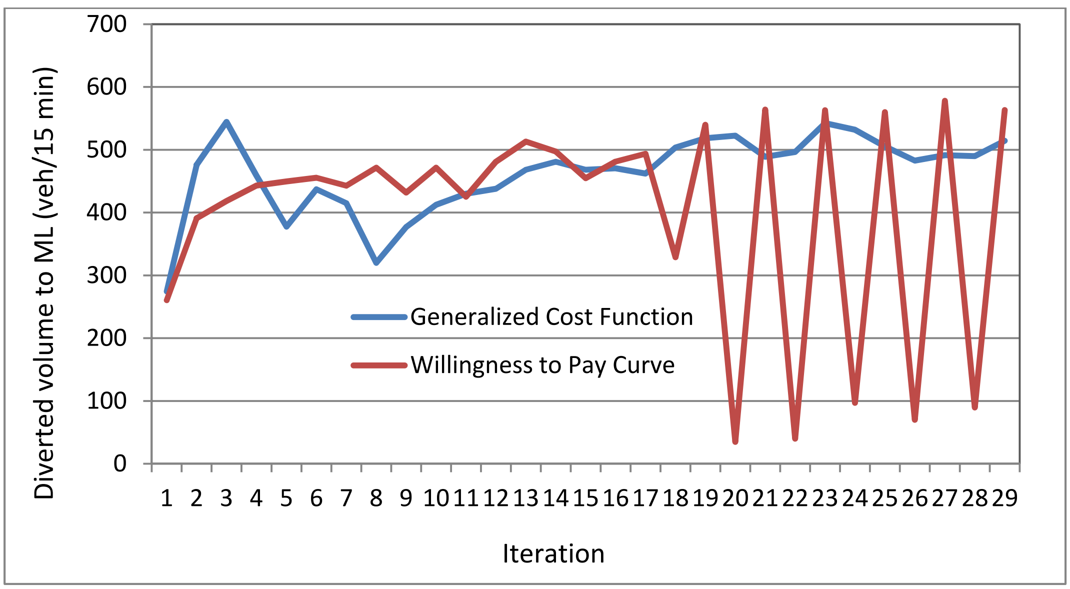


Figure 7-12 Diverted volume to ML for different assignment approaches

It should be noted that in general, but particularly in the case of managed lane modeling, trip-based or route-based measures of convergence are more important to be checked compared to link-based convergence measures. The current versions of Highway and Avenue assignment modules report a link-based convergence measure and do not report trip-based measures. Additional criteria such as the variation in the volume utilizing the ML between sequential iterations should be used as additional criteria for checking convergence, as is done in this study.

The calibration procedure adapted in this study is an iterative process between demands, network and assignment parameters. The final iteration is conducted when the assignment and route choice behavior is calibrated. When the diversion volumes to ML are calibrated, the OD estimation procedure needs to be run one more time. With the final estimated OD patterns, the network parameters should be fine-tuned again to replicate real-world congestion patterns. This will complete the iterative supply-demand-assignment processes.

7.6 Model Validation and Sensitivity Analysis

The demand and route choice parameters in this research were calibrated based on the volume averaged over representative days. The median day or any other day could have been used. In most locations, the volumes vary with a coefficient of variance (variance/ mean) of 3% to 7% between days. For validation and sensitivity analyses, Cube Avenue was run with different demand values from low to high, to see if it can replicate days with lower/higher congestion.

Figure 7-13 shows the speed contours for general purpose lanes, resulting from the generalized cost function assignment approach, for a demand multiplied by a scale factor. It can be seen that this approach can reasonably respond to the change in the demand level in terms of increased congestion patterns, meaning that the higher demands produce more congested networks.

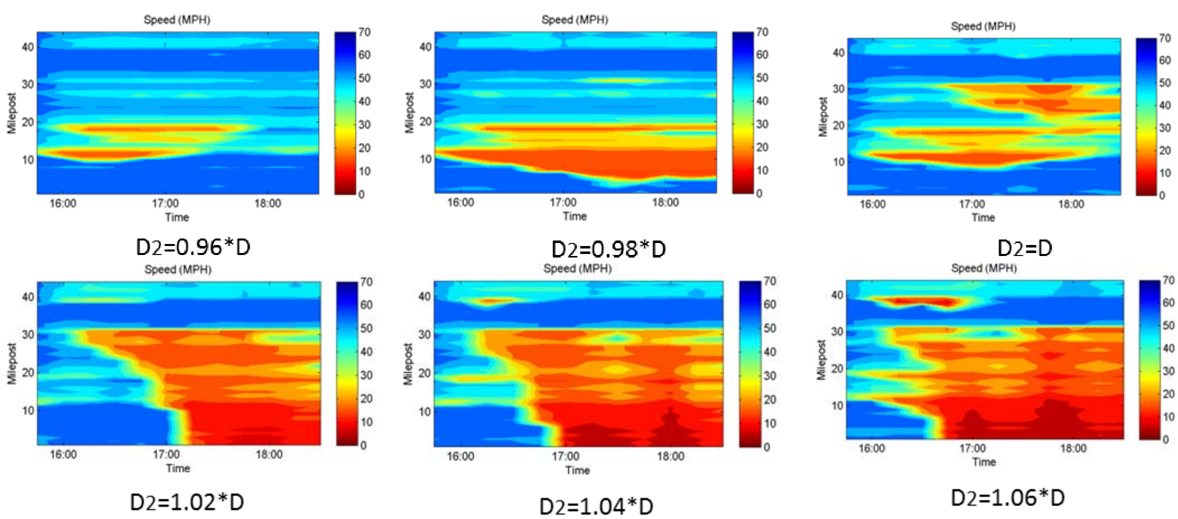


Figure 7-13 Speed contour for GPL with different demand levels when using the generalized cost function assignment

Figure 7-14 demonstrates the trend when using the willingness-to-pay approach. Unexpectedly, demands that are 96 percent of the original demands produced a high congestion. The reason is that the network is not stable, as is shown in Figures 7-11 and 7-12. Figure 7-15 shows the trend of VMT and VHT with changing demands using the generalized cost approach.

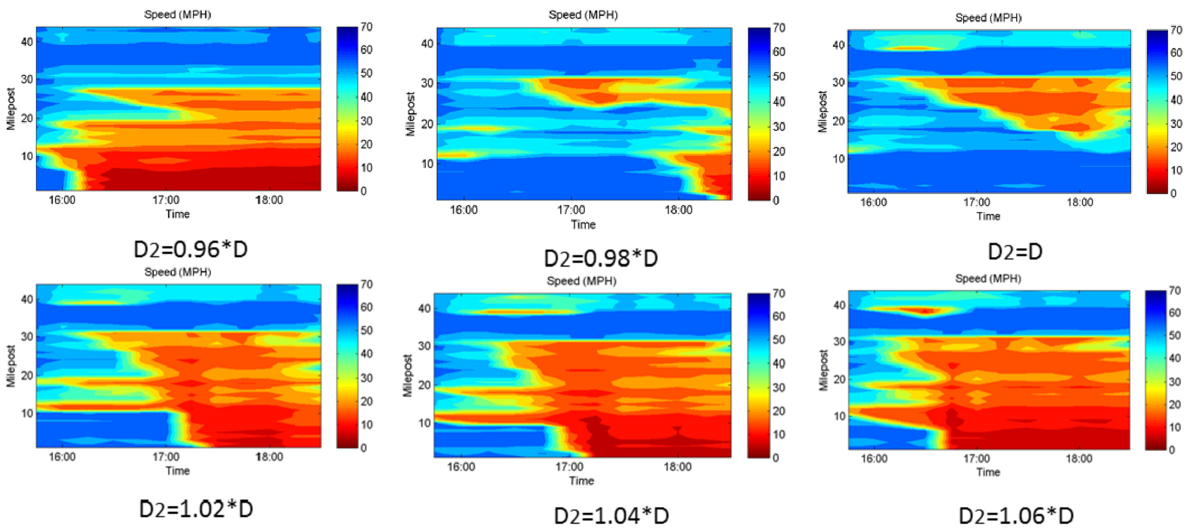


Figure 7-14 Speed contours for GPL with different demand levels when using the willingness-to-pay assignment

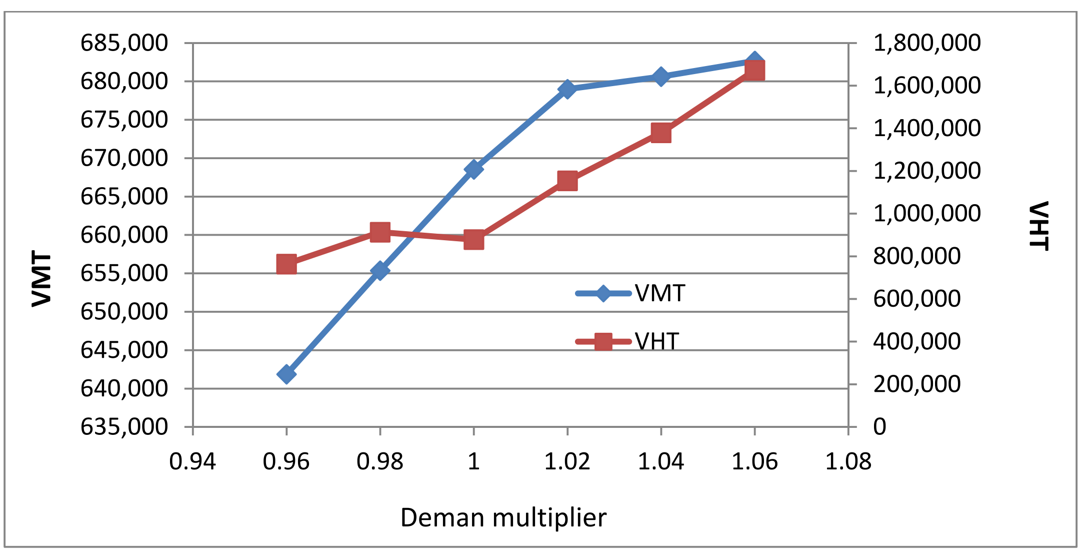


Figure 7-15 Changing in VMT and VHT on GPL whith changing demand

7.7 Assessment of Performance Measures

During the calibration, extensive use of data visualization and statistics analysis of volumes, speeds, and queues were conducted. A number of state and FDOT standards were consulted in this process including:

- The FHWA Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Micro-simulation Modeling Software to determine the quality of the resulting solution (Dowling, et al., 2004).
- FSUTMS-Cube Framework Phase II Model Calibration and Validation Standards (Cambridge Systematics, Inc., 2008).

Ohio RMSE Curve, which offers a target percent root mean squared error by volume group (Cambridge Systematics, Inc., 2010).

The final results for a converged and calibrated generalized cost function assignment are presented below. Figure 7-16 shows the scatter plot of observed versus simulated volumes at the screenline locations. This figure indicates that the coefficient of determination (R2) between the simulated and observed data is high (0.9761) indicating very high correlation with about 5% overestimation of the volumes on average as indicated by the 1.0515 coefficient value. It is interesting to compare this figure, with Figure 7-17, which is scatter plot for the same screenlines (without ramps), for two different representative days. This figure shows the day-to-day variation in real-world volume. The coefficient of determination between the volume measurements for these two days (0.7715) is lower than that between observed and simulated values.

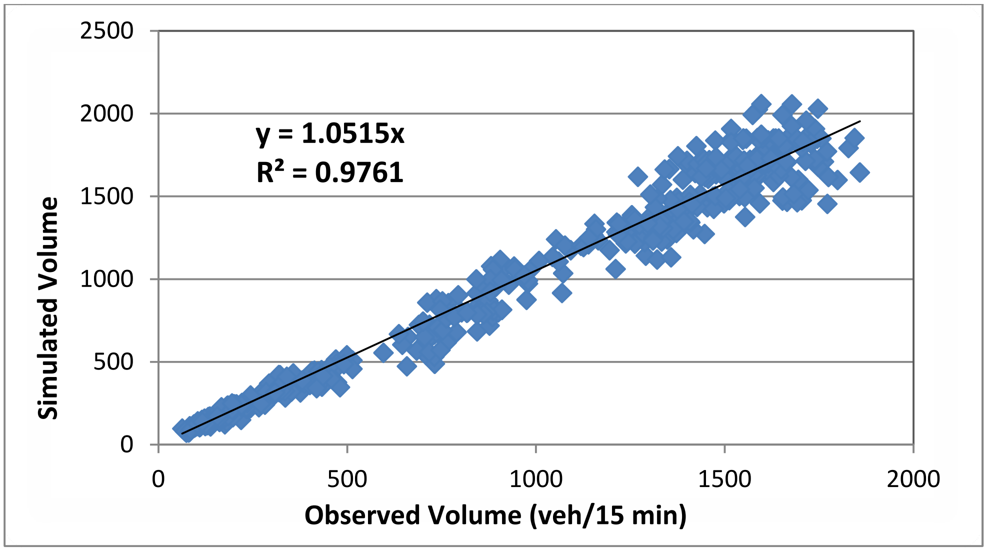


Figure 7-16 Scatter plot of observed versus simulated volume of screenlines

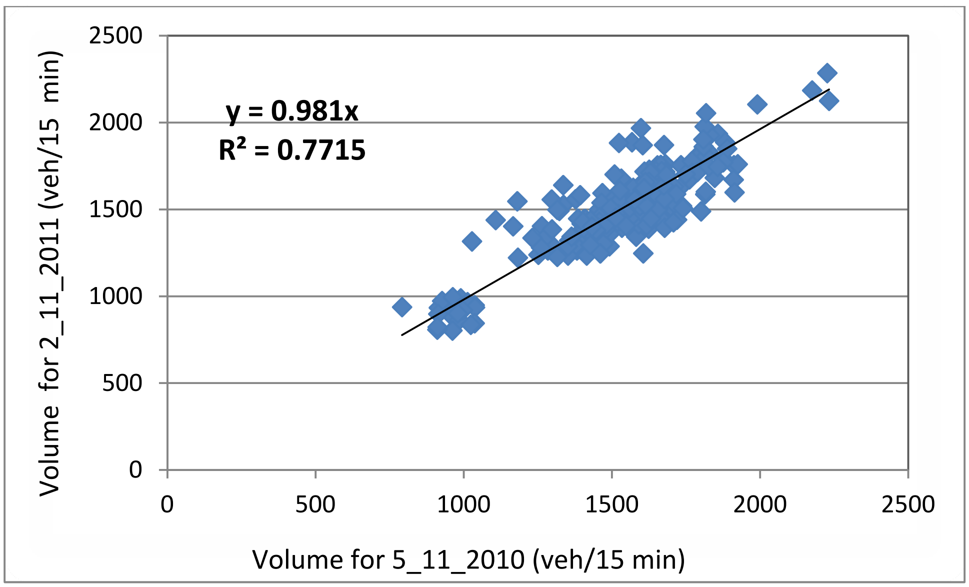


Figure 7-17 Scatter plot of observed volume of screenlines for different days

Figure 7-18 compares the observed and simulated flow rate on ML, which is an important indicator of route choice behavior calibration.

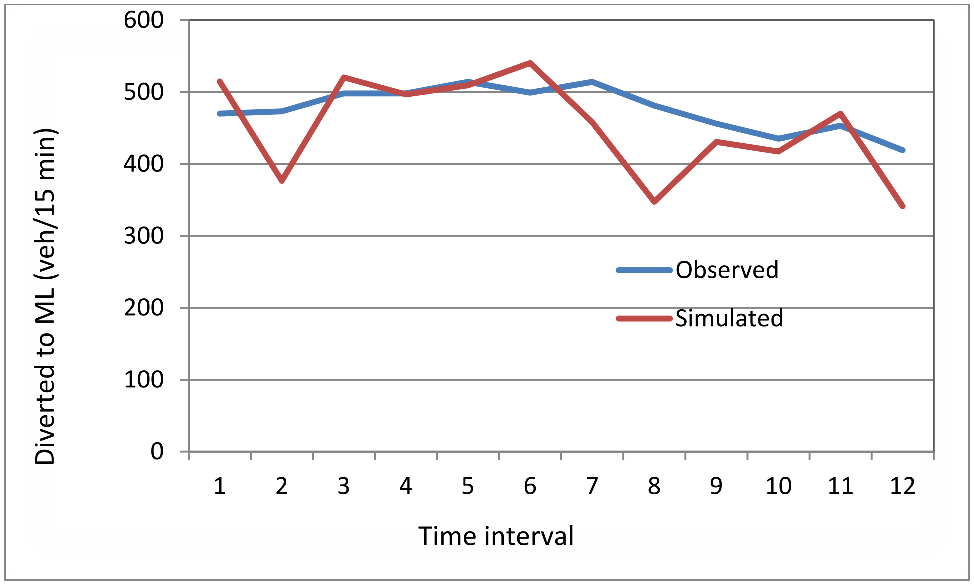


Figure 7-18 Comparison between observed and simulated flows on ML

Table 7-4 presents calculated goodness of fit measures of volume replication. These measures have been defined previously in section 5-5. The results in this table indicate a good goodness of fit between the measured and simulated volumes.

Table 7-4 Goodness of Fit Sstatistics for Volume Replication

Goodness of Fit Statistics	Value
RMSE	113
% RMSE	13.6
MAE	77
R squared	0.976
GEH <5	87
GEH <10	100

Examining RMSE curves for different ranges of volume showed that the resulting RMSE curves are well below the Ohio RMSE curve.

Figure 7-19 shows speed contour for simulated and one observed representative day.

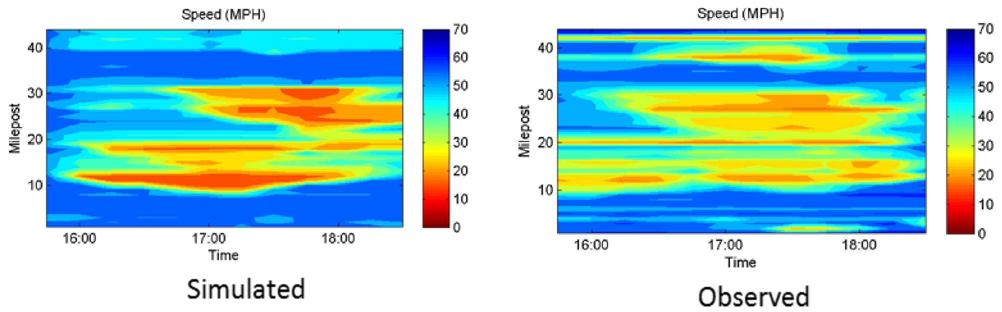


Figure 7-19 Comparison between simulated and observed speed contour

Figure 7-20 is scatter plot of the simulated speed, versus one observed representative day. This figure indicates not as good correlation between the measured and simulated speeds, although Figure 7-19 that the model was able to model the queues relatively well. Again it is interesting to compare the results in Figure 7-20 with those in Figure 7-21, which shows the relationship between the speeds for two different representative days. As can be seen, due to the probabilistic nature of traffic breakdown, there is a great variation in day-to-day congestion pattern and the correlation of speeds between these days is also low.

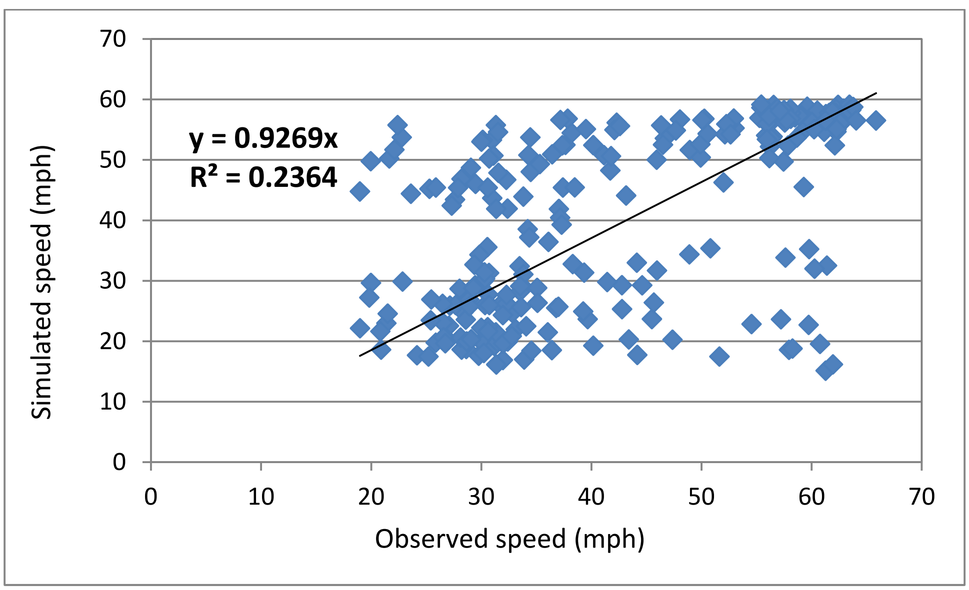


Figure 7-20 Scatter plot of observed versus simulated speed of screenlines

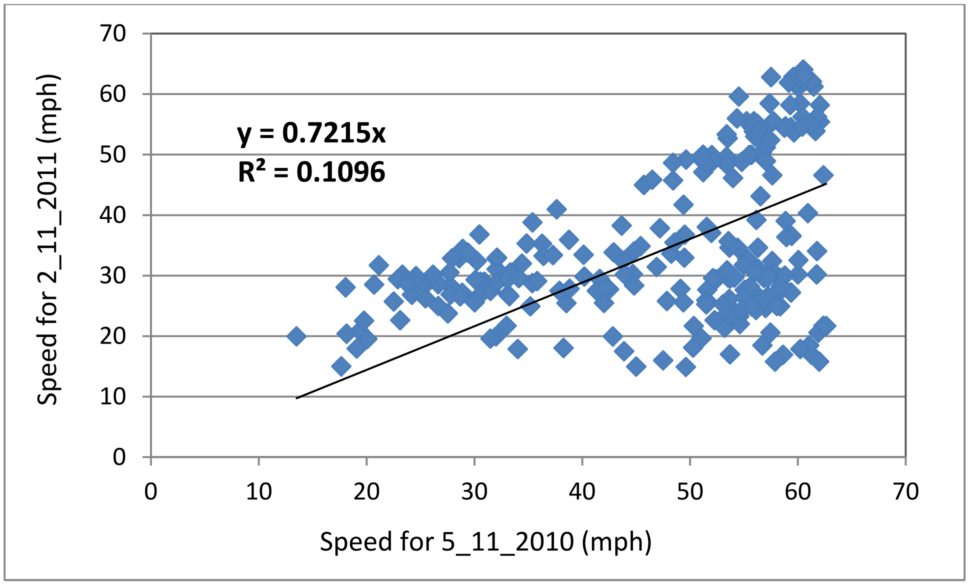


Figure 7-21 Scatter plot of observed speed of screenlines for different days

8. Subarea Managed Lane Modeling

8.1 Introduction

As stated earlier, the dynamic traffic assignment (DTA) modeling processes identified in this study were tested on the linear network of the I-95 corridor in Miami-Dade, Florida. URS, a member of the research team with extensive experience in toll modeling, was assigned the tasks of reviewing the model and investigating expansion into a wider subarea that includes parallel arterials east and west of I-95. The effort included:

- Independent model applications review, including evaluating model inputs, scripts, and results.
- Expanding the linear I-95 Managed Lane (ML) and General Purpose Lane (GPL) facility model to parallel arterials.

The specific objectives of the effort are as follows:

- Test the applications and processes developed in this project, providing general quality assurance/quality control (QA/QC) and ensure replication of results.
- Evaluate matrix estimation techniques and DTA modeling on the expanded highway network.

A secondary objective of this effort is to support an initial step of transferring the applications and lessons learned in this project to the modeling community, considering URS' role in the modeling of Florida's Turnpike toll facilities.

8.2 Linear Model Applications

As described earlier in this document, the developed model is based on Citilabs' applications, which include "Cube Voyager" (version 6.1 SP1) standard static equilibrium assignment, matrix estimation (ME) tools ("Analyst" for static and "Analyst Drive" for dynamic matrix estimation), and "Avenue" for the dynamic traffic assignment portion of the model. The highway network was extracted from the latest available Southeast Regional Planning Model (SERPM), as well as the daily trip matrices for all travel modes: Drive Alone (DA), Shared Ride of Two Persons (SRP2), Shared Ride of Three or More Persons (SRP3) and Truck. The trip tables were then factored to 15-minute time segments, based on factors derived using corresponding traffic counts to include as screenlines along the I-95 facility.

Once the highway network was populated with the screenline counts, the static matrix estimation process was used to refine the trip matrices for each of the 15-minute time segments for the dynamic traffic assignment, which included the peak 3 hours of the PM peak (12 time segments in total) for the four travel modes mentioned above.

The linear model only includes I-95 (managed lanes and general purpose lanes plus ramps) and has 303 highway links and 57 traffic analysis zones (TAZs). The model was created by the Florida International University (FIU) research team members and reviewed by the URS team members. Some key statistics are presented in Table 8.1. The URS team ran the model using the same inputs and individual program scripts used by the FIU team, but due to Citilabs' licensing agreements with URS, the "Analyst" portion had to be run in one computer, and the "Avenue" portion in another. The results of the URS model run are also summarized in Table 8.1.

Table 8-1 Linear Model Matrix Estimation – Cube Avenue Results

	All Vehicles Except Trucks			Trucks Only		
	GEH<5	GEH<10	R^2	GEH<5	GEH<10	R^2
FIU	71.84%	91.57%	0.985	94.83%	98.73%	0.954
URS	86.34%	93.75%	0.994	93.56%	98.81%	0.945
	All Vehicles					
	Vehicles Departed	Vehicles Arrived	Gap			
FIU	126,219	125,461	0.0102			
URS	123,426	123,415	0.0039			
Source: From model output files DOESS00A.DAT, DOESS00B.DAT , DOAVN00A.PRN						

Both model runs were quite similar, however, they are not exactly the same, due to slight changes/edits made to the screenline counts, which in turn affected the number of iterations to achieve convergence. In this model, there was even a loop in place to refine Cube Avenue assignments using an extra dynamic matrix estimation application from Citilabs ("Analyst Drive"), which was then input again into Cube Avenue to improve convergence.

8.3 Subarea Model Application

The next step was the development of a subarea network model, which included major arterials that run parallel to I-95. The model was evaluated following the same process used for the linear model. Figure 8-1 shows the 284 TAZs and 3,025 highway links of this subarea network extracted from the same SERPM as the linear model. The linear model is also shown in Figure 8-1 to appreciate the difference.

The static ME Analyst application was able to handle this subarea network size, but the dynamic ME application Analyst Drive failed, and after running the software for about 30 hours (including the Cube Avenue model run), it generated an error message and listed output files that were corrupted. This failure was replicated by Citilabs' developers, and since then, Citilabs included new options in a beta version of Analyst Drive to allow for a model this size to run properly. This beta version was not tested due to project time constraints. The DTA model as executed by Cube Avenue had convergence issues, which was also reported to Citilabs. Table 8-2 shows the results of this Cube Avenue model run.

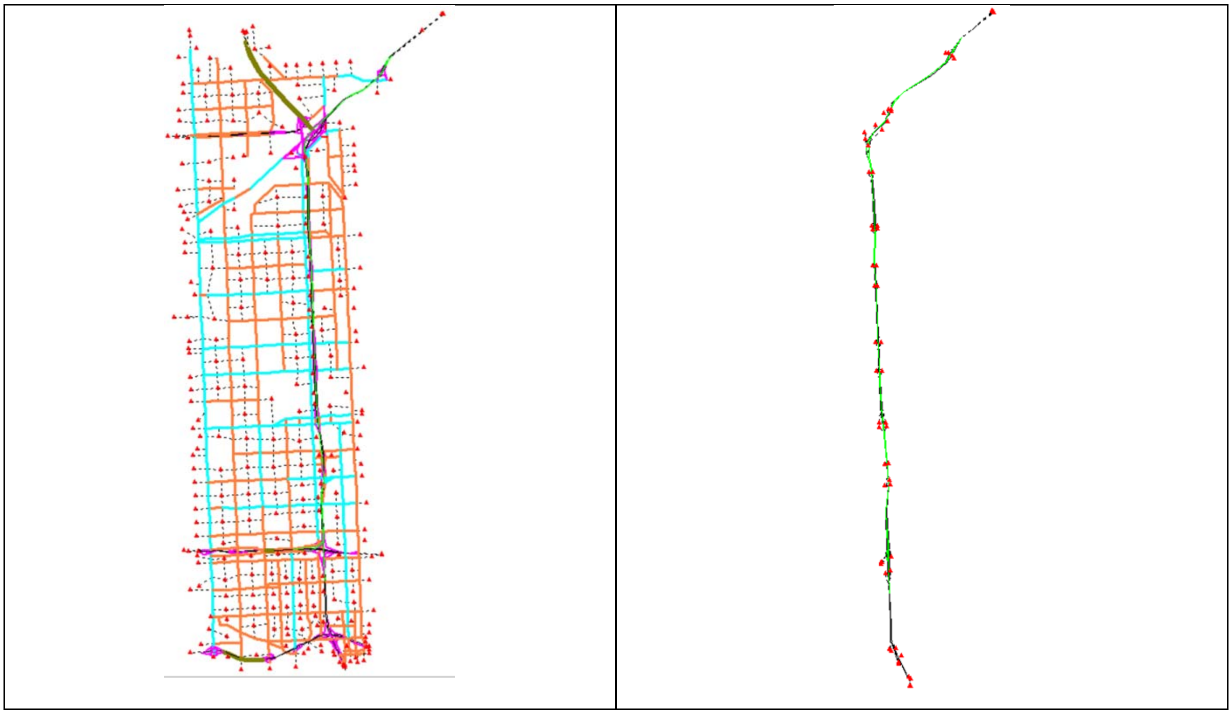


Figure 8-1 Subarea and linear corridor highway networks

Table 8-2 Subarea Model – Cube Avenue Results

Vehicles Departed	Vehicles Arrived	Gap	Number of Iterations
330,487	310,174	0.01042	50

Source: From model output file DOAVN00A.PRN

Note: The GAP cutoff value was 0.005, and the maximum number of iterations was set to 50. The average model running time was around 15+ hours for 50 iterations.

At this point in the process, it became obvious that a simplified network/trip table were needed so that the Cube applications could manage it. The original subarea was reduced by manually combining TAZs and modifying the highway network accordingly to match the new TAZ boundaries. Trip tables were also combined accordingly, keeping the original total trips unchanged. The result was a new network of 149 TAZs and 2,387 highway links. Figure 8-2 presents both subarea networks (before and after modifications) side by side, and a small insert of each network to graphically understand the differences in the levels of detail in the highway network and traffic analysis zones. It should be pointed out that combining the zones was only done to allow the model to run, and in general, is not recommended, since as stated in Chapter 3, it is recommended that the regional model zones are disaggregated rather than aggregated when transferred to DTA applications. This analysis indicates that Cube Avenue may not be appropriate for larger networks.

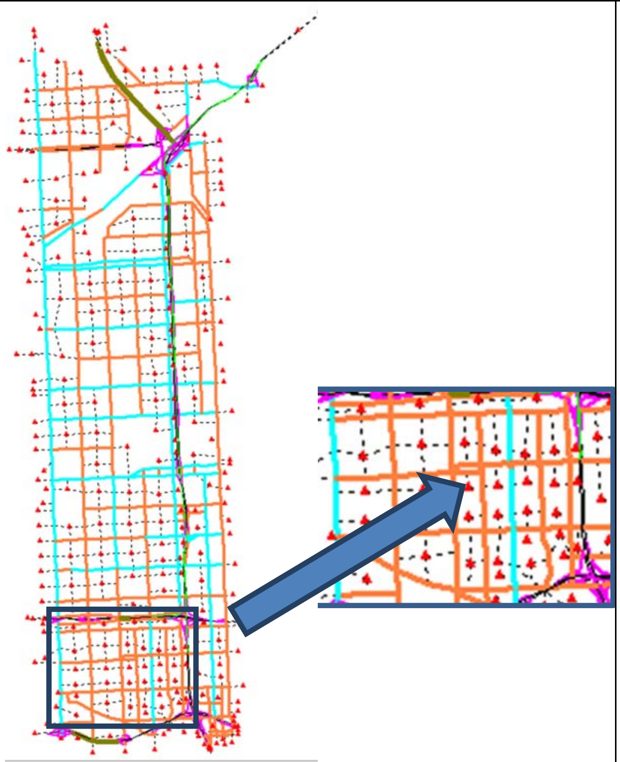
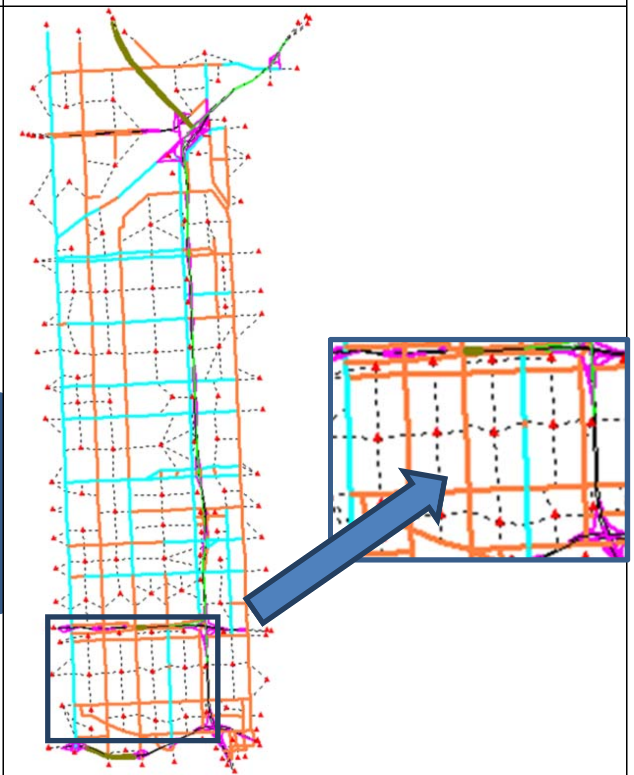
Original Subarea	Modified Subarea
	

Figure 8-2 Original and modified subarea highway networks

The "modified" subarea model still had issues with the Analyst Drive dynamic origin-destination (OD) estimation, even with the aggregation of the zones. Thus, it was decided to drop the dynamic ME refinement loop altogether and only use the static trip matrices as Cube Avenue inputs. Convergence was never achieved using the default gap cutoff (gap=0.005), although several attempts were made to refine the traffic counts, adding weights to differentiate count stations' level of trust, modifying the transition zone between HOV and managed lanes at the north end of the project, or including toll costs in the path development (not just the willingness-to-pay curves to split the trip tables). Table 8.3 displays the last model run statistics. Figures 8-3 and 8-4 are typical representations of Root Mean Square Error (RMSE) and GAP values variation throughout the 50 iterations.

Table 8-3 Modified Subarea Model – Cube Avenue Results

Vehicles Departed	Vehicles Arrived	Gap	Number of Iterations
333,931	318,708	0.02803	50

Source: From model output file DOAVN00A.PRN

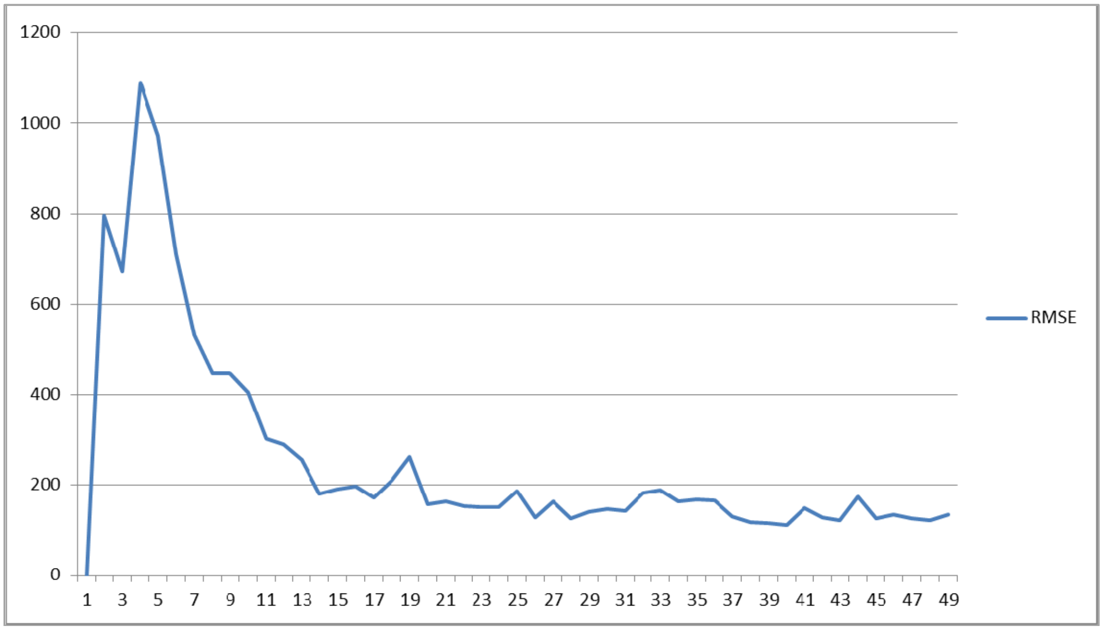


Figure 8-3 RMSE variation by model iteration

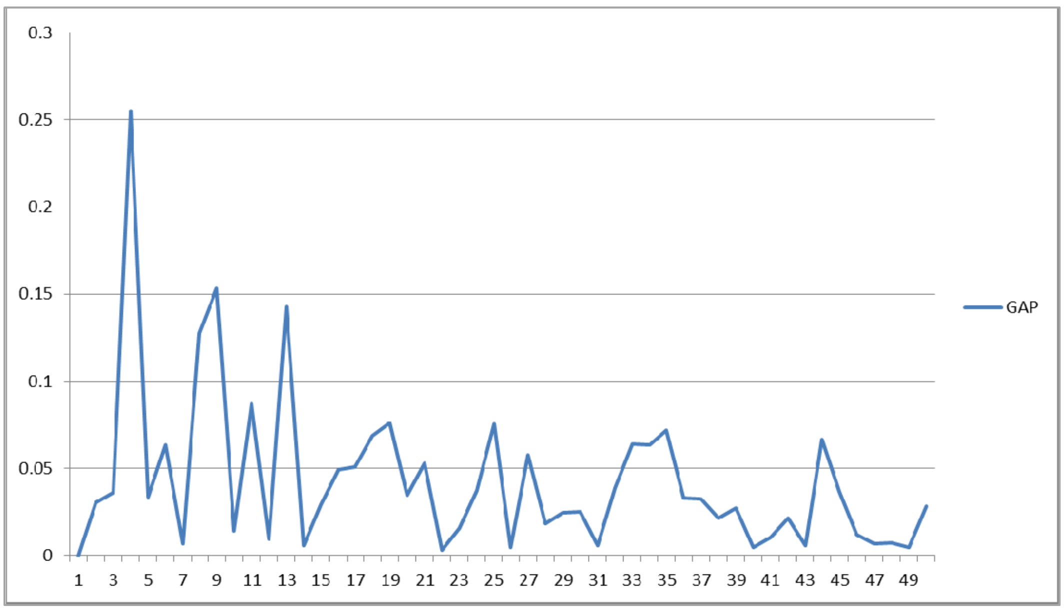


Figure 8-4 Gap variation by model iteration

The results in Figure 8-5 and Figure 8-6 were generated after considering all trips assigned by the Cube Avenue DTA and the static equilibrium assignment for all time intervals. The lack of convergence previously noted in the Cube Avenue assignment may be one of the reasons why the static assignment shows a better correlation between observed counts and assigned volumes (R^2=0.94) than the dynamic traffic assignment (R^2=0.79).

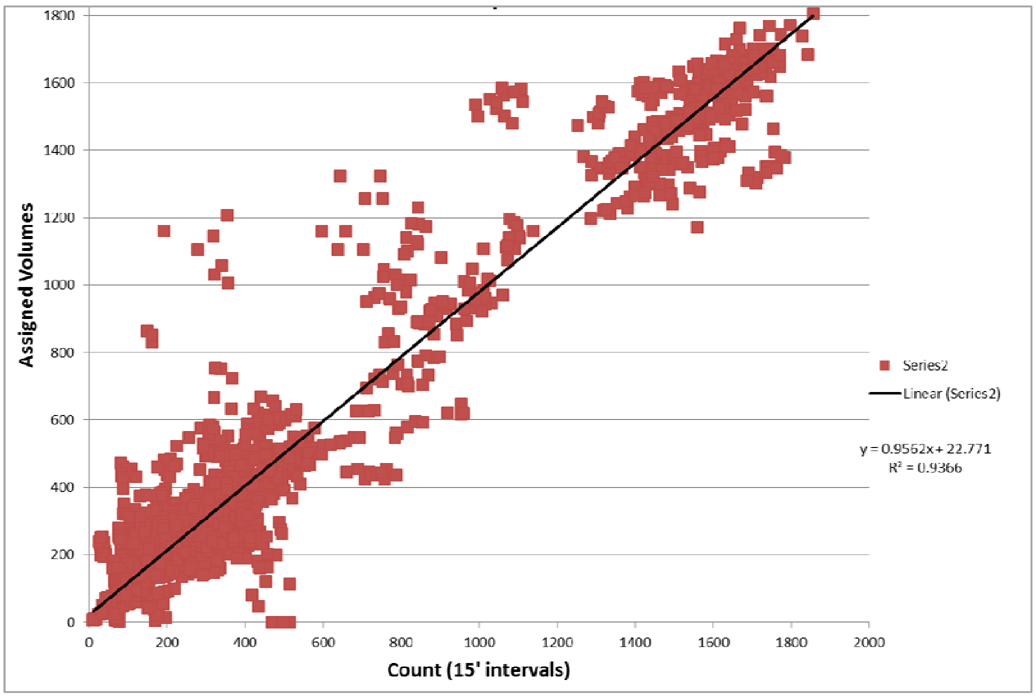


Figure 8-5 Counts versus static assignment estimated volumes for all time intervals

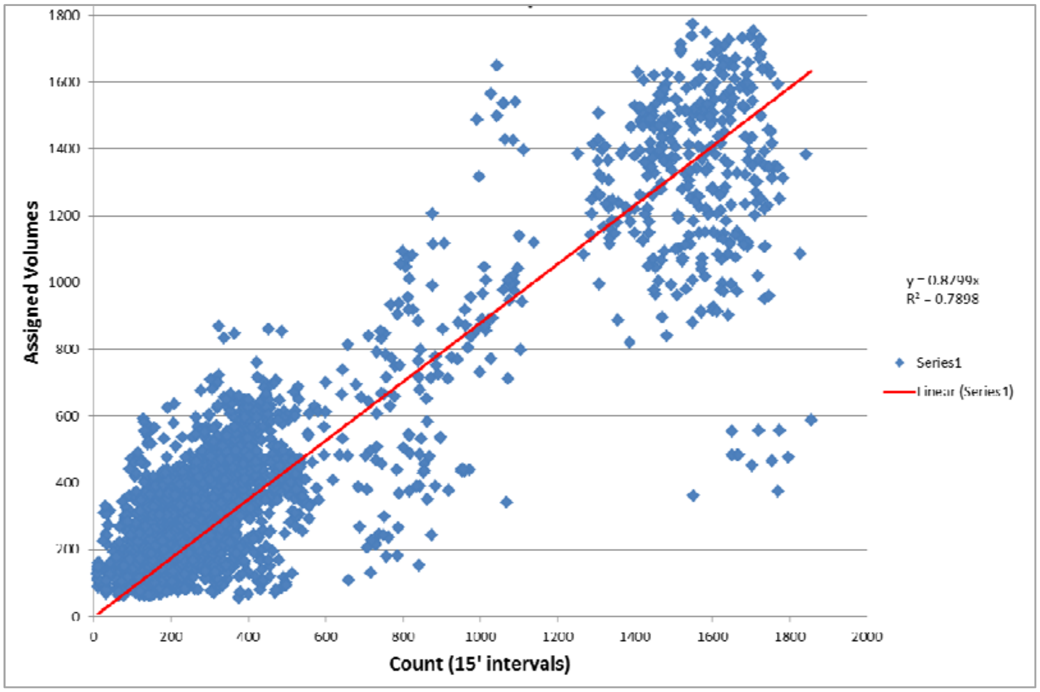


Figure 8-6 Counts versus DTA estimated volumes for all time intervals

However, due to the convergence uncertainty, it is not possible to draw any conclusions in terms of the advantages of the static versus dynamic assignment from a planning perspective, or even the static matrix estimation over the dynamic matrix estimation due to previous software limitations.

8.4 Findings

Based on the results of the analysis, the URS team member offered the following recommendations to promote further research and improve model users' understanding of these models:

- Network size. A linear facility with small-size trip matrices can work well, and are easier to replicate than larger networks; a subarea extraction from Regional Models required additional manipulation in order for the software to function.
- Software capability. A software version can make a difference, as well as licensing agreements: model complexity can create an impassable situation depending on software capability. The original release of Analyst Drive did not work for this project, and

convergence issues on the subarea are still unresolved. Cube Voyager, Cube Analyst, and Cube Avenue require separate licenses if they will be used either alone or together in a model application. This project required all of these applications to run together, which is quite new in Florida modeling practices; hence, the emergence of licensing conflicts and the need for a bypass to the problem.

- Traffic count issues. Factoring trip tables to obtain short-length travel time intervals and trusted sources for counts have compounding effects on model calibration/validation of results.
- Static versus dynamic assignment. It was not possible to reach a final conclusion on the advantages/disadvantages of DTA versus static assignment on larger than normal linear corridors from this analysis. Similar issues in terms of matrix estimation techniques (static versus dynamic) were encountered that require more guidance from the software developer's side.

9. Conclusions

Simulation-based dynamic traffic assignment (DTA) has been increasingly utilized to evaluate traffic management strategies, including managed lane (ML). Compared to traditional methods that normally utilize static traffic assignment (STA) and simple analytical traffic flow equations, simulation-based DTA better captures the dynamics of traffic operations by modeling time variant system measures (including queuing and travel times), demands, advanced management strategies, and the associated responses of travelers. The following can be stated regarding DTA applications in general, and their use in ML modeling in particular, based on the findings from this study.

- A variety of modeling approaches have been proposed to assess managed lane implementations. These approaches range from high-level sketch planning tools to micro-level modeling of individuals' behaviors and traffic operations. Simulation-based DTA should be considered an effective modeling approach to support the planning and operation of ML.
- The modeling network can be extracted as a subarea network from the regional models. For this purpose, the subarea boundary can be specified using the Cube Polygon feature or a GIS tool. However, it was found that the modeled network geometry needs to be updated to better represent the existing real-world network geometry, since the details and accuracy of modeling the network in demand forecasting models are not sufficient for DTA applications.
- Advanced modeling tools such as DTA requires more detailed and higher quality data to ensure that the developed model accurately replicates real-world conditions. This study successfully and extensively utilized detector data collected from an existing ITS system operated by the regional traffic management center, combined with portable traffic monitoring sites (PTMS) ramp counts and measurements from other sources of data to satisfy the DTA data needs. However, significant efforts were needed to process, fuse, and validate the data to allow for use in the modeling processes.
- Managed lane modeling was successfully implemented and tested in Cube Avenue using two approaches. The first approach involves adding the equivalent value of time of the

toll cost value to the travel time function within the assignment, resulting in a generalized cost that considers the ML toll. In this approach, referred to as the "generalized cost function" approach in this study, vehicle use of ML is solely governed by the user equilibrium (UE) assignment procedure, based on the generalized costs of the competing paths. The second approach is referred to as the "willingness-to-pay curve" approach. In this method, prior to the assignment, travelers are divided into two groups: A group that will not choose to pay the toll and is limited to using the general purpose lane (GPL), and a group that is willing to pay and use the ML. The latter other group is eligible to use the ML lanes based on the willingness-to-pay curve, but the final decision to use either ML or GPL depends on its origin and destination points (if there is a managed lane in their paths) and on the difference in the generalized costs between ML and alternative routes according to the UE process.

- In this research, several issues were identified that limited Cube Avenue's modeling abilities, in particular, as it relates to ML modeling using the willingness-to-pay curve approach. Citilabs, which is a member of this research, addressed the identified issues and updated Cube Avenue during the course of the projects to reflect the project's findings. Although these problems were solved and reasonable results were produced with the final version of the model, it is recommended that the user examine the results and report any issue to the program developer.
- A sequential procedure that iterates between network (supply) calibration, demand estimation, and route choice parameter estimation is recommended in this study. Despite the existence of mathematical formulations and solutions for simultaneous supply and demand estimation, their implementations in the real world are not straightforward and have not been executed.
- Supply or network calibration in Cube Avenue entails estimating capacity, free-flow, and traffic flow model parameters for each link in the network. These parameters affect the travel time, congestion time, queue formation, and queue spillback when the demand is loaded. A systematic multilevel approach to network (supply) calibration is recommended in this study, with an increasing calibration scope in each level. The process starts at the level of separated bottlenecks where the capacity is estimated by

various methods based on field data. The network is gradually extended to connect the bottlenecks, and then to the whole corridor and subarea coverage.

- The supply calibration performed in this study illustrates the importance of coding capacity based on detector measurements in DTA tools, particularly when there is evidence that the modeled corridor capacity is lower than the HCM-based estimates. In the case explored in this study, it was found that the free-flow speed and more importantly, the capacity, were overestimated by the HCM procedures, resulting in incorrect travel times and congestion when used in the DTA model.
- One of the important congestion spots in the modeled network is caused by a spillback from an off-ramp that causes low speeds in the two left lanes (the I-95 Northbound off-ramp to the Turnpike). Since the utilized DTA tool (Cube Avenue) does not support lane-by-lane modeling, it is not possible to correctly replicate that location, because the queue in the model first fills up the whole segment (including 5 lanes) before backing up to the upstream link. In the real-world, only the two left lanes are blocked. If replicating the congestion at such locations is important to a study, a tool that better handle this situation or multi-resolution analysis should be considered.
- Dynamic traffic assignment requires trip matrices specified for short time intervals (e.g., 15 minutes or 30 minutes). The derivation of these matrices is performed in this study using a sequential process that starts from matrix factorization based on count data, followed by static assignment-based OD matrix estimation, and finally followed by dynamic assignment-based OD matrix estimation. However, identified limitations, tool immaturity, and the results of this study indicate that at the current stage, the dynamic OD estimation process in Cube Analyst should be used with caution until further enhancements and testing of these enhancements are completed so as to confirm that the tool is able to produce good results.
- During the matrix estimation process with the currently available tools, several manual adjustments and iterations are required to ensure joint calibration of demand, supply, and route choice behaviors. Adjustments and fine-tunings are also needed to avoid unrealistic deviation from the initial matrix and trip pattern.
- Important specific enhancements to the OD estimation process in Cube Avenue are recommended, as listed in Chapter 6.

- When calibrating supply, demand, and assignment parameters, a distance function between simulation outputs and field measurements is minimized. This function can include different measures, such as link volumes, OD demands, link speeds and/or densities, etc. Limiting the function to replicating link volumes, as is the case in many studies, can be misleading and fail to produce the correct demands or congestion patterns. Most OD matrix estimation methods are based on link traffic volumes and initial OD matrices. If enough data on speeds, densities, queue lengths, OD routes or zonal trip end rates are available, they should be incorporated into the calibration process to better replicate real-world traffic conditions.
- Calibrating the toll curve, value of time, and willingness-to-pay curve parameters are important aspects of DTA utilization for ML modeling.
- There is evidence that the value of time used in the SERPM model ($12.60 per hour) is low. The value identified in a previously conducted study of the I-95 Express ($22.00) produces better results.
- There is evidence that motorists perceived additional benefits of ML not accounted for by the raw value of travel time. Thus, a factor is used in this study, as a multiplier of the saved travel time, to improve the results by magnifying the saved travel time, originally obtained from skimming toll and toll-free routes. This is to account for other factors initially not accounted for, such as reliability, comfort, and safety.
- The findings from this study highlight the shortcomings of utilizing static assignment for assessing managed lanes, even when the measured capacity values and correct volumes are coded, illustrating the need to utilize DTA modeling for such assessments. The calibrated DTA model was able to produce results that are similar to real-world results. However, the Cube static assignment module was not able to replicate real-world data.
- For the case study of this project, it was found that the generalized cost approach and the willingness-to-pay approach produce comparable results, although the generalized cost approach is much simpler to implement, calibrate, and converge.
- For the case study of this project, the generalized cost approach appears to be able to achieve converged and stable solutions. However, the willingness-to-pay approach was not able to achieve converged and stable solutions.

- Initial attempts were made to expand the network to a larger subarea, compared to the linear model used in this study. It was found that a linear facility with small-size trip matrices can work well and are easier to replicate than larger networks. In addition, the original release of Analyst Drive (dynamic OD estimation) did not work for the larger area.
- Cube Voyager, Cube Analyst, and Cube Avenue require separate licenses if they will be used either alone or together in a model application. This project required all of these applications to run together, which is quite new in Florida's modeling practices; hence, the emergence of licensing conflicts and the need for a solution to the problem.

10. References

Akcelik, R. (2003). Speed-Flow and Bunching Relationship for Uninterrupted Flows Presented at 25th Conference of Australian Institute of Transport Research, University of South Australia.

Alvarez, P. (2012). A Methodology to Estimate Time Varying User Responses to Travel Time and Travel Time Reliability in a Road Pricing Environment. Ph.D. Dissertation, Florida International University

Antoniou, C., Ben-Akiva, M., and Koutsopoulos, H. (2007). Non–linear Kalman Filtering Algorithms for On–line Calibration of Dynamic Traffic Assignment Models. IEEE Transactions on Intelligent Transportation Systems, Vol. 8, No. 4, pp. 661 –670.

Ashok, K. and Ben-Akiva, M. (1993). Dynamic OD Matrix Estimation and Prediction for Real-Time Traffic Management Systems. In Daganzo, C., editor, Transportation and Traffic Theory, Elsevier Science Publishing, pp. 465–484.

Ashok, K., and Ben-Akiva,M. (2002). Estimation and Prediction of Time-Dependent Origin–Destination Flows with a Stochastic Mapping to Path Flows and Link Flows, Transportation Science., Vol. 36, No. 2, pp. 184–198.

Balakrishna, R. (2002). Calibration of the Demand Simulator in Dynamic Traffic Assignment System, Master Thesis, Massachusetts Institute of Technology.

Balakrishna, R., Ben-Akiva, M., and Koutsopoulos, H. (2007). Off-line Calibration of Dynamic Traffic Assignment: Simultaneous Demand-Supply Estimation. Transportation Research Record: Journal of the Transportation Research Board, No. 2003, pp. 50-58.

Bar-Gera, H., Boyce, D. (1999). Route Flow Entropy Maximization in Origin-Based Traffic Assignment. In Ceder, A. (Ed.), In Proceedings of the 14th International Symposium on

Transportation and Traffic Theory, Jerusalem, Israel, 1999. Elsevier Science, Oxford, UK, pp. 397–415.

Bar-Gera, H. (2010). Traffic Assignment by Paired Alternative Segments. Transportation Research Part B, Vol. 44, pp.1022-1046.

Bassan, S. and Polus, A. (2010). Meaning of Actual Capacity of Freeways and Its Estimation. Canadian Journal of Civil Engineering, Vol. 37, No. 1, pp. 77-87.

Bertini, R. L., Fernandez, R., Wieczorek, J., Li, H. (2008). Using Archived ITS Data to Automatically Identify Freeway Bottlenecks in Portland, Oregon. In: Proceedings of the 15th World Congress on ITS, New York.

Bertini, R.L. and Myton, A. (2005). Using PeMS Data to Empirically Diagnose Freeway Bottleneck Locations in Orange County, California. Transportation Research Record: Journal of the Transportation Research Board, No. 1925, pp. 48-57.

Boyce, D., Nie, Y., Bar-Gera, H., Liu, Y., and Hu, Y. (2010). Field Test of a Method for Finding Consistent Route Flows and Multiple-Class Link Flows in Road Traffic Assignments. Final Report to the Federal Highway Administration, Washington, DC.

Boyce, D., O'Neill, C. R., & Scherr, W. (2008). Solving the Sequential Travel Forecasting Procedure with Feedback. Transportation Research Record: Journal of the Transportation Research Board, No. 2077, pp. 129-135.

Boyce, D., Ralevic-Dekic, B. and Bar-Gera, H. (2002). Convergence of Traffic Assignments: How much is Enough? Journal of Transportation Engineering, Vol. 130, No. 1, pp. 49-55.

Brandiss, J. (2001). Estimation of Origin-Destination Flows for Dynamic Traffic Assignment. Master Thesis, Massachusetts Institute of Technology.

Brilon, W., Geistefeldt, J., and Regler, M. (2005). Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity, In: Proceedings of the 16th Symposium on Transportation and Traffic Theory.

Burris, M., Nelson, S., Kelly, P., Gupta, P. and Cho, Y. (2012). Willingness-to-pay for HOT Lanes - Empirical Analysis from I-15 and I-394. Transportation Research Record: Journal of the Transportation Research Board, No. 2297, pp. 47-55.

Cambridge Systematics, Inc. (2008). FSUTMS-Cube Framework Phase II Model Calibration and Validation Standards. Prepared for Florida Department of Transportation Systems Planning Office, http://www.fsutmsonline.net/images/uploads/reports/FR2_FDOT_Model_CalVal_Standards_Final_Report_10.2.08.pdf, Accessed Oct 12, 2013.

Cambridge Systematics, Inc. (2010). Travel Model Validation and Reasonableness Checking Manual (Second Edition). Prepared for Federal Highway Administration, http://www.camsys.com/pubs/FHWA-HEP-10-042.pdf, Accessed Oct. 12, 2013.

Cambridge Systematics, Inc. (2012). Value of Travel Time Reliability, Synthesis Report and Workshop Working Paper. Prepared for SHRP 2 Workshop on Value of Travel Time Reliability. http://onlinepubs.trb.org/onlinepubs/shrp2/L35RFP/DraftSynthesis-ValuingReliability.pdf, Accessed Oct 12, 2013.

Chao, C., Varaiya, P., and Kwon, J. (2005). An Empirical Assessment of Traffic Operations. In Proceedings of the International Symposium on Transportation and Traffic Theory.

Chen, C., Skabardonis, A., and Varaiya, P. (2004). Systematic Identification of Freeway Bottleneck, In: Proceedings of the 83rd Transportation Research Board Annual Meeting, Washington, D.C.

.

Chen, Z. (2003). Bayesian filtering: From Kalman filters to particle filters, and beyond. Journal of Statistics, Vol. 182, No.1, pp: 1-69.

Chi, H., (2010). An Improved Framework for Dynamic Origin-Destination Matrix Estimation. PhD. Dissertation, Florida International University.

Chiu, Y.C. (2012). Modeling Express Lanes Using Dynamic Traffic Assignment Models. FDOT Managed Lane Workshop. May 22, 2013, http://www.fsutmsonline.net/images/uploads/12_Dynamic_Traffic_Modeling_-_Yi-Chang_Chiu.pdf, Accessed Oct. 15, 2013.

Chiu, Y. and Bustillos, B. (2009). A Gap Function Vehicle Based Solution Procedure for Consistent and Robust Simulation Based Dynamic Traffic Assignment. 88th Annual Meeting of the Transportation Research Board.

Chiu,Y., Bottom,J., Mahut,M., Paz,A., Balakrishna,R., Waller,T. and Hicks, J. (2011). Dynamic traffic assignment: A primer. Transportation Research E-Circular, No. E-C153.

Chiu, Y. C., Zhou, L., and Song, H. B. (2010). Development and Calibration of the Anisotropic Mesoscopic Simulation model for Uninterrupted Flow Facilities. Transportation Research Part B No. 44 pp.152-174.

Citilabs, Inc. (2013). Citilabs Cube User Manual.

Davidson, K.B. (1966). A Flow–Travel Time Relationship for Use in Transportation Planning. In Proceedings of the 3rd ARRB Conference, Vol. 3, No. 1, pp. 183-194.

Davidson, K.B. (1978). The Theoretical Basis of a Flow–Travel Time Relationship for Use in Transportation Planning. Australian Road Research, Vol. 8, No. 1, pp. 32-35.

Dehghani, Y., Adler, T., Doherty, M. W., and Fox, R. (2003). Development of a new toll mode-choice modeling system for Florida's Turnpike Enterprise. Transportation Research Record: Journal of the Transportation Research Board, No. 1858, pp. 9-17.

Dehman, A. (2012). Breakdown Maturity Phenomenon at Wisconsin Freeway Bottlenecks. In: Proceedings of the 91st Annual Meeting of Transportation Research Board.

Dervisoglu, G., Gomes, G., Kwon, J., and Varaiya, P. (2009). Automatic Calibration of the Fundamental Diagram and Empirical Observations on Capacity, In Proceedings of the 88th Annual Meeting of Transportation Research Board.

Dixon, M. P. and Rilett, L. R. (2002). Real-time OD estimation using automatic vehicle identification and traffic count data. Journal of Computer-Aided Civil and Infrastructure Engineering, Vol. 17, No.1, pp. 7-21.

Doan D.L., Ziliaskopoulos, A. and Mahmassani, H.S. (1999). An On-line Monitoring System for Real-Time Traffic Management Applications. Transportation Research Record: Journal of the Transportation Research Board, No.1678, pp. 142-149.

Doblas, J. and Benitez, F. (2005). An Approach to Estimating and Updating Origin–Destination Matrices Based Upon Traffic Counts Preserving the Prior Structure of a Survey Matrix. Transportation Research Part B, Vol. 39, No. 7, pp. 565–591.

Dowling, R., Skabardonis, A., and Alexiadis, V. (2004). Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Software. Final Report, FHWA-HRT-04-040.

Dowling, R. G., Kittelson, W., Skabardonis, A., and Zegeer, J. (1997). Techniques for Estimating Speed and Service Volumes for Planning Applications. NCHRP Report 387:TRB, National Research Council, Washington, D.C.

Elefteriadou, L., Roess, R.P., and McShane, W.R. (1995). Probabilistic Nature of Breakdown at Freeway Merge Junctions. Transportation Research Record, No.1484.

Elefteriadou, L. and Lertworawanich, P., (2003). Defining, Measuring, and Estimating Freeway Capacity. Presented at the 82nd Annual Meeting of the Transportation Research Board.

Federal Highway Administration (FHWA). (2008a). Policy Options Evaluation Tool for Managed Lanes (POET-ML) Users Guide and Methodology Description, Federal Highway Administration HOV Lane Performance. http://ops.fhwa.dot.gov/publications/fhwahop09031/, Accessed Oct. 10, 2013.

Federal Highway Administration (FHWA). (2008b). Value Pricing Pilot Program TRUCE 3.0 Users Guide, Federal Highway Administration HOV Lane Performance. http://ops.fhwa.dot.gov/tolling_pricing/value_pricing/tools/truce_model_guide.htm, Accessed Oct. 10, 2013.

Federal Highway Administration (FHWA). (2008c). Managed Lanes: A Primer. http://ops.fhwa.dot.gov/publications/managelanes_primer/managed_lanes_primer.pdf, Accessed Oct. 20, 2013.

Florida Department of Transportation (FDOT). (2013). Southeast Florida Regional Planning Model, v. 6.5.3 (SERPM). Florida Department of Transportation, Tallahassee, Florida. http://www.fsutmsonline.net/index.php?/model_pages/model_pages/, Accessed Oct. 20, 2013.

Florida's Turnpike Enterprise, (2005) 2005 Annual Customer Service Satisfaction Survey, http://www.floridasturnpike.com/downloads/CustomerSurveys/2005custsat_web.pdf, Accessed Nov 4th, 2013.

Goodall, N., and Smith, B. L. (2010). What Drives Decisions of Single-Occupant Travelers in High-Occupancy Vehicle Lanes? Transportation Research Record: Journal of the Transportation Research Board, No. 2178, 156-161.

Gupta, A. (2005). Observability of Origin-Destination Matrices for Dynamic Traffic Assignment, Master's thesis, Mass. Inst. Technol

Hadi. M, Ozen, H, Shabanian, S, Xiao, Y, Zhao, W, Ducca, F. (2012). Use of Dynamic Traffic Assignment in FSUTMS in Support of Transportation Planning in Florida. Technical report submitted to Florida Department of Transportation.

Hall, F. L., and Agyemang-Duah, K. (1991). Freeway Capacity Drop and the Definition of Capacity, Transportation Research Record: Journal of the Transportation Research Board, No. 1320, pp. 91–98.

Hansen, S., Byrd, A., Delcambre, A., Rodriguez, A., and Bertini, R.L. (2005). Using Archived ITS Data to Improve Regional Performance Measurement and Travel Demand Forecasting 2005 CITE Quad/WCTA Regional Conference Vancouver, BC, Canada.

Hu, S.-R. and Chen, C.-B. (2004). Dynamic Estimation of Freeway Origin-Destination Demand and Travel Time Using Extended Kalman Filtering Algorithm, In Proceedings of IEEE Conference on Networking, Sensing and Control, pp.1329-1334.

Huntsinger. L., and Rouphail, N. (2011). Calibrating Travel Demand Model Volume-Delay Function Using Bottleneck and Queue Analysis. Presented at 90th Annual Meeting of the Transportation Research Board, Washington DC.

Janson, M. and Levinson, D. (2013). HOT or Not, Driver Elasticity to Price on the MnPASS HOT Lanes. http://nexus.umn.edu/Papers/HOTorNOT.pdf, Accessed Oct. 30, 2013.

Jia, A., Williams, B.M., and Rouphail, N.M. (2010). Identification and calibration of site specific stochastic freeway breakdown and queue discharge. Transportation Research Record: Journal of the Transportation Research Board, No. 2188, pp. 148-155.

Kunde, K. K. (2002). Calibration of Mesoscopic Traffic Simulation Models for Dynamic Traffic Assignment. Master Thesis. Massachusetts Institute of Technology.

Lin, W. (2006). A Robust Model for Estimating Freeway Dynamic Origin-Destination Matrices. PhD. Dissertation, Department of Civil and Environmental Engineering University of Maryland, College Park. MD.

Lin, D.Y., (2009). A Dual Approximation Framework For Dynamic Network Analysis: Congestion Pricing, Traffic Assignment Calibration, and Network Design Problem, Ph.D. Dissertation, University of Texas at Austin.

Lu, S., Nie, Y., (2010). Stability of User-equilibrium Route Flow Solutions for the Traffic Assignment Problem. Transportation Research Part B, Vol. 44, No. 4, pp.609–617.

Loudon, W. R. (2007). Improving the Estimation of Potential Travel-Time Savings from HOV Lanes. In Transportation Research Board Annual Meeting. Washington, DC.

Mahmassani, H., Qin, X., and Zhou, X. (2004). DYNASMART-X Evaluation for Real-Time TMC Application: Irvine Test Bed. Technical report, Maryland Transportation Initiative, University of Maryland.

Minderhoud, M.M., Botma, H., and Bovy, P.H.L. (1997). Roadway Capacity Estimation Methods Explained and Assessed. Transportation Research Record: Journal of the Transportation Research Board, No. 1572.

National Cooperative Highway Research Program (NCHRP). (2012). Assessing Highway Tolling and Pricing Options and Impacts, Volume II: Travel Demand Forecasting Tools. Transportation Research Board, Washington, D.C. http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_722v2.pdf, Accessed Oct., 20, 2013.

Nava, E., Milster, K., Chiu, Y.C., (2013). Modeling Managed Lane/Express Lane with Dynamic Traffic Assignment, Submitted for presentation at the 93nd Transportation Research Board Annual Meeting, Washington D.C.

Nguyen S. (1982). Estimating Origin-Destination Matrices from Observed Volumes. In Proceedings of the First Course on Transportation Planning Models of the International School of Transportation Planning, Amalfi, Italy.

Ortúzar, J.D. and Willumsen, L.G. (2001). Modeling Transport (3rd Edition). John Wiley & Sons, Chichester, England.

Parsons Brinckerhoff. (2013). Managed Lane Modeling Application for FSUTMS, Phase II, Memorandum to Advanced Traffic Assignment Subcommittee. Florida Department of Transportation, http://www.fsutmsonline.net/images/uploads/ManagedLane_Phase_II_Draft_2_2.docx, Accessed Oct. 23, 2013.

Peeta, S., Kumar, A., and Sharma, S. (2011). Post-Processing Techniques to Enhance Reliability of Assignment Algorithm Based Performance Measures. Publication FHWA/IN/JTRP-2011/19. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, doi: 10.5703/1288284314643.

Rakha, H., and Arafeh, M. (2010). Calibrating Steady-State Traffic Stream and Car-Following Models Using Loop Detector Data. Transportation Science, Vol. 44, No. 2, pp. 151-168.

Resource Systems Group, (2012). South Florida Expressway Stated Preference Survey Report. Prepared for URS Corporation.

Resource Systems Group, (2010). Convergence Peer Exchange Read Ahead Information.

Ruegg, S., Kulshrestha, A., and Puppala, S. (2013). Phase I Managed Lane Assignment Model Testing I-95 Corridor, Orlando and Tampa Regional Model. Memorandum to Advanced Traffic Assignment Subcommittee, Florida Department of Transportation, http://www.fsutmsonline.net/images/uploads/ManagedLaneTesting_PhaseI.docx, Accessed Oct. 20, 2013.

Saberi, M. (2010). Investigating Freeway Speed-Flow Relationships for Traffic Assignment Applications M.S. Thesis, Portland State University.

Safirova, E., Gillingham, K., Harrington, W., and Nelson, P. (2003). Resources for the Future, Are HOT Lanes a Hot Deal? http://www.rff.org/rff/Documents/RFF-IB-03-03.pdf, Accessed Oct. 20, 2013.

Schrank, D. L. and Lomax, T. J. (2007). The 2007 Urban Mobility Report, Texas Transportation Institute, Texas A & M University.

Singh, R. (1995). Beyond the BPR Curve: Updating Speed-Flow and Speed-Capacity Relationships in Traffic Assignment. Presented at 5th Conference on Transportation Planning Methods Applications, Seattle, Washington.

Sloboden, J., Lewis, J., Alexiadis, V., Chiu, Y., and Nava, E. (2012). Traffic Analysis Toolbox Volume XIV: Guidebook on the Utilization of Dynamic Traffic Assignment in Modeling. Final Report, FHWA-HOP-13-015, U.S. Department of Transportation, Federal Highway Administration.

Spiess, H. (1990a). Conical Volume-Delay Functions. Transportation Science, Vol. 24, No. 2.

Spiess, H. (1990b). A Gradient Approach for the OD Matrix Adjustment Problem. Vol. 693. Publication.

Tavana, H. (2001). Internally-consistent Estimation of Dynamic Network Origin-Destination Flows from Intelligent Transportation Systems Data Using Bi-level Optimization. Ph.D. Dissertation, University of Texas at Austin.

Transportation Research Board. (2000). Highway Capacity Manual, Washington, D.C.

Transportation Research Board. (2010). Highway Capacity Manual, Washington, D.C.

Transportation Research Board Planning Applications Conference, Reno, Nevada.

Tung, R. and Chiu, Y.C. (2011). Integration of Dynamic Traffic Assignment in a Four-Step Model Framework, A Deployment Case Study in PSRC (Puget Sound Regional Council) Model. Transportation Research Board Planning Applications Conference, Reno, Nevada.

URS, Inc. (2011). Toll Road Modeling Techniques, Published by the Maricopa Association of Governments, Headquartered in Phoenix, AZ. http://www.azmag.gov/addons/MAG/download.asp?ID=10890, Accessed Nov. 3, 2013.

Van Aerde, M., and Rakha, H. (1995). Multivariate Calibration of Single Regime Speed-Flow-Density Relationships VNIS/Pacific Rim Conference, Seattle, WA.

Van Arem, B. and van der Vlist, M. J. M. (1992). An On–line Procedure for Estimating Current Capacity. Technical Report INRO-VVG 1991-17, TNO Institute of Spatial Organization, Delft.

Vaze, V. S. (2007). Calibration of Dynamic Traffic Assignment Models with Point-to-Point Traffic Surveillance. Master thesis. Massachusetts Institute of Technology.

Vovsha, P., Davidson, W., and Donnelly, R. (2006). Making the State of the Art the State of the Practice: Advanced Modeling Techniques for Road Pricing. In Expert Forum on Road Pricing and Travel Demand Modeling.

Vovsha, P., Donnelly, B., and Bradley, M. (2013). Impact of Congestion Pricing and Travel Time Reliability on Travel Demand, Managed Lane Modeling Practice Workshop. Miami, Florida.

Wang, Y. et al. (2009). An adaptive freeway traffic state estimator, Automatica, Vol. 45, pp.10-24.

Wang, Y., Liu, X., and Rouphail, N. (2012). Analysis of Managed Lanes on Freeway Facilities, Final Report for NCHRP Project 03-96.

Wang, Y., Papageorgiou, M., Messmer, A., Coppola, P., Tzimitsi, A., and Nuzzolo. A. (2009). An Adaptive Freeway Traffic State Estimator. Automatica. Vol. 45, pp.10-24.

Washburn, Scott S., Modi, V., and Kulshrestha, A. (2010). Investigation of Freeway Capacity: a) Effect of Auxiliary Lanes on Freeway Segment Volume Throughput and b) Freeway Segment Capacity Estimation for Florida Freeways. No. TRC-FDOT-73157-2010.

WilburSmith Associates, (2011). Investment Grade Traffic and Revenue Study US. 36 Managed Lane. Final Report submitted to Colorado Department of transportation, http://www.coloradodot.info/library/studies/us-36-managed-lanes-investment-grade-traffic-and-revenue-study/Final-20Report_080911.pdf, Accessed Oct 25, 2013.

Xiao, Y., Hadi, M., Ozen, H., and Mysore, V. (2010). An ITS Evaluation Tool in the FSUTMS Regional Demand Modeling Environment. Transportation Research Record: Journal of the Transportation Research Board, No. 2176, pp. 76-83.

Yeon J., Hernandez, S., and Elefteriadou, L., (2007). Differences in Freeway Capacity by Day of the Week, Time of Day, and Segment Type. In Proceedings of the 86th Annual Meeting of the Transportation Research Board, Washington, D.C.

Zhang, L. and Levinson, D. (2004). Some Properties of Flows at Freeway Bottlenecks. In: Proceedings of the 83rd Transportation Research Board Annual Meeting, Washington, D.C.

Zhou, X. (2004). Dynamic Origin-Destination Demand Estimation and Prediction for Offline and Online Dynamic Traffic Assignment Operation. Ph.D. Dissertation, University of Maryland.

Zijpp, N. J. and Hamerslag, R. (1994). Improved Kalman Filtering Approach for Estimating Origin-Destination Matrices for Freeway Corridors. Transportation Research Record: Journal of the Transportation Research Board, No. 1443, pp. 54-64.

Appendix A: Cube Avenue Scripts for Prototype Managed Lane Model

```
RUN PGM=AVENUE PRNFILE="{CATALOG_DIR}\Model\MLAVN01A.PRN" MSG='DTA - Managed Lanes'
FILEI NETI = "{SCENARIO_DIR}\DTA_Highway.NET"
FILEI MATI[1] = "{SCENARIO_DIR}\SOV_TS_12.MAT"
FILEI MATI[2] = "{SCENARIO_DIR}\SR3_TS_12.MAT"
FILEI MATI[3] = "{SCENARIO_DIR}\Truck_TS_12.MAT"
FILEI LOOKUPI[1] = "{CATALOG_DIR}\Input_Data\NoWilling_to_Pay_Proportions.dbf"
FILEI LOOKUPI[2] = "{CATALOG_DIR}\Input_Data\TollCost_by_Density.dbf"

FILEO NETO = "{SCENARIO_DIR}\DTA_Managed_Lanes_Loaded.NET",
      INCLUDE=LW.TOLL1,LW.TOLL2,LW.TOLL3,LW.TOLL4,LW.TOLL5,LW.TOLL6,
              LW.TOLL7,LW.TOLL8,LW.TOLL9,LW.TOLL10,LW.TOLL11,LW.TOLL12
FILEO PACKETLOG = "{SCENARIO_DIR}\DTA_Managed_Lanes_Loaded.LOG"
FILEO MATO[1] = "{SCENARIO_DIR}\SKIM_FreeRoute.MAT",
 MO=201-224, DEC=24*9, COMBINE=F
FILEO MATO[2] = "{SCENARIO_DIR}\SKIM_TollRoute.MAT",
 MO=261-284, DEC=24*9, COMBINE=F
FILEO MATO[3] = "{SCENARIO_DIR}\SOV_TollTrips.MAT",
 MO=321-332, DEC=12*9, COMBINE=F
FILEO MATO[4] = "{SCENARIO_DIR}\SOV_FreeTrips.MAT",
 MO=341-352, DEC=12*9, COMBINE=F
FILEO PRINTO[1] = "{SCENARIO_DIR}\TEMP_1.PRN"

ARRAY _MAX_DENSITY=100                   ; maximum density for each direction in each corridor (11=SB & 12=NB for I-95 HOT)
ARRAY TOLL_SEG1_SB=20, TOLL_SEG1_NB=20   ; toll values for each time segment (12) in I-95 HOT lanes (segment=1)

;--- look-up table for no willingness-to-pay curve (percents by average toll-cents)
LOOKUP LOOKUPI=1, NAME=DIVERT,                      ; VOT distribution as reported by NuStats
       LOOKUP[1]=AVG_TOLL, RESULT=FREEPER_A,        ; A - urban & rural shot
       LOOKUP[2]=AVG_TOLL, RESULT=FREEPER_B,        ; B - long-distance business
       LOOKUP[3]=AVG_TOLL, RESULT=FREEPER_C,        ; C - long-distance tourist
       LOOKUP[4]=AVG_TOLL, RESULT=FREEPER_D,        ; D - short, cross-border EI
       LOOKUP[5]=AVG_TOLL, RESULT=FREEPER_E,        ; E - long-distance US, Canada
       LOOKUP[6]=AVG_TOLL, RESULT=FREEPER_F,        ; F - medium truck
       LOOKUP[7]=AVG_TOLL, RESULT=FREEPER_G,        ; G - unused
       LOOKUP[8]=AVG_TOLL, RESULT=FREEPER_H,        ; H - light truck
       INTERPOLATE=Y, FAIL=5,100

;--- look-up table for toll policy - dynamic tolls (cents)
LOOKUP LOOKUPI=2, NAME=TOLL_HOT,                    ; LOS-toll table
       LOOKUP[1]=DENSITY, RESULT=TOLLCOST,          ; min=$0.25 & max=$7
       INTERPOLATE=Y, FAIL=0.01,7

;seed the random number generator to ensure repeatability
seed = RANDSEED(12345)
;set total number of zones
PARAMETERS ZONES={ZONES}
;set dynamic traffic assignment methodology
PARAMETERS COMBINE=AVE, PACKETS=PA, GENPKTBYITER=T, ITERLOADINC={ITERLOADINC}, MAXITERS={MAXITER}
;set model period and time segment list
PARAMETERS MODELPERIOD=180, SEGMENTS=12*15 ; fifteen-minute time interval
;set assumptions for default storage (vehicles per lane per mile)
PARAMETERS VEHPERDIST={VEHPERDIST}
PARAMETERS PRESERVEMW=201-224,261-284,321-332,341-352
REPORT SPEED=YES CAPACITY=YES       ;report speed/capacity tables in network

PROCESS PHASE=LINKREAD
SPEED=LI.SPEED
T0=LI.TIME
C=LI.CAPACITY
STORAGE=LI.STORAGE

IF (STORAGE=0) STORAGE=9999999
IF (C=0)  C=9999999

;--- link class
IF (LI.FT=51,52)                  ; zonal centroid connector
   LINKCLASS= 1
ELSE                              ; other roads
   LINKCLASS= 2
ENDIF

;--- link activations
IF (LI.TOLL_LINK=1)               ; fixed toll lanes
   ADDTOGROUP=1
ELSEIF (LI.HOV_LINK=1)            ; HOV lanes (free)
   ADDTOGROUP=2
ELSEIF (LI.HOT_LINK=1)            ; HOT lanes
   ADDTOGROUP=3
ENDIF

;--- toll costs
IF (ITERATION=0)
   T1=LI.TIME
   LW.TOLL1 =LI.HOT_SEG_T         ; HOT entrance link- initial toll ($0.25 - time segment 1 )
   LW.TOLL2 =LI.HOT_SEG_T         ; HOT entrance link- initial toll ($0.25 - time segment 2 )
   LW.TOLL3 =LI.HOT_SEG_T         ; HOT entrance link- initial toll ($0.25 - time segment 3 )
   LW.TOLL4 =LI.HOT_SEG_T         ; HOT entrance link- initial toll ($0.25 - time segment 4 )
```

```
   LW.TOLL5 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 5 )
   LW.TOLL6 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 6 )
   LW.TOLL7 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 7 )
   LW.TOLL8 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 8 )
   LW.TOLL9 =LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 9 )
   LW.TOLL10=LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 10)
   LW.TOLL11=LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 11)
   LW.TOLL12=LI.HOT_SEG_T          ; HOT entrance link- initial toll ($0.25 - time segment 12)
ENDIF
ENDPROCESS

PROCESS PHASE=ILOOP
;--- input trip matrices
FILLMW MW[301]=MI.1.1(12)           ; SOV+SR2  for 12 time segments
FILLMW MW[401]=MI.2.1(12)           ; SR3P  for 12 time segments
FILLMW MW[451]=MI.3.1(12)           ; truck for 12 time segments

;--- update LW toll variables in ILOOP
IF (ITERATION>1 & TIMESEGMENT=0 & I=1)  ; before performing skimming
   LINKLOOP
      IF (LI.HOT_SEG_R=11)          ; toll for each time segment in I-95 HOT SB ramp
         LW.TOLL1 =TOLL_SEG1_SB[1]
         LW.TOLL2 =TOLL_SEG1_SB[2]
         LW.TOLL3 =TOLL_SEG1_SB[3]
         LW.TOLL4 =TOLL_SEG1_SB[4]
         LW.TOLL5 =TOLL_SEG1_SB[5]
         LW.TOLL6 =TOLL_SEG1_SB[6]
         LW.TOLL7 =TOLL_SEG1_SB[7]
         LW.TOLL8 =TOLL_SEG1_SB[8]
         LW.TOLL9 =TOLL_SEG1_SB[9]
         LW.TOLL10=TOLL_SEG1_SB[10]
         LW.TOLL11=TOLL_SEG1_SB[11]
         LW.TOLL12=TOLL_SEG1_SB[12]
      ENDIF
      IF (LI.HOT_SEG_R=12)          ; toll for each time segment in I-95 HOT NB ramp
         LW.TOLL1 =TOLL_SEG1_NB[1]
         LW.TOLL2 =TOLL_SEG1_NB[2]
         LW.TOLL3 =TOLL_SEG1_NB[3]
         LW.TOLL4 =TOLL_SEG1_NB[4]
         LW.TOLL5 =TOLL_SEG1_NB[5]
         LW.TOLL6 =TOLL_SEG1_NB[6]
         LW.TOLL7 =TOLL_SEG1_NB[7]
         LW.TOLL8 =TOLL_SEG1_NB[8]
         LW.TOLL9 =TOLL_SEG1_NB[9]
         LW.TOLL10=TOLL_SEG1_NB[10]
         LW.TOLL11=TOLL_SEG1_NB[11]
         LW.TOLL12=TOLL_SEG1_NB[12]
      ENDIF
   ENDLINKLOOP
ENDIF

;--- building paths
DYNAMICLOAD PATH=COST, EXCLUDEGRP=2-3,                         ; SOV free roads
            MW[201]=TRACE(0,TIME),             NOACCESS=0,   ; time segment 1
            MW[202]=TRACE(0,LW.TOLL1),         NOACCESS=0,
            MW[203]=TRACE(15,TIME),            NOACCESS=0,   ; time segment 2
            MW[204]=TRACE(15,LW.TOLL2),        NOACCESS=0,
            MW[205]=TRACE(30,TIME),            NOACCESS=0,   ; time segment 3
            MW[206]=TRACE(30,LW.TOLL3),        NOACCESS=0,
            MW[207]=TRACE(45,TIME),            NOACCESS=0,   ; time segment 4
            MW[208]=TRACE(45,LW.TOLL4),        NOACCESS=0,
            MW[209]=TRACE(60,TIME),            NOACCESS=0,   ; time segment 5
            MW[210]=TRACE(60,LW.TOLL5),        NOACCESS=0,
            MW[211]=TRACE(75,TIME),            NOACCESS=0,   ; time segment 6
            MW[212]=TRACE(75,LW.TOLL6),        NOACCESS=0,
            MW[213]=TRACE(90,TIME),            NOACCESS=0,   ; time segment 7
            MW[214]=TRACE(90,LW.TOLL7),        NOACCESS=0,
            MW[215]=TRACE(105,TIME),           NOACCESS=0,   ; time segment 8
            MW[216]=TRACE(105,LW.TOLL8),       NOACCESS=0,
            MW[217]=TRACE(120,TIME),           NOACCESS=0,   ; time segment 9
            MW[218]=TRACE(120,LW.TOLL9),       NOACCESS=0,
            MW[219]=TRACE(135,TIME),           NOACCESS=0,   ; time segment 10
            MW[220]=TRACE(135,LW.TOLL10),      NOACCESS=0,
            MW[221]=TRACE(150,TIME),           NOACCESS=0,   ; time segment 11
            MW[222]=TRACE(150,LW.TOLL11),      NOACCESS=0,
            MW[223]=TRACE(165,TIME),           NOACCESS=0,   ; time segment 12
            MW[224]=TRACE(165,LW.TOLL12),      NOACCESS=0
DYNAMICLOAD PATH=COST, EXCLUDEGRP=2,                           ; SOV HOT-lane toll roads
            MW[261]=TRACE(0,TIME),             NOACCESS=0,
            MW[262]=TRACE(0,LW.TOLL1),         NOACCESS=0,
            MW[263]=TRACE(15,TIME),            NOACCESS=0,
            MW[264]=TRACE(15,LW.TOLL2),        NOACCESS=0,
            MW[265]=TRACE(30,TIME),            NOACCESS=0,
            MW[266]=TRACE(30,LW.TOLL3),        NOACCESS=0,
            MW[267]=TRACE(45,TIME),            NOACCESS=0,
            MW[268]=TRACE(45,LW.TOLL4),        NOACCESS=0,
            MW[269]=TRACE(60,TIME),            NOACCESS=0,
            MW[270]=TRACE(60,LW.TOLL5),        NOACCESS=0,
            MW[271]=TRACE(75,TIME),            NOACCESS=0,
            MW[272]=TRACE(75,LW.TOLL6),        NOACCESS=0,
            MW[273]=TRACE(90,TIME),            NOACCESS=0,
            MW[274]=TRACE(90,LW.TOLL7),        NOACCESS=0,
            MW[275]=TRACE(105,TIME),           NOACCESS=0,
            MW[276]=TRACE(105,LW.TOLL8),       NOACCESS=0,
            MW[277]=TRACE(120,TIME),           NOACCESS=0,
```

```
            MW[278]=TRACE(120,LW.TOLL9),       NOACCESS=0,
            MW[279]=TRACE(135,TIME),           NOACCESS=0,
            MW[280]=TRACE(135,LW.TOLL10),      NOACCESS=0,
            MW[281]=TRACE(150,TIME),           NOACCESS=0,
            MW[282]=TRACE(150,LW.TOLL11),      NOACCESS=0,
            MW[283]=TRACE(165,TIME),           NOACCESS=0,
            MW[284]=TRACE(165,LW.TOLL12),      NOACCESS=0

;--- applying willingness-to-pay proportions for free vs. toll
IF (TIMESEGMENT>0)
   _TS=TIMESEGMENT
   _IDX1=300+_TS                             ; input SOV total trip index
   _IDX2=320+_TS                             ; for SOV free trip index
   _IDX3=340+_TS                             ; for SOV toll trip index
   _T_IDX=200+_TS*2-1                        ; index for free time skim
   _C_IDX=260+_TS*2                          ; index for toll skim
   IF (I=1) _ODFLAG=0                        ; flag to check out zero OD trips for toll trips
   ;--- SOV diversion process for HOT-lanes
   JLOOP
      IF (MW[_IDX1]>0)                       ; if there exist SOV trips
         TSAVE=MW[_T_IDX]-MW[_T_IDX+60]      ; time saving (free time - toll time)
         IF (TSAVE>0 & MW[_C_IDX]>0)         ; if positive time saving & positive toll cost($)
            TCPS=(MW[_C_IDX]*100)/TSAVE      ; average toll cost (cents) per minute saved
            PWPT=100-DIVERT(4,TCPS)          ; proportion (%) willing to pay toll by average toll-cents
         ELSE
            PWPT=0
         ENDIF
         MW[_IDX3]=MW[_IDX1]*(PWPT/100)      ; SOV toll trips
         MW[_IDX2]=MW[_IDX1]-MW[_IDX3]       ; SOV free trips
         IF (_ODFLAG=0 & MW[_IDX3]>0) _ODFLAG=1   ; if there are toll trips, change the flag for toll OD trips
      ELSE
         MW[_IDX3]=0
         MW[_IDX2]=0
      ENDIF
   ENDJLOOP
ENDIF

;--- loading free and toll trips
DYNAMICLOAD PATH=COST, BUILDALLPATHS=T VOL[1]=MW[(320+__TS__)], PACKETSIZE={PacketSize}, EXCLUDEGRP=2-3   ;  SOV   free trips
DYNAMICLOAD PATH=COST, BUILDALLPATHS=T VOL[2]=MW[(340+__TS__)], PACKETSIZE={PacketSize}, EXCLUDEGRP=2     ;  SOV   HOT-lane toll
trips
DYNAMICLOAD PATH=COST, BUILDALLPATHS=T VOL[3]=MW[(401+__TS__)], PACKETSIZE={PacketSize}                   ;  HOV   free trips
DYNAMICLOAD PATH=COST, BUILDALLPATHS=T VOL[4]=MW[(451+__TS__)], PACKETSIZE={PacketSize}, EXCLUDEGRP=2-3   ;  Truck free trips
ENDPROCESS

PROCESS PHASE=ADJUST
FUNCTION {
   V=VOL[1]+VOL[2]+VOL[3]+VOL[4]
   TC[1]=T0
   TC[2]=T0*(1.0 + LI.COEF*((V/C)^LI.EXPO))
   COST=TIME}

;--- updating tolls based on maximum density for each directional corridor for I-95 (segment 1)
IF (TIMESEGMENT>0)           ; for each time segment
   _CTS=TIMESEGMENT          ; current time segment
   IF (LINKNO=1)
      LOOP _I=11,12          ; maximum density for each HOT direction (1=SB & 2=NB) for I-95 (segment 1)
         _MAX_DENSITY[_I]=0
      ENDLOOP
      TOLL_SEG1_SB[_CTS]=0    ; toll value for each time segment for each HOT direction in I-95 (segment 1)
      TOLL_SEG1_NB[_CTS]=0
   ENDIF

   ;--- computing link density (vpmpl=vehicle per mile per lane)
   ; density = (vehicle/hour)/(mile/hour) = hourly-vehicle/speed = vehicle per mile per lane
   IF (V>0 & LI.DISTANCE>0 & TIME>0 & LI.LANES>0)
      _SPEED=(LI.DISTANCE/TIME)*60
      _LINK_DEN=((V*(60/{IntTimeSeg}))/_SPEED)/LI.LANES
   ELSE
      _SPEED=0
      _LINK_DEN=0
   ENDIF

   ;--- updating toll values based on maximum density for each direction in each HOT corridor
   IF (LI.HOT_SEG_C=11)       ; I-95 (segment 1) HOT SB corridor links (11=SB & 12=NB)
      IF (_MAX_DENSITY[11]<_LINK_DEN)
         _MAX_DENSITY[11]=_LINK_DEN
         TOLL_SEG1_SB[_CTS]=MIN(7,TOLL_HOT(1,_MAX_DENSITY[11]))   ; toll in I-95 HOT SB ramp
     ENDIF
   ELSEIF (LI.HOT_SEG_C=12)   ; I-95 (segment 1) HOT NB corridor links (11=SB & 12=NB)
      IF (_MAX_DENSITY[12]<_LINK_DEN)
         _MAX_DENSITY[12]=_LINK_DEN
         TOLL_SEG1_NB[_CTS]=MIN(7,TOLL_HOT(1,_MAX_DENSITY[12]))   ; toll in I-95 HOT NB ramp
      ENDIF
   ENDIF
ENDIF
ENDPROCESS

ENDRUN
```