ABSTRACT

Title of Dissertation:	ONLINE and REAL-TIME TRANSPORTATION SYSTEMS MANAGEMENT and OPERATIONS DECISION SUPPORT WITH INTEGRATED TRAVEL BEHAVIOR and DYNAMIC NETWORK MODELS
	Zheng Zhu, Doctor of Philosophy, 2018
Dissertation directed by:	Lei Zhang, Professor Department of Civil & Environmental Engineering

The acceleration of urbanization is witnessed all around the world. Both population and vehicle ownership are rapidly growing, and the induced traffic congestion becomes an increasingly pervasive problem in people's daily life. In 2014, transportation congestion caused $160 billion economic loss in 498 U.S. urban areas, which is 5.5 more than that in 1982. Without effective reactions, this number is expect to grow to $192 billion in 2020. In order to mitigate traffic congestion, many transportation demand management (TDM) strategies (e.g. bus rapid lanes, and flextime policy), and active traffic management (ATM) strategies (e.g. real-time user guidance, and adaptive traffic signal control) have been proposed and implemented.

Although TDM and ATM have proved their values in theoretical researches or field implementations, it is still hard for transportation engineers to select the optimal strategy when faced with complex traffic conditions. In the

science of transportation engineering, mathematical models are usually expected to help estimate traffic conditions under different scenarios. There have been a number of models that help transportation engineers make decisions. However, many of them are developed for offline use and are not suitable for real-time applications due to computational time issues. With the development of computational technologies and traffic monitoring systems, online transportation network modeling is getting closer and closer to reality. The objective of this dissertation is to develop a large-scale mesoscopic transportation model which is integrated with an agent-based travel behavior model. The ultimate goal is to achieve online (real-time) simulation to estimate and predict the traffic performance of the entire Washington D.C. area. The simulation system is expected to support real-time transportation system managements and operations.

One of the most challenging issue for this dissertation is the calibration of online simulation models. Model parameters need to be estimated based on real-time traffic data to reflect the reality. Literature review of previous relevant studies indicates a trade-off between computational speed and calibration accuracy. In order to apply the model onto a real-time horizon, experts usually ignore the inherent mechanism of traffic modeling but rely on fast converging technologies to approximate the model parameters. Differently from previous online transportation simulation approaches, the method proposed in this dissertation focuses more on the mechanism of transportation modeling. With the fundamental understanding of the modeling mechanism, one can quickly determine the gradient of model parameters such that the gap between real-time traffic measures and simulation results is minimized.

This research is one of the earliest attempts to introduce both agent-based modeling and gradient-based calibration approach to model real-time large-scale networks. The contribution includes: 1) integrate an agent-based travel behavior model into dynamic transportation network models to enhance the behavior realism; 2) propose a fast online calibration procedure that quickly adjusts model parameters based on real-time traffic data. A number of real-world case studies are illustrated to demonstrate the value of this model for both long-term and real-time applications.

ONLINE and REAL-TIME TRANSPORTATION SYSTEMS
MANAGEMENT and OPERATIONS DECISION SUPPORT WITH
INTEGRATED TRAVEL BEHAVIOR and DYNAMIC NETWORK
MODELS

by

Zheng Zhu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
  Professor Lei Zheng, Chair
  Professor Benjamin Kedem
  Professor Ali Haghani
  Professor Paul Schonfeld
  Professor Xuesong Zhou

Acknowledgements

It would not have been possible for me to finish this research without the help and support from so many people. It is a great pleasure to thank all the people that contributed to the preparation and the completion of this dissertation.

I am deeply indebted to my advisor, Dr. Lei Zhang, for his continuous support, inspiring advice and enthusiastic encouragement throughout the years of my Ph.D. study at the University of Maryland, College Park. I am grateful for the opportunity he gives me to work on this emerging and interesting topic, and his strong support for my second master in statistics. He is not only an ideal advisor in research, but also a great mentor in my pursuit of career and a good friend that cares about my personal life.

I would like to extend my thanks to the rest of my dissertation committee: Professor Benjamin Kedem, Professor Ali Haghani, Professor Paul Schonfeld and Professor Xuesong Zhou, for their encouragement, insightful comments and valuable suggestions to my research. Professor Benjamin Kedem is also my committee chair for my master degree in statistics. I would like to thank him again together with Professor Paul Smith for the kindly help on my second master.

My sincere thanks also go to Dr. Chenfeng Xiong and Dr. Xiqun (Michael) Chen, for the knowledge and insight they have shared. A lot of ideas in this research come from the fruitful discussion between us.

I thank all my colleagues in the National Transportation Center at the University of Maryland. The ideas and suggestions they shared with me during our group meetings have greatly contributed to the completion of this research. I would also

like to thank my fellow graduate students from other groups in the transportation program. I appreciate the opportunities to take class, cooperate on projects and have discussions with them.

Finally, I would like to express my gratitude towards my family and dedicate this dissertation to them. They have been encouraging me, supporting me and loving me all the time. I am especially thankful to my wife, Can Dong, for her understanding and love, and my daughter Zoe Zhu, for her coming to bring us endless happiness.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AgBM	Agent-based travel behavior model
AgBM-DTA	The system transportation network model integrated with agent-based travel behavior model and dynamic traffic assignment model
API	Application programming interface
ARPA-E	Advanced Research Projects Agency-Energy
ATIS	Advanced traffic information system
ATM	Active traffic management
BMC	Baltimore Metropolitan Council
BUE	Behavioral user equilibrium
CATT	Center For Advanced Transportation Technology
CTP	Consolidated transportation program
DOT	Department of Transportation
DTA	Dynamic traffic assignment
DUE	Dynamic user equilibrium
FFTT	Free-flow travel time
FHWA	Federal Highway Administration
IAP	Implementation assistance program
I-TMS	Internet traffic monitoring system
ICM	Integrated corridor management
ITS	Intelligent transportation system

MAS	Multi-agent system
SHRP2	The second Strategic Highway Research Program
SMTM	Maryland Statewide Transportation Model
MWCOG	Metropolitan Washington Council of Governments
OD	Origin-destination
ODME	OD estimation procedure in DTALite (the DTA package used in this dissertation)
PC	Principal component
PCA	Principal component analysis
RITIS	Regional integrated transportation information system
SB	Southbound
SBO	Surrogate-based optimization
SHA	State Highway Administration
SILK	Searching, information, learning, and knowledge behavior theory
TAZ	Traffic analysis zone
TDM	Transportation demand management
TPB	Transportation Planning Board
VMT	Vehicle miles traveled
VOT	Value of time
WB	Westbound
WISE	Work zone impacts and strategies estimator

WMSE   Weighted mean squared error

UE   User equilibrium

UMD   University of Maryland

Chapter 1: Introduction

Section 1.1 Background

The acceleration of urbanization is witnessed all around the world. In today's world, the pace of life is speeding up with technological advancements such that people's desire for travel is increasing rapidly. Meanwhile, due to the growth of economy and productivity technologies, vehicles and other transportation tools become more and more affordable and necessary. The induced traffic congestion becomes an increasingly pervasive problem in people's daily life. In 2014, transportation congestion caused $160 billion economic loss in 498 U.S. urban areas, which was 5.5 times more than that in 1982 (in 2014 dollars) (Schrank et al. 2015). The number is expect to become $192 billion in 2020 (in 2014 dollars). Travel demand is not only increasing in magnitude, but also getting more complex. With the multi-modal design of urban transportation systems, people have more alternative modes for their trips. For example, they can drive their own car to a transit station, and then take public transit to the destination; or, they can call Uber for a ride share. With limited transportation infrastructure and people's diverse desire for travel, it can be challenging for traffic management and congestion mitigation.

There are two ways to deal with traffic congestion: demand side management and supply side operation. The first one relies on some policies to either limit the access of users to highly demanded transportation infrastructures during peak periods, or encourage travelers to switch to lowly demanded infrastructures to avoid unnecessary

congestion. These approaches include vehicle plate restrictions (Yang et al. 2014), flextime work schedule policy (Zhu et al. 2015), bus rapid transit lane (Levinson et al. 2002), variable message signs (Mammar et al. 1996), and real time user guidance (Kaysi 1993). The second one attempts to adjust transportation infrastructure systems to improve urban mobility, e.g. freeway ramp metering (Papageorgiou et al. 1991), adaptive traffic signal control (Lowrie 1982), bus signal control, and congestion-based road pricing (Ison 1996). With the help of modern detection technology and information transformation technology, a number of real time demand or supply side traffic management strategies, referred as active traffic management (ATM), have been introduced into the field of transportation. As parts of intelligent transportation systems (ITS), ATM strategies play important roles to better manage the increasing travel demand and improve the reliability of transportation systems. ATM strategies may better utilize the capacity or provide additional capacity to accommodate peak-hour traffic, improve the detection and response to incidents, reduce delays resulted from recurrent congestion, and thus enhance the transportation network's performance in safety, efficiency, reliability, and sustainability. Real-time traffic management techniques are currently under operation in many metropolitan areas in the U.S., including adaptive ramp metering, hard shoulder running, dynamic lane use control, reversible lanes, and variable speed limit. Since transportation systems are getting more and more complex with their corresponding traffic transportation demand management (TDM) strategies and ATM strategies, advanced transportation modeling systems are required to evaluate different congestion mitigation strategies and even take real-time

estimations for decision support.

Section 1.2. Traffic modeling approaches

In order to evaluate the performance of transportation systems and to determine the best strategy for implementation, it is crucial to have a reliable evaluation model that fully incorporates human's travel behavior and traffic flow movement mechanism. It is important that such an evaluation model reflects network wide traffic condition under the interactions between users and transportation infrastructure. A transportation system was first modeled as a network in the 1950s, when Wardrop's equilibrium principle was introduced for static network performance (Wardrop 1952). The static equilibrium is usually used in a four-step transportation planning model, which assumes static supply and demand conditions for typical days or time periods within a day. Such a network model can hardly reflect the dynamic performance of a transportation system. Static traffic models can be used for policy oriented analysis, e.g. for the impact of vehicle mileage traveled (VMT) fee on traffic congestion, or for the effectivity of value of time (VOT) on travel behavior. However, the limitation in dynamic traffic analysis impedes its further applications in ATM. In addition, some important issues in demand pattern shifts, such as trip chaining and scheduling, are difficult to describe by a four-step model due to the lack of solid behavioral foundation (Zhang 2007).

Dynamic traffic assignment (DTA) models have been used to capture time dependent traffic phenomena (Janson 1991). DTA models have been a valuable and

powerful approach in the transportation modeling. Compared with static traffic assignment models, the major benefit of using DTA is its advance in calculating spatial and temporal traffic conditions, determining departure time choices, mode choices, and time dependent route choices. The capabilities of obtaining time-dependent traffic phenomena allow the testing of a multitude of transportation conditions, e.g. bottleneck studies (Gentile et al. 2007); integrated corridor management (ICM) strategies (Zhou et al. 2008); TDM strategies (Yang et al. 2000); additional capacity and land development studies (Zhu et al. 2015); and adaptive control studies (Gartner and Stamatiadis 1998). The benefits of using DTA in transportation analysis increase with the size of the study area because researchers can obtain the performance under different zooming levels (e.g., specific links, intersections, corridors, or even regions).

With modern computation technologies, DTA models have been developed for large-scale network studies. However, the increase of network size also brings up computational time issues, so that real-time traffic analysis for large-scale transportation networks is very difficult. The calibration of a DTA model can be one of the most time consuming and most important processes when applying DTA to real-world analysis. A DTA model needs calibration and validation to reflect realistic real-world conditions. For large networks, calibration can be time consuming because the number of parameters to adjust grows rapidly with the network size. A sound calibration needs to consider both supply parameters and demand parameters in the DTA model components. Another limitation of many DTA models lays in their travel behavior foundation. The pursuit of user equilibrium (UE) in DTA may not be advanced in

depicting how people actually learn about traffic information and make decisions. This is because DTA usually relies on the rational behavior foundation, which assumes that travelers have perfect information about traffic conditions and are able to make optimal decisions. There is still room for improving DTA's travel behavior foundations to reflect people's actual reactions towards ATMs. In addition, real-time DTA can be a promising direction since traffic data is getting more and more abundant. Online applications of DTA models can play important roles in real-time transportation condition estimation, prediction, and ATM evaluation.

Section 1.3. Real-time agent-based traffic modeling

With the development of computational technologies, mathematical models and software become increasingly agent-oriented. Agent-oriented computing is one of the powerful technologies for the development of distributed complex systems (Zambonelli and Parunak 2002). Agent-oriented techniques enable the design, analysis and implementation of large and complex systems by providing simpler logics to solve complex problems in these systems. The concept of "agent" refers to a single component of the distributed system that is operated via some simple rules that will interact with other agents as well as the system. An agent is intelligent to learn the current system's condition via the environment and the communication with other agents. Then the agent makes decisions and execute its functionality in the system. A multi-agent system (MAS) is a modeling system composed of multiple agents. The condition of the MAS is obtained via the modeling of the agents and the environment.

Figure 1-1 basically illustrates how a MAS works, different kinds of agents (type A, B, and C) may have specific functionalities and goals. These agents may have one-directional information communication (A to B), or two directional information exchange (B and C). For instance, in Figure 1-1, type A can be farmers who want to sell vegetables to wholesalers (type B); wholesalers will contact factory (type C) to make products, and these products can be sold back to wholesalers and brought to the market (the environment). Agents work in the environment with the information they obtain to pursue their goals, which are to maximize their own profits in the example.
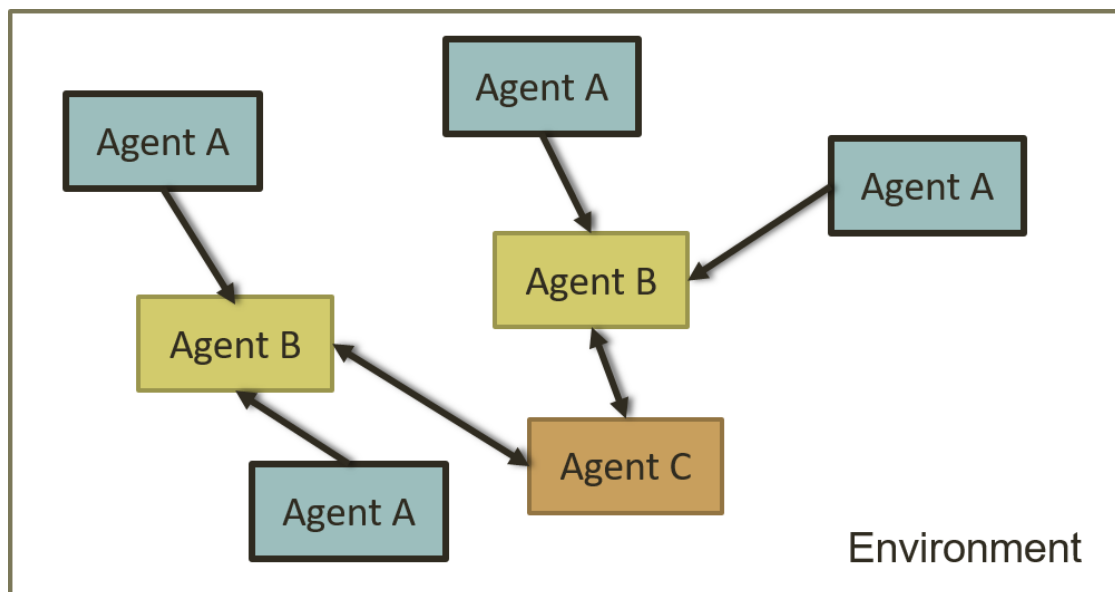
Figure 1- 1 How a MAS works

A transportation system, which has geographically distributed nature and stochastic operation characteristics, is well suitable to be modeled as a MAS (Chen and Cheng 2010). In a transportation modeling system, the demand can be modeled as traveling agents; and the supply side infrastructure such as roadway, traffic detectors,

traffic signals, traffic management agencies can all be represented by different types of agents. Because agents are autonomous, collaboration, and reactivity, applying agent-based technics is beneficial for both modeling and the implementation of ITS. From the traffic and transportation management perspective, agents can operate without the direct intervention of humans. This feature helps to model and implement automated traffic control and management systems. Agents communicate with other agents to collaborate with each other to perform traffic control and management based on real-time traffic conditions. In a transportation modeling system, advanced demand and behavior modeling can be achieved by treating travelers as intelligent agents. These intelligent agents can learn knowledge about their travel experience and make pre-trip or en-route decisions. Recently, increasingly agent-based traffic and transportation models and applications have been developed. In next chapter, a comprehensive literature review will show that MAS have been applied to many aspects of ITS, including modeling and simulation, intelligent traffic control and management, dynamic routing and congestion management, driver-infrastructure collaboration, and decision support (Chen and Cheng 2010).

Although there have been attempts of integrating MAS into traffic modeling systems, the illustrations under large-scale networks or real-time applications are rarely seen in literature. One major reason is because the computational time required for large-scale network is so high for researchers to conduct online (real-time) real-world applications. With the motivation of the online application of large-scale agent-based transportation modeling systems, this dissertation attempts to propose an online

applicable MAS, in which travelers are treated as agents that can adjust their behavior via an information updating and learning process.

Section 1.4. Summary

Based on the introductions above, I have discussed the limitations of current transportation modeling approaches and the advantages of using agent-based technologies. Therefore, I would develop an advanced transportation modeling system that is able to incorporate agent-based modeling and real-time simulation for both transportation planning and operation applications. The objective of this dissertation is to develop a large-scale mesoscopic transportation model which is integrated with an agent-based travel behavior model. The ultimate goal is to achieve online (or real-time) simulation to help estimate and predict the performance of the whole transportation system. This integrated model can not only assist in transportation planning impact analysis, but also work along with ATM for real-time decision support.

This research is one of the earliest attempts to introduce both agent-based modeling and online network modeling for large-scale networks. The contributions include: 1) integrate an agent-based travel behavior model into DTA models to enhance the behavior realism; 2) propose a gradient-based fast online calibration procedure that quickly adjusts model parameters based on real-time traffic data. The practical value of this tool can match its theoretical value because the implementation of the proposed tool is straightforward. A number of real-world and real-time case studies are illustrated to demonstrate the value of this model for both long-term and real-time applications.

The rest of the dissertation is organized as follows. Chapter 2 provides a comprehensive literature review on transportation modeling, offline/online calibration, and MAS modeling approaches in transportation engineering. Chapter 3 introduces the agent-based transportation system model in terms of its components and model integration. Chapter 4 proposes a gradient-oriented online calibration approach. Detailed methodologies on data processing, demand and supply calibration mechanism are discussed. Chapter 5 mainly illustrates the applications of this proposed transportation network model. Medium-scale case studies on both general days and incident days are conducted to illustrate the performance of the online calibration approach. The online modeling of a large-scale network that covers the whole Washington D.C. area is also shown in Chapter 5. Different ATMs are evaluated as a case study for real-time decision support. Chapter 6 extends the online system model to another important application: work zone schedule and operation. This is done to illustrate the capability of the model in a transportation planning application. Conclusions and future work are discussed in Chapter 7.

Chapter 2: Literature Review

In this chapter, I will provide a comprehensive literature review on different topics that are incorporated in this dissertation. Firstly, I will talk about transportation modeling approaches; then, the offline and online calibration of transportation models, especially the calibration of DTA models will be introduced; finally, transportation related MAS will be discussed.

Section 2.1. Transportation modeling approaches

Transportation modeling aims at estimating travel demand and the performance of transportation systems. Transportation modeling can be used both in planning and management applications. Within transportation planning frameworks, four-step models are among the earliest transportation modeling approaches. Four-step models were introduced during the 1950s by Detroit Metropolitan Area Traffic Study (Area 1955) and Chicago Area Transportation Study (Hoch 1959). The four steps are: trip generation, trip distribution, mode choice, and route assignment. Trip generation is conducted via economic based analysis to estimate the traffic analysis zone (TAZ) based trip frequency for different trip purposes (Hoch 1959). Trip distribution matches the TAZ to TAZ productions and attractions (Schneider 1959). Mode choice calculates the fraction of different travel modes among TAZ to TAZ trips (Ben-Akiva 1974; Daganzo 1979). Route assignment assigns the trips to the transportation network via Wardrop's equilibrium principle (Wardrop 1952). Four-step models usually represent traffic as a static phenomenon because traditional travel demand is considered to be

static during a specific time period; and traffic assignment solves a static programming problem. Traffic dynamics such as queuing and routing changes are not adequately taken into account.

DTA, which considers traffic flow based link costs, has been used to address the static traffic issue (Daganzo 1977; Merchant and Nemhauser 1978; Janson 1991; Wu et al. 1998). The pursuit of dynamic user equilibrium (DUE) enables the analysis time-dependent transportation performance. There are three categories of DTA models: macroscopic level models, mesoscopic level models, and microscopic models. Macroscopic level DTA models formulate traffic as mathematical functions between speed, volume and density (Inose 1967). Microscopic level DTA models attempt to model individual vehicles' movements via car following models and lane changing models (May 1967). Mesoscopic models utilize macro traffic models to represent transportation performance, but they track individual travelers' trip information (Ben-Akiva et al 1997). There have been several DTA packages in the academic area for time dependent transportation modeling. For macroscopic level DTA models, we have TransCAD (Caliper 2005) developed by Caliper, VISUM (PTV 2009) developed by PTV, etc. There are also mesoscopic DTA models such as DynaMIT (Ben-Akiva 2002), Dynasmart (Mahmassani et al. 2004), and DTALite (Zhou and Taylor 2014). Both macroscopic models and mesoscopic models are designed for corridor-level up to regional-level analysis of a wide range of transportation planning and traffic management applications. VISSIM (PTV 2008), SUMO (Behrisch et al. 2011), and TransModeler (Caliper 2009) are widely used in microscopic DTA applications.

Even though DTA gains its popularity by presenting detailed traffic dynamics, the pursuit of DUE may not be advanced in depicting how people actually learn about traffic information and make decisions (Zhang 2007). This is because DTA usually relies on the rational behavior foundation, which assumes that travelers have perfect information about traffic conditions and are able to make the optimal decisions. In addition, some important issues in demand pattern shifts, such as trip chaining and scheduling, are difficult to describe by these models due to the lack of solid behavioral foundation (Zhang, 2007). The modeling of traffic dynamics in DTA is not equivalent to modeling transportation because travelers' behavior is not fully considered. Chen et al. (2016) conducted a comprehensive review on how to apply different transportation data for behavior analysis. Chen's research team also did a several research on activity pattern prediction (Kitamura et al. 2000), mode choice prediction (Chen et al. 2008), activity duration prediction (Mokhtarian and Chen 2004), etc. However, these studies are not integrated with traffic-related models, such that time is still needed for traffic modeling to benefit from these behavior findings.

Activity-based travel demand models (Koppelman and Pas 1980; Carpenter and Jones 1983) and agent-based travel behavior models (Zhang 2007) have been proposed to enhance the behavior foundation in transportation models. These models theoretically promise a stronger behavioral foundation for demand modeling and are expected to provide more accurate time-dependent estimates of origin-destination (OD) demand than four-step models. The integration between activity/agent-based demand models and DTA techniques enhances the time-dependent advantage on both the

demand side and the supply side (Lin et al. 2008; Hao et al. 2010), which will be a new trend for modeling transportation systems.

Section 2.2. Calibration of transportation models

Calibration is one of the most important processes when applying DTA to the real-world analysis because the base model needs to reflect real-world conditions. For large networks, calibration can be time consuming because the number of parameters to adjust grows rapidly with the network size. A sound calibration needs to consider both supply parameters and demand parameters in the DTA model components. On one hand, the supply model components capture traffic dynamics via detailed representations of the capacities and the mobility of transportation facilities. Supply parameters include the features (e.g., free flow speed, capacity, and jam density) of different types of roadways, timing plans of traffic signals, merging/diverting rules when traffic flows arrive at a merging/diverting point, weaving behavior, etc. On the other hand, the demand model components provide travel patterns obtained from destination choice models, pre-trip departure time choice models, and mode choice models. Therefore, time-dependent OD matrices for different modes are obtained as the demand parameters. For each application, the supply and demand parameters need to be identified so that the model outputs match real-world traffic conditions observed via traffic detectors or probe vehicles in the study area. Today's information technologies provide rich sources of real-time and historical traffic data for ITS. As a result, sound online calibration approaches are required for large-scale real-time simulations to

implement ATMs.

The calibration of DTA models can be formulated as an optimization problem that adjusts the supply and demand parameters so that the gap between available traffic measures and the models' outputs is minimized. Since the DUE condition is already an optimization problem, the calibration of DTA can be regarded as a bi-level optimization with: 1) the upper level to optimize the supply and demand parameters to reach the minimal gap between simulation outputs and real-world data; 2) the lower level to converge to DUE given the supply and demand parameters. There are no closed formulations for the objective function (i.e. the gap between simulation and real-world traffic condition), and the objective value needs to be evaluated through DTA outputs under the DUE constraint in the lower level. In addition, large networks make the optimization more challenging because it can take from hours to days for large networks to converge to DUE.

DTA calibration began to attract researchers' attention from the late 1980s, when a generalized least-squares model was applied to calibrate the OD matrix based on traffic counts (Cascetta and Nguyen 1988; Cascetta 1993). The authors estimated the linear relationship between the time-dependent traffic counts and the OD flows. This linear relationship could also help the closed formulation of the objective function in the DTA OD calibration. Some researchers also attempted to manually adjust the model parameters, which could be time consuming for large networks (Chu et al. 2004). There was a trend that researchers calibrated supply and demand parameters separately, with iterative steps assuming the remaining parameters stayed constant (Doan et al.

1998; Hawas 2002). Balakrishna et al. (2005) proposed a systematic calibration method that treats different parameters independently and identified supply parameters and demand parameters simultaneously. Using the simultaneous perturbation stochastic approximation (SPSA) method, they calibrated a mid-size mesoscopic DTA model with 606 links in Los Angeles (Balakrishna et al. 2006; Balakrishna et al. 2007). Vaze et al. (2009) applied SPSA to calibrate over 6,000 demand and supply parameters, and they showed the superiority of SPSA over genetic algorithms (GA). In these studies, SPSA was proven a fast converging method, but the limitation was that SPSA is a local optimization method that might not guarantee global optima (He 2014). Compared with calibrating demand parameters and supply parameters separately, the joint calibration ignored the presence of interaction among the demand and supply parameters (Omrani and Kattan 2012).

One common issue with DTA calibration is that the number of parameters to calibrate is far more than that in traditional optimization problems. Therefore, the calibration of medium-scale or large-scale models is usually time consuming. Another issue is that the calibrated model may only refer to a "typical" scenario (e.g. typical weekday, typical weekend), which impedes its further applications in day-to-day or even real-time applications.

Section 2.3. Online calibration of transportation models

With the fast development of computational technologies and traffic monitoring systems, online calibration of transportation models has drawn increasingly

attentions. Online calibrations are necessary for real-time traffic modeling systems in terms of current traffic condition estimations and near-future predictions. Unlike optimization-oriented offline calibrations, online calibrations usually utilize some adaptive approaches to adjust model parameters.

There have been attempts to make online traffic condition estimations and predictions. van Arem and van der Vlist (1992) utilized real-time measurements of traffic flow, occupancy and speed to estimate the online capacity of a motorway cross-section. They first estimated the fundamental diagram via a maximum occupancy that may be achieved under free-flow conditions; then calculated the capacity using this fundamental diagram and the notion of "maximum" occupancy. Doan et al. (1998) proposed a reactive traffic propagation adjustment module to periodically adjust a traffic management simulation model, such that the model performed consistently with real world network. The module was formulated as a PID (proportional, integral and derivative) controller to adjust the inconsistencies between simulated and measured densities. He et al. (1999) listed a number of major sources of the inconsistences between DTA model outputs and real-world observations, i.e. link travel time functions, route choices, and flow propagation models. They attempted to calibrate the outputs of a modified Greenshields' model to make the analytically computed travel times match the detected travel times. An iterative approach that sequentially considers the three components until convergence was proposed.

There are some general approaches to estimate real-time traffic conditions under DTA frameworks. Tavana and Mahmassani (2000) proposed transfer function

methods (bivariate time series) to estimate speed-density relations from typical detector data. Time series analysis was used as a basis to estimate speed based on densities. Huynh et al. (2002) extended the work by incorporating the transfer function model into a DTA simulation-based framework. The parameters were updated periodically based on prevailing traffic conditions and nonlinear least squares optimization, making it an adaptive process. A similar approach was completed by Qin and Mahmassani (2004), who utilized equilibrium/actual speed as the input/output of the transfer function, respectively. The objective of building a transfer function is to estimate and predict traffic measurements that both considered historical traffic dynamics and real-time detections. The models demonstrated their accuracy compared with static macroscopic models. Wang and Papageorgiou (2005) conducted real-time traffic estimation by formulating a stochastic macroscopic traffic flow model as a state-space model. The model considers traffic model's parameters as time-varying state variables. All these works are about supply side online calibration, i.e. estimating real-time free-flow speed, critical density, and capacity. State-space models were also used for online OD matrices estimation and prediction problems (Ashok and Ben-Akiva 1993; Ashok 1996; Ashok and Ben-Akiva 2000). A Kalman Filtering algorithm was used to solve these state-space models, making it computational effective for real-time applications. The model was extended to calibrate both OD matrices and supply parameters by incorporating a nonlinear Kalman Filtering Algorithm (Antoniou et al. 2007). Hashemi et al. (2017) utilized Q-learning to formulate consistency checking and online calibration procedure. The procedure will estimate whether or not to adjust the DTA

parameters under the reinforcement learning framework. If the procedure determines to calibrate the model, least square method will be used to estimate the new OD tables. However, these research did not consider much about historical data.

Cantelmo et al. (2014) used modified SPSA to estimate dynamic OD under a bi-level DTA framework. Cantelmo et al. (2017) also proposed a bi-level dynamic OD estimation model for DTA demand calibration. The model embedded utility optimization into the formulation of the optimization problem. However, due to the computational speed issue, only a small network was illustrated in the paper. Zhou and Mahmassani (2005) developed an online OD matrices estimation approach that minimizes the deviation between simulated information and real-world information as well as demand adjustment magnitude. In their paper, only real-time deviation was considered, such that historical data was not adequately considered to save computational time. Zhou and Mahmassani (2007) enhanced their method by using a polynomial trend filter to estimate deviation from a prior estimated demand. Day-to-day updating was also incorporated. A possible issue is that it can hardly consider the differences in demand patterns on weekdays and weekends.

Currently, the focus of online calibration of DTA models is still on how to estimate the current traffic condition via fast linear or non-linear state models. The advantage of these methods is fast computation, but there is still room to improve the calibration accuracy as well as efficiency, especially for large-scale networks. In addition, historical data and "typical" (or prior) models may play more important roles in online calibration approaches.

Section 2.4. Multi-agent technologies and models

Today's computational and information technologies make it possible to utilize large-scale DTA models and advanced travel demand models for transportation system analysis. The transportation modeling platforms provide a testbed for many ATM strategies and transportation oriented policies under ITSs. As a complex system, an ITS is suited for agent-oriented technologies. The concept "agent-based" was first proposed in the 1970s during the dynamic modeling of segregation process (Schelling 1971), in which autonomous agents interacted with each other in a shared environment. In a MAS, agents learn and accumulate knowledge under the environment and make decisions to pursue their own goals. In ITSs, the environment can be the transportation system, and agents can vary from travelers to ATM facilities. In the academic field of transportation, MAS have been applied to many aspects of traffic and transportation systems.

Agent technology enhances interoperability and distributed computing capability of existing centralized information systems in ITSs. The distributed agent systems can combine information from multiple detection stations and systems, evaluate traffic flow, respond to traffic flow changes, and evaluate operational responses to traffic flow changes in real-time (Chen and Cheng 2010). In the field of ATM, Roozemond (2001) applied a MAS for urban real time traffic signal control. The MAS consists of signal agents, segment agents and authority agents, such that they make reactions to the change of traffic condition by coordinating with each other to reach the global optimal performance. Similar agent-based methods were proposed to achieve

urban signal adaptive coordination and control (Kosonen 2003; France and Ghorbani 2003; Yang et al. 2005; Grégoire et al. 2007). Logi and Ritchie (2002) proposed freeway agents and arterial agents to help manage traffic facilities (e.g. signals, ramp metering, dynamic massage signs) to mitigate local congestion. Some researchers proposed routing agents to help with real time road guidance or assistance to avoid traffic congestion (Shi et al. 2005; Weyns et al. 2007; Rothkrantz 2009). There is also integration between traffic facility control and road guidance. Li and Shi (2003) proposed a MAS that integrated signal control with real time road guidance.

The research motioned above usually regards the infrastructures of transportation systems as intelligent agents. Another trend is to model travelers as agents. Rossetti et al. (2000) model drivers as learning agents to extend a microscopic traffic simulation model called dynamic route assignment combining user learning and microsimulation. Wahle et al. (2002) utilized a tactical layer and a strategic layer to model drivers' reactions under dynamic traffic information. Panwai and Dia (2007) applied neural agents to model car-following behavior. Doniec et al. (2008) proposed a multi-agent behavior model to coordinate drivers' actions at intersections. Zhang (2007) developed a searching, information, learning, and knowledge (SILK) theory to model travelers' long term behavior on route choice. Xiong and Zhang (2013) extended this SILK for departure time choice. Zhu et al. (2015) applied this departure time agent-based model to analyze the impact of flextime schedule policies and cumulative land development (Zhu 2014).

Section 2.5. Summary

This chapter provides a comprehensive literature review on transportation modeling, offline line and online calibration of transportation models, and agent-based systems in transportation engineering. Firstly, it is found that a comprehensive transportation system modeling approach is difficult to build, especially when incorporating individual travel decision making process with network modeling. The ongoing activity-based models usually take days or weeks to model a typical day's demand pattern and the corresponding traffic performance. Secondly, even though there are successful large-scale transportation models, it is still unrealistic to apply these models for real-time applications. The computational speed of online calibration methods is the major bottleneck that prevents the models from real-time applications.

With these two concerns, this dissertation aims at developing an agent-based transportation modeling system that is capable for online applications. Compared with machine learning or data driven methods, transportation network models may not be the most accurate in predicting near future traffic conditions. However, an online transportation model system allows agency planners to evaluate different traffic management strategies, which is infeasible for machine learning or data driven prediction methods. One major contribution of this dissertation is to develop a fast enough online system model calibration approach. Based on the online calibration approach and an agent-based system model, one can achieve real-time simulation for even large-scale transportation networks for traffic management decision supports. The practical value of this tool can match its theoretical value because the implementation

of the proposed tool is straightforward. And there is still room for improvements once

more advanced simulation or computation technologies are ready.

Chapter 3: Multi-Agent Transportation Simulation System

The major approach of this chapter is to explain the model components, and procedures required to build the multi-agent transportation modeling system. I will introduce an open source DTA model and an agent-based behavior model, then integrate them as a system model to overcome the limitations of the traditional travel demand modeling. The integrated model can be viewed as a convenient transportation modeling tool for agencies and planners to investigate future demand forecasting issues such as peak spreading, route shifts under recurrent congestion, etc. It can also be used for real-time decision support once the online calibration issue has been addressed.

Section 3.1. Regional planning model

In the Northern Washington, D.C. area, there is a regional planning and traffic demand model called the MWCOG model. The model is developed to forecast travel demand and its impacts from the base year (2010) to decades in the future (2030). It is able to analyze land activities on the TAZ level, and assess transportation network development plans and transportation policy assumptions to estimate the travel demand for different modeling years and scenarios. The TAZ level land activity is forecasted through the cooperative forecasting program called "Round". Round estimates TAZ level activities based on the econometric projections of employment, population and household information from local governments. Local governments offer their land development details to "Round" to provide these social demographic data. The transportation network data, which cover freeways, most major/minor arterials, some

connectors, and transit-only links (e.g., rail), are developed by the COG staff. The transportation policy assumptions include the change in transit fare/auto operating costs over years, the change of parking cost and the amount of through traffic from other regions. As a trip-based, four-step travel demand model, the MWCOG model produces several outputs at both the aggregate level and disaggregate level. For aggregate level outputs, the trip generation module estimates the trip production and attraction for each TAZ. Disaggregate level outputs include TAZ-to-TAZ trips (OD matrices) for different travel modes from the trip distribution module and mode choice module, auto trips/volumes on road links from the static traffic assignment module, and transit person volumes on transit links from the transit assignment module. The proposed system model takes the network and the demand outputs of MWCOG to develop the DTA model. Therefore, the TAZ level travel demand is still modeled by MWCOG, but the travel behavior and traffic dynamics can be modeled by using the proposed integrated system model.

Section 3.2. Mesoscopic DTA model

Many DTA integrated traffic simulators, such as DYNASMART, TRANSIMS and DTALite (Zhou and Lu 2010) have been used in previous studies. They are all equipped with good features for real-world applications, and there is no consensus on the superiority of any single simulator. DTALite is selected in the system model for four reasons: 1) it is able to generate a simulation network directly from the network file of a regional planning model or a geographical information system (GIS)

database; 2) it is a mesoscopic simulator with parallel computing that increases the speed of the simulation-based DTA; 3) the embedded OD calibration module and the subarea cut module allow a detailed analysis for zooming into specific subareas; and 4) it is well-supported for agent-based modeling.

Agent-based modeling means the TAZ-to-TAZ demand input (OD matrices) for DTALite is not necessary. One can also provide an input agent list file for the traffic assignment. The agent list contains every traveler's detailed travel information, including origin, destination, departure time, and VOT. After producing traffic assignment, DTALite will also generate an output agent list that provides each traveler's trip information (e.g., origin, destination, departure time, arrival time, and trip trajectory). Therefore, the integration between DTALite and agent-based behavior models can be implemented via these agent list files. The output agent list from DTALite can be used as travel experience to update travelers' knowledge, then the behavior model is employed to forecast behavior changes; based on the forecasted behavior changes, one can modify the input agent list to conduct traffic assignment again to investigate the traffic impact.

Supported by Maryland State Highway Administration (SHA), a mesoscopic DTALite model has been developed for the entire Montgomery County, MD, including all freeways, most major/minor arterials, and some local connectors/streets. The major commuting corridors—I-270, North I-495 and MD 355—are located in the middle of this study area, as shown in Figure 3-1. This large-scale network is taken as an example to illustrate how to integrate the proposed agent-based travel behavior model (AgBM)

with DTA. The simulation network, which contains 470 TAZs, 5,481 links and 1,921 nodes, is directly obtained from the MWCOG traffic demand model via the DTALite's subarea cut procedure. I transform MWCOG's static OD to time-dependent OD as seed OD matrices for calibration. Section 3.4 will illustrate how the subarea cut procedure works, and how the time-dependent seed OD is obtained.
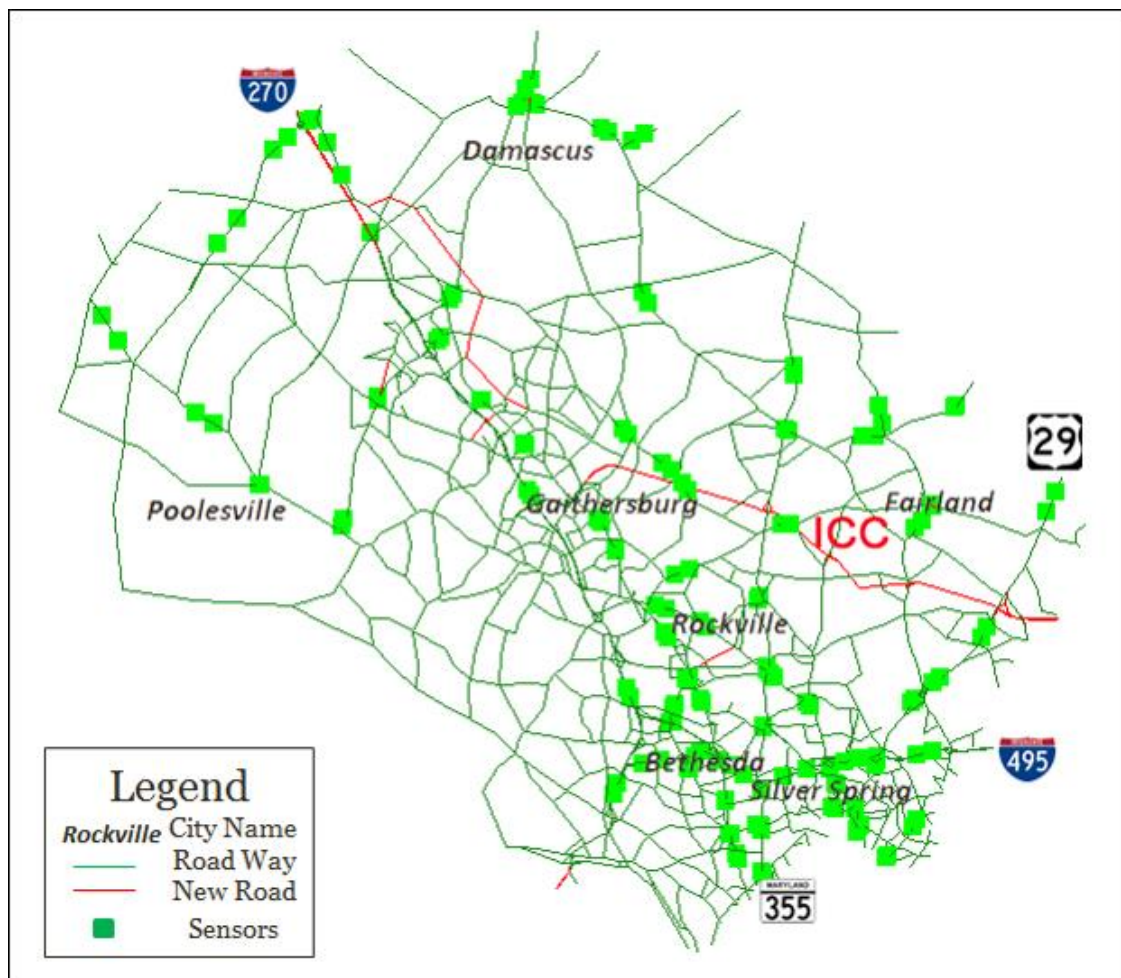
Figure 3- 1 Mesoscopic simulation network with traffic count stations

The base year model (2010, typical weekday morning peak period from 5:00 AM to 10:00 AM) is calibrated with hourly volume data measured by 160 sensors from the Maryland SHA's traffic monitoring system (denoted by green points in Figure 3-1).

The calibration is conducted via the "OD estimation" procedure (called ODME) in DTALite, which adjusts the time-dependent seed OD matrices based on the volume differences of the 160 sensors. That is, the "OD estimation" procedure first conducts DTA with the seed OD to obtain DUE, and then begins to adjust the seed OD based on the route choice under DUE. If the assigned flow on a link with sensor is higher than the real-world counts, the procedure will decrease the demand of the corresponding OD pairs; otherwise, it will increase the demand. After comparing all the sensors, the adjusted OD matrices are set to be new seed OD matrices for the next iteration of calibration. The calibration process will terminate when there is no improvement in terms of the matching of the sensor counts. Readers may refer to Lu et al. (2013) for more detailed information for this OD calibration procedure. More details about ODME will be discussed in Chapter 4. Since the OD estimation procedure only makes OD adjustments, we also need to calibrate the supply parameters (e.g., link capacity, jam density, speed limit, and congestion wave propagation speed) by trial and error. After the calibration, the weighted mean squared error (WMSE) is 0.162. The WMSE, which is a measurement usually used for DTA calibration (Xiong et al. 2015a), is defined as:

$$WMSE = \sqrt{\frac{\sum_i \sum_h \left(x_{i,h} - x_{i,h}^*\right)^2}{\sum_i \sum_h \left(x_{i,h}\right)^2}} \tag{3-1}$$

where $x_{i,h}$ is field detected counts of station i during hour h, $x_{i,h}^*$ is the simulated counts.
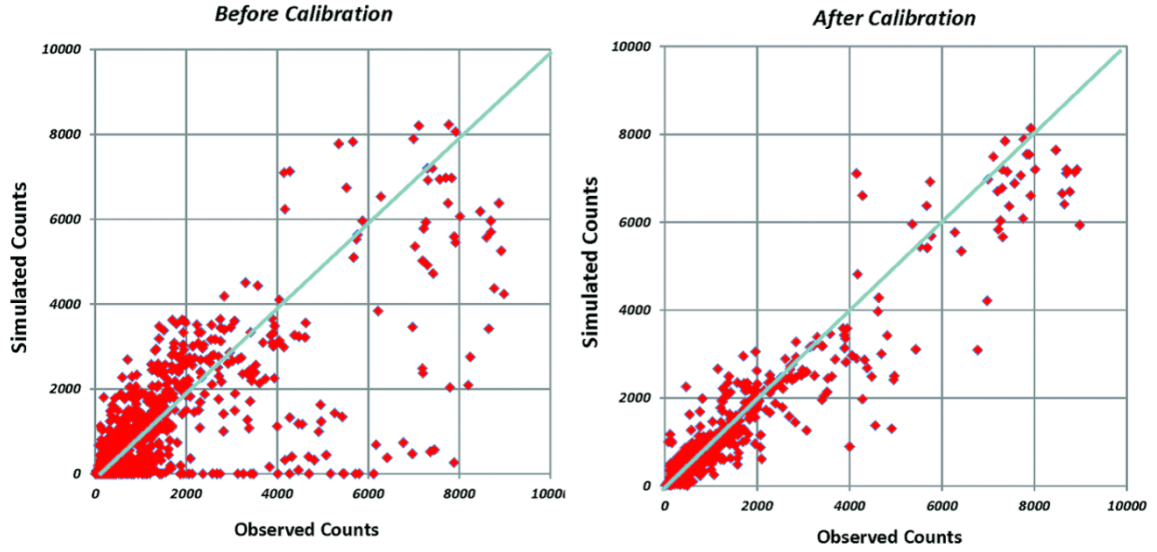
Figure 3- 2 Calibration results of the simulation-based DTA.

Except for WMSE, I also check the GEH statistic which is used to compare two sets of traffic volumes (Chitturi et al. 2014).

$$G_{i,h} = \sqrt{\frac{2\left(x_{i,h} - x_{i,h}^{*}\right)^2}{x_{i,h} - \overline{\overline{x_{i,h}}}}} \qquad (3\text{-}2)$$

where $G_{i,h}$ is the GEH statistic of station i during hour h, and $\overline{x_{i,h}}$ denotes the average field detected counts across all the sensors and time intervals. The average value of all the 160 sensors across 5 hours' time intervals (800 $G_{i,h}$ in total) is 3.18; and 703 (87.9%) of the GEH statistics are under 5, which indicates a good match. In the calibrated model, there are 426,958 total trips, including a warm-up period from 4:00 AM to 5:00 AM and a clear-up period from 10:00 AM to 11:00 AM.

Section 3.3. Agent-based travel demand model

The positive agent-based travel behavior model is integrated with DTALite to incorporate travelers' cognition in the mesoscopic DTA model. In this dissertation,

I try to model the departure time dimension in one's travel decision making process, other dimensions e.g. mode choice, destination choice can also be included. The framework of the positive departure time model was developed by Zhang and Xiong (2012). The model attempts to estimate how people actually behave without the assumption of perfect rationality. In the positive departure time model, each traveler is described as a learning agent. Each agent in the model has an initial subjective belief on his/her trip, which is the ideal travel condition with free-flow travel time (FFTT) and zero schedule early arrival/delay. While it is almost impossible to reach the ideal condition in reality, the positive model forms a perception and learning process to model agents' learning and behavior adjustments toward their actual experience and beliefs, see Figure 3-3. Agents are able to acquire and learn traffic information from their travel experience (or other sources, such as traveler information systems) and then update their knowledge. Both the subjective belief and the learnt knowledge are quantified to determine personal attitudes towards his/her current experienced traffic conditions:

$$P(\tau) = \phi_{TT} \cdot TT(\tau) + \phi_{ASDE} \cdot ASDE(\tau) + \phi_{ASDL} \cdot ASDL(\tau) \qquad (3\text{-}3)$$

$$ASDE(\tau) = \max\left(0, PAT - \tau - TT(\tau)\right) \qquad (3\text{-}4)$$

$$ASDL(\tau) = \max\left(0, \tau + TT(\tau) - PAT\right) \qquad (3\text{-}5)$$

where τ is the departure time, $P(\tau)$ is the quantified attitude (also called payoff) towards the current traffic condition, $TT(\tau)$ is the travel time, $ASDE(\tau)$ and $ASDL(\tau)$ are the schedule early arrival/delay, respectively, and PAT denotes preferred arrival time, ϕ_{TT}, ϕ_{ASDE}, and ϕ_{ASDL} are the coefficients in payoff function. The gap

between the experienced best situation and the ideal situation is theorized as subjective search gain, given by:

$$g_T = \left(p^s - p_{weight,T} \right) / \left(T + 1 \right)$$ (3-6)

where g_T is the search gain on day T, p^s is the ideal payoff under the subjective belief (payoff under FFTT with zero schedule early/delay), $p_{weight,T}$ is the payoff under the best traffic condition the agent has ever met up to day T. This subjective search gain measures how much the agent expects to benefit if he/she shifts departure time. Here, the model uses $(T+1)$ as the denominator, because agents' expectation of finding a more desirable departure time will decrease as they keep searching and continue making decisions in the system. Since individuals can select from all tried departure times and pick the one with the highest weight (impression), the weight of day t's travel experience is calculated as follows:

$$w_t = \eta_t \cdot \zeta_t$$ (3-7a)

$$\eta_t = (1 - \frac{p^* - p_t}{p^*})^{\theta_1}$$ (3-7b)

$$\zeta_t = \left(\frac{1}{T - t + 1} \right)^{\theta_2}$$ (3-7c)

$$p_{\text{weight},T} = p_m, 1 \leq m \leq T, w_m = \max \left(w_1, w_2, ..., w_T \right)$$ (3-7d)

where w_t denotes the weight for day t among T days that has passed. The weight is determined by a representativeness weight η_t and a recentness weight ζ_t; p_t is the payoff on day t; θ_1 and θ_2 are factors for representativeness and recentness. Thus, $p_{weight,T}$ is the payoff of the day with the highest weight (3-7c).

This captures agents' satisfying behavior, meaning that agents will become satisfied and stop the search for new, alternative departure times. Correspondingly, since agents may take effort to search, e.g., time, monetary, mental efforts, and risk, the

search cost is proposed to measure their perceived loss in each round of search, given by

$$c_{LOW} = g_T = \left(p^s - p_{\max,T}\right)/(T+1) \tag{3-8a}$$

$$c_{HIGH} = g_{T-1} = \left(p^s - p_{\max,T-1}\right)/T \tag{3-8b}$$

$$c = \left(c_{LOW} + c_{HIGH}\right)/2 \tag{3-8c}$$

where c_{LOW} and c_{HIGH} are the lower and upper boundaries of the search cost of an agent, respectively.

To investigate these two boundaries, a local survey was conducted, where every respondent was asked to search repeatedly until he/she was satisfied with the most recent departure time (Zhang and Xiong, 2012). If T is the total time of this searching process for the respondent, then it is reasonable to assume the search cost is between the gain of the last search (c_{LOW}) and the second-to-last search (c_{HIGH}). Thus, I assume the cost to be the average of c_{LOW} and c_{HIGH}. The start and termination of an agent's searching process are determined by the trade-off between the subjective search gain and the perceived search cost.

In the aforementioned explanations, I have illustrated how travel experience is converted to a quantified attitude by using subjective belief, search gain and search cost. Assume that one has a traffic simulation model or travel information system so that every agent's travel conditions, such as origin, destination, shortest travel time, departure time and actual arrival time on day t, are known. Then, one is able to estimate the departure time choice on day $(t + 1)$ for all the agents in the network via the flowchart in Figure 3-3.
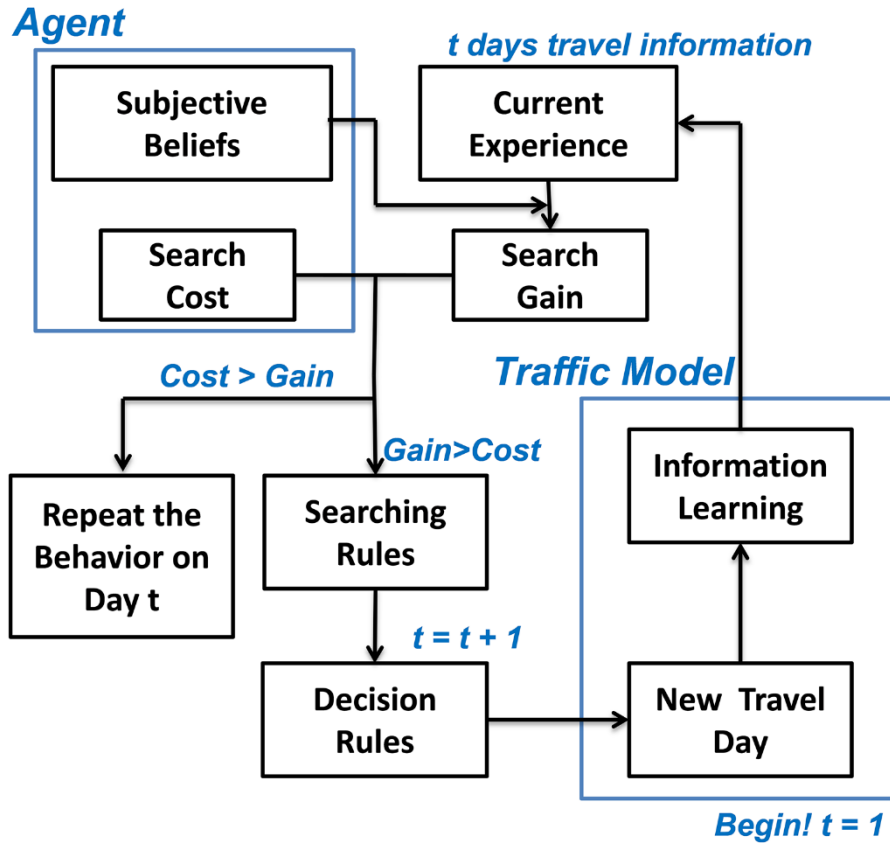
Figure 3- 3 Flowchart of the positive departure time model for one agent

After generating the agents' subjective beliefs and search cost based on their travel information, the departure time decision-making process begins with updating the agents' travel experience from the output of the simulation model on day t. For each agent, the positive model calculates the search gain from his/her updated travel condition and his/her subjective beliefs. If search gain is greater than search cost, the searching begins; the agent then follows a series of searching rules and decision rules. The searching rules guide the identification of an alternative departure time for each agent. The decision rules govern the switching behavior between the new alternative and the original departure time. Once the search gain is smaller than search cost, the agent will stop searching, and he/she would repeat his/her current departure time for

the rest of the simulation. After all the agents have made decisions on day t, the individual-level behaviors are aggregated as the demand for day $(t + 1)$ to the traffic simulation model. The simulation output for day $(t + 1)$ is updated again for agents to estimate the departure time decisions for day $(t + 2)$. This loop process will continue until all agents stop searching and changing departure times, which is referred as behavior user equilibrium (BUE) (Zhang 2007).

The parameters of the travel time and schedule early arrival/delay in the search gain function are taken from the departure time study done by Small (1982). The search rules and decision rules have been estimated using decision tree algorithms and were cross-validated with stated preference survey data (Zhang and Xiong, 2012). Given the framework of the proposed agent-based behavior model, one can conduct local travel behavior surveys and use decision trees or other technologies to estimate search/decision rules or reformulate the payoff functions.

Section 3.4. Model integration

This section presents the integration of the aforementioned model components (regional planning model, mesoscopic DTA model, and positive agent-based behavior model) into a system network modeling tool for transportation planning and management applications. Unlike traditional trip-based transportation models, this integrated model treats every traveler in the network as an intelligent agent who can update knowledge and make decisions. Agents follow the positive model and make travel decisions based on their previous travel experience and subjective beliefs. After

modeling travelers' behavior change, the individual decisions are aggregated as a new

demand pattern for traffic impact modeling. The integration flowchart is shown in

Figure 3-4.
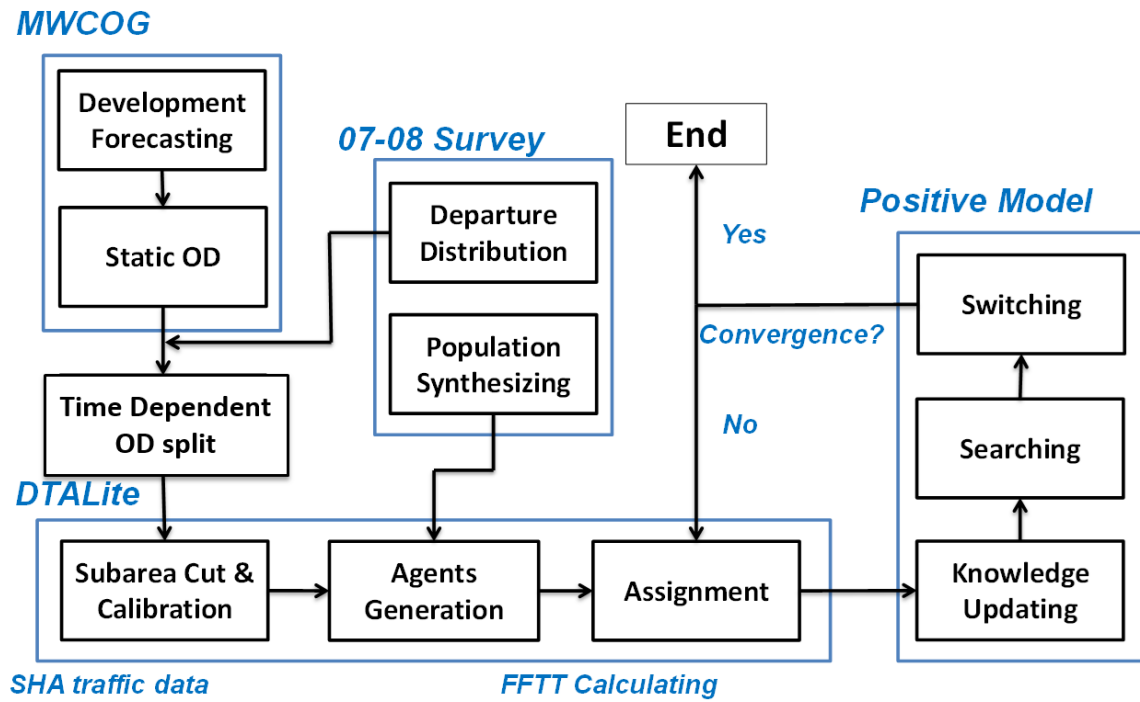
Figure 3- 4 Flowchart of the integrated system model

The integration of the proposed MAS simulation tool in this dissertation starts

with the MWCOG regional planning model. The "Round" model first takes land

development inputs and estimates TAZ based travel demand, then, using the four-step

travel demand model in MWCOG, obtains static OD matrices. The static OD matrices

are converted to time-dependent OD matrices, based on the departure time distribution

in the 2007-2008 Transportation Planning Board (TPB)/Baltimore Metropolitan

Council (BMC) Household Travel Survey. Let S_h denote the number of auto trips

departing during hour h, summarized from the TPB/BMC survey, OD_k denotes the

static demand for OD pair *k* from MWCOG, then the time-dependent OD is:

$$OD_{k,h}=OD_k \cdot \frac{S_h}{\sum_{h'=1}^{H} S_{h'}}$$

(3-9)

where $OD_{k,h}$ is the trips for OD pair *k* during hour *h*, *H* is the total traffic assignment horizon.

The open source mesoscopic DTA package DTALite involves a network converting procedure that coverts and imports the MWCOG transportation planning network to a simulation-based DTA network. With the MWCOG DTA network and the time-dependent OD matrices, I conduct DTA to assign the regional auto demand onto the MWCOG network. Since the study area is a subarea within the MWCOG network, I derive the subarea network and its time-dependent OD matrices via the subarea cut procedure in DTALite. The subarea cut procedure defines the boundary of the study area and summarizes its internal and through demand, based on traffic assignment results of the entire MWCOG network. The cut DTA model is then calibrated via the traffic counts data mentioned in the previous section. Until now, one has already developed a mesoscopic DTA model from the regional planning model.

To integrate such a traditional DTA model with a positive agent-based behavior model, one needs to take advantage of the output agent list mentioned in the mesoscopic DTA model subsection. The travelers in the output agent list of this subarea DTA model are considered all the agents to model. Therefore, one knows the origin, destination, departure time, arrival time and other detailed travel information for these agents. For each agent, his/her payoff under the ideal travel condition can be found by

capturing the FFTT of the shortest path connecting his/her OD pair, assuming the schedule early arrival/delay to be zero. Based on the social demographic distributions in 2007-2008 TPB/BMC survey, each agent in the output list is assigned with demographic information such as income, and gender. The social demographic attributes are necessary because they are included as explanatory variables for searching and decision rules in the positive behavior model (Zhang and Xiong 2012). This process can be regarded as the initialization of the AgBM-DTA model.

Under the initialization, DTA is conducted to simulate agents' travel experience for the daily traffic knowledge learning process. One can take the output agent list from DTALite to update each agent's travel experience, and follow the flowchart in Figure 3-3 to model the knowledge learning, searching and decision-making process. After predicting each agent's departure time change, one can update the input agent list with the new departure times. DTA is conducted again to simulate agents' travel experience for a new day. For long-term transportation planning impact studies, the iterative loops of departure time prediction and DTA will not finish until only a small portion of individuals (5% in this dissertation) are still searching for alternative departure times; or, for day-to-day real-time transportation simulations, there is no need to stop the iterative loops. The AgBM is not only targeted on day-to-day behavior shifts, but also allows agents to make en-route behavior adjustments, which will be fulfilled in future research.

Section 3.5. Model calibration via surrogate-based optimization (SBO)

The positive departure time model aims at estimating how travelers adjust departure times via day-to-day learning and knowledge updating process. The integrated model requests calibration so that the BUE can reflect travelers' actual choices in reality. It is assumed that the framework of this positive departure time choice model correctly describes how people change their departure time as they obtain more knowledge. It is also assumed that travelers in a real-world network have no knowledge at the beginning. Then it is reasonable to deduce that the process of approaching BUE correctly describes how people switch to today's departure time from an initial zero-knowledge departure time. The initial zero-knowledge departure time is assumed to be the preferred arrival time minus FFTT, which is estimated as the arrival time minus FFTT in the integrated flowchart (Figure 3-4). In this way, the goal of the calibration is an optimization solving process to find the best decision variables so that the departure profile of the BUE condition matches the real world data. Surrogate-based optimization (SBO) is applied to achieve the calibration work (Chen et al. 2015).

SBO is an optimization method that mainly deals with optimization problems that do not have an analytical solution and request simulation to obtain the value of objective function. In these optimization problems, the relationship between decision variables and objective function is not derivable directly through mathematical formulations. Once a vector of decision variables $X = (x_1, x_2, ..., x_l)$ is given, simulation approach is usually required to observe the performance of X on the objective variable Y. In SBO, a list of initial decision variables $(X_1, X_2, ..., X_N)$ will

be tested to approximate a list of objective variables ($Y_1, Y_2, ..., Y_N$). These N initial sample points will be utilized via the SBO algorithm to regress the black-box relationship between X and Y to some parameterized basis functions. Once this relation is built, a new vector of decision variables X_{N+1}, which may have increased potential to reach better objective function values, will be generated. With the existing N initial points and one (or more) infill point, the relationship will be regressed once again to generate infill vector X_{N+2} and more infill vectors. The termination criterion of this infill process is often set as a maximum infill number (*M*) of evaluations of objective function.

In this research, the reason to select SBO is that the optimal solution cannot be obtained in analytical ways; one must rely on the integrated model framework to obtain the travel departure profile under BUE. Once the real-world traffic network and an initial demand have been obtained, I will utilize the integrated model to simulate every traveler's travel experience and estimate their departure time switching process until reaching a BUE. The outputs from DTALite can provide the detailed travel data to calculate the objective function.

Since the search cost, search gain, and decision rules have been validated in previous work (Zhang and Xiong 2012), this SBO calibration mainly considers representativeness factor (θ_1) and recentness factor (θ_2) mentioned in Section 3.3. There are five parameters to calibrate in this section:

1) representativeness factor (commuting trips), ranges from 0 to 1

2) recentness factor (commuting trips), ranges from 0 to 1

3) representativeness factor (non- commuting trips), ranges from 0 to 1

4) recentness factor (non- commuting trips), ranges from 0 to 1

5) search Scale factor (commuting trips), ranges from 0.6 to 1

The first four parameters are the factors in experience weight function in equation (3-7) for both commuting trips and non-commuting trips; the last parameter is a scale factor for commuting trips, which means the commuting travelers will have smaller switches in search rules. Details about search rules can be found in Zhang and Xiong (2012). The objective function is shown below:

$$Y = \sqrt{\frac{\sum_{h=1}^{H}(d_h - d_h^{BUE})^2}{\sum_{h=1}^{H}d_h^2}} + \sqrt{\frac{\sum_{h=1}^{H}(cd_h - cd_h^{BUE})^2}{\sum_{h=1}^{H}cd_h^2}} \qquad (4\text{-}10)$$

where the simulation horizon is divided to H time periods; d_h is the demand percentage of period h over the whole horizon in real-world data; cd_i is the cumulative demand percentage from period 1 to period h in real world data; d_h^{BUE} and cd_h^{BUE} are the corresponding percentages of the BUE condition from the integrated model.

Another point of calibrating these five parameters is for the transferability of this TIA tool. Since the searching rules and decision rules were obtained and validated based on a local survey in Zhang and Xiong (2012), repeating such a survey in other study areas may be difficult; it is easier for agencies to obtain the departure profiles of their own study areas. Therefore, different agencies and experts can calibrate and validate the behavior model for their own studies without re-estimating the searching and decision rules by adjusting these five parameters.

The flowchart of the SBO calibration is shown in Figure 3-5. First, *N* initial

decision vectors will be generated by using the Latin hypercube sampling method (Chen et al., 2015) to fill the five-dimensional range space of the decision variables. These initial vectors of behavior parameters will be loaded into the integrated AgBM-DTA model to approximate the objective functions one-by-one. Then I apply the surrogate model improved by Chen et al. (2015) to regress the relation between X and Y. Based on the estimated relation, a new infill point with a large potential to reach better objective value will be generated for the integrated TIA tool to approximate its performance. After re-estimating the relationship with the infill point, the system will generate one more infill point and repeat the infill process until the number of total infill points reaches M. After the termination of this SBO process, the system will pick the point with the lowest objective value to be the optimal solution. The capability of SBO to obtain the global optimal decision variables has been explained in this article.
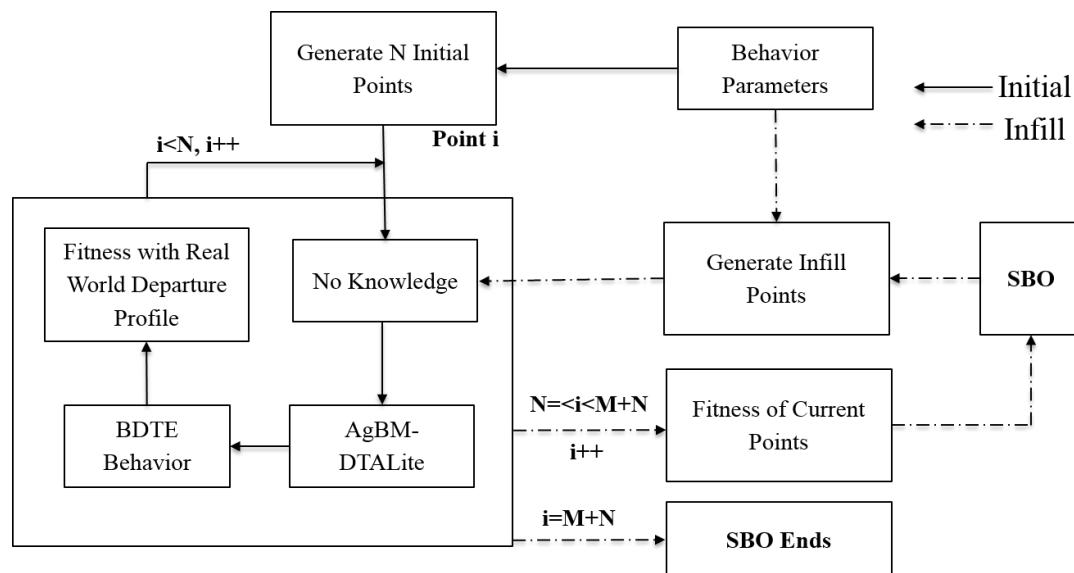
Figure 3- 5 Flowchart of SBO integrated model calibration

The integrated model (Montgomery County network) with the transportation network presented in Section 3.2 is calibrated and validated following the framework

above. The calibration horizon is from 5:00 AM to 10:00 AM. The real-world departure profile is obtained from the 2007-2008 TPB/BMC survey, as well as the ratio of commuting and non-commuting travelers. I assign the demand calibrated (agent list) in Section 3.2 randomly to be commuters or non-commuters based on this ratio. I then generate 80 initial points and 40 infill points to complete the calibration. For each point, it takes around 20 simulation days (AgBM iterations) to reach BUE. Before calibration, when the 5 parameters are all set to 1, the value of objective function is 7.559. After 120 iterations of SBO, the objective value drops down to 3.211 with the optimal decision vector [0.9925, 0.7724, 0.0247, 0.6123, 0.9953]. This calibration result indicates that commuting travelers pay more attention to representativeness; and non-commuting travelers care more on recentness. The search scale factor is closed to 1, indicating that the search rules are well-validated in previous research (Zhang and Xiong 2012). The comparisons of the departure profiles are shown in Figure 3-6 (a) and Figure 3-6 (b).
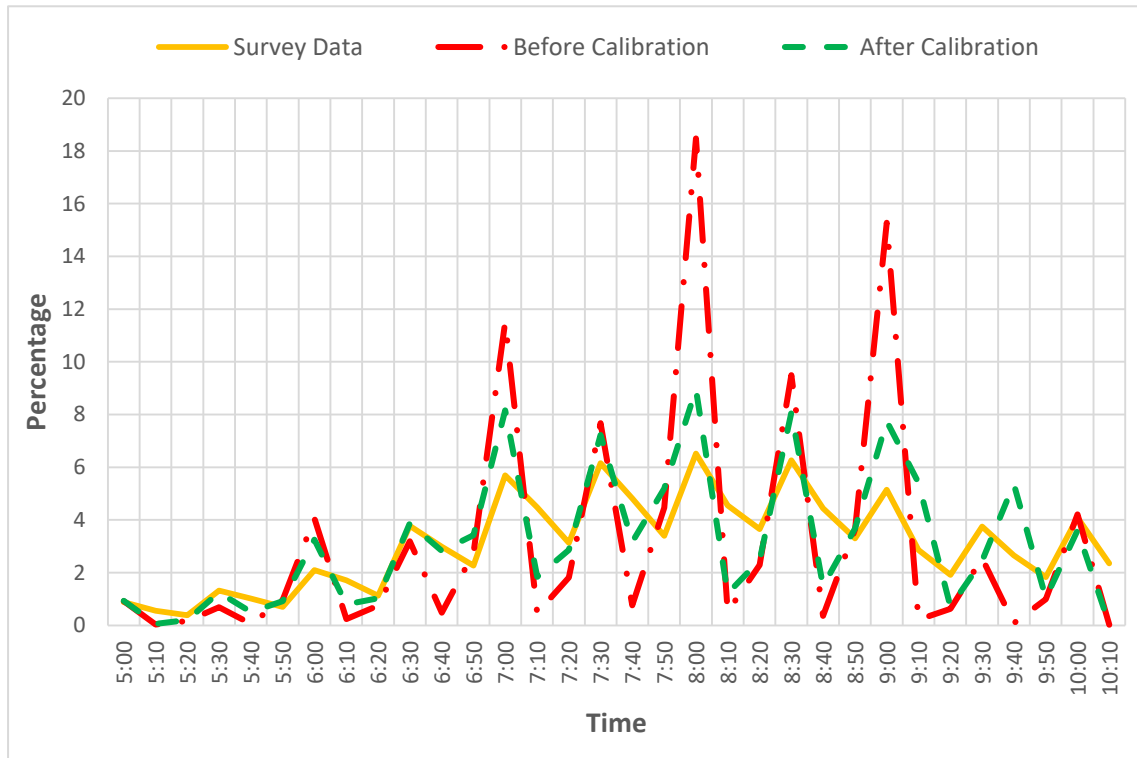
(a) Timely demand percentage profile
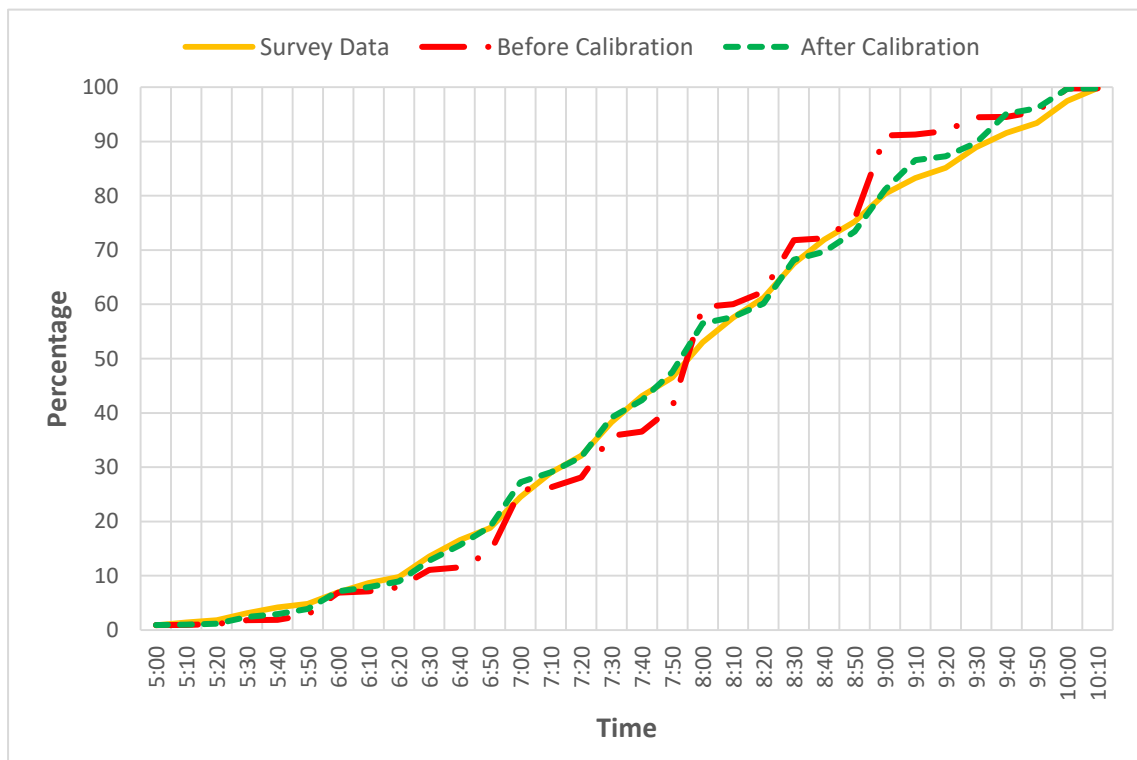
(b) Cumulative demand percentage profile

Figure 3- 6 AgBM calibration results

The calibrated model has been used to analyze the traffic impact due to cumulative land developments in my master thesis (Zhu 2014). The details of this traffic impact study will not be presented in this dissertation.

Section 3.6. Large-scale network application

The ultimate goal of this dissertation is to develop an online agent-based simulation tool for large-scale networks. Supported by Advanced Research Projects Agency-Energy (ARPA-E) of the department of energy, a large-scale simulation model has been developed. Figure 3-5 displays the entire system model network, which covers the whole of Washington, D.C.; Montgomery, Prince George's and Frederick Counties in Maryland, as well as parts of Baltimore County; and Arlington and Fairfax Counties in Virginia. All the geographic information and traffic infrastructure information of this network comes from the Maryland Statewide Transportation Model (MSTM), which is another planning four-step model similar with MWCOG. The large-scale DTA network contains 1,228 traffic analysis zones, 16,563 nodes, and 42,240 links. All the interstate freeways, highways, most of the major and minor arterials, and some of the connectors and local roadways are included in this network. The DTA model is also coded in DTALite due to its built-in parallel computing capability dramatically speeds up the traffic assignment and OD estimation process when using multi-core CPU hardware. At the current stage, the University of Maryland (UMD) research team has been working on the calibration for the morning peak, mid-day, and afternoon peak period (6:00 AM to 7:00 PM).
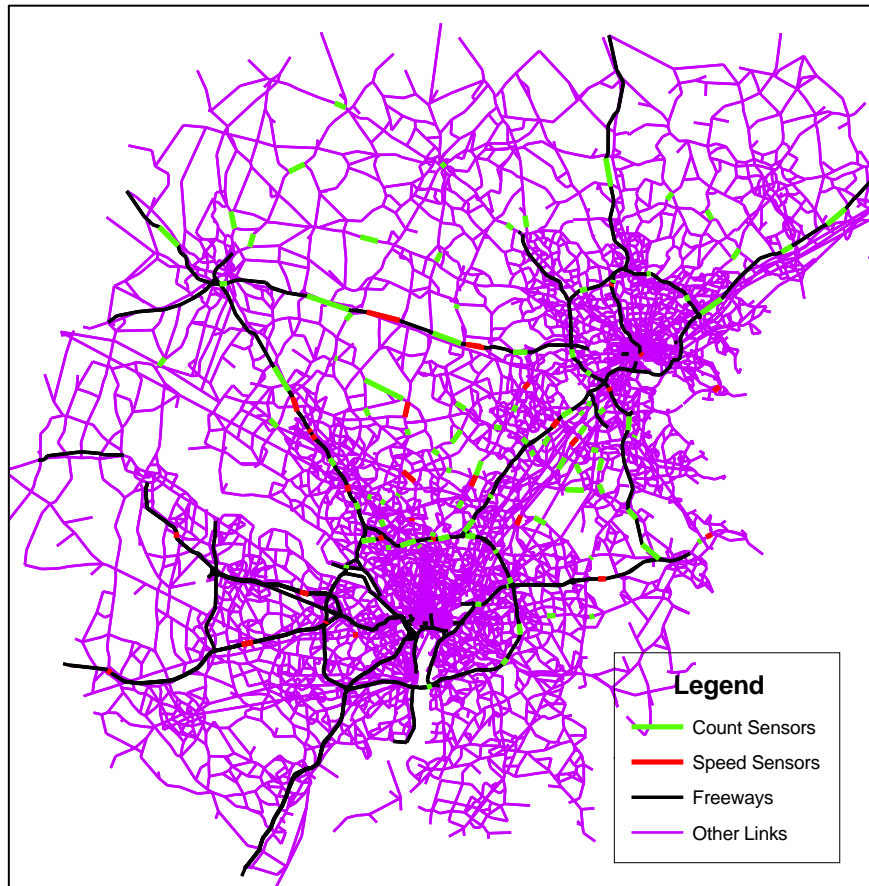
Figure 3- 7 Washington D.C. system model transportation network

The traffic counts calibration of this large-scale DTA model is an offline process with hourly volume data of 182 loop detectors located on most of the freeway corridors of the network (green points in Figure 3-7, 88 on freeways, 94 on other roadways). The data is obtained from the Maryland SHA internet traffic monitoring system (I-TMS). I-TMS provides volume data and turning movement data that covers freeway and major/minor arterials within the state of Maryland. The counts data is historical data detected during the continuous 48 hours of Tuesday and Wednesday. For the 182 detectors, I select the volume data detected near April 2015 for the base DTA model. The seed demand comes from MSTM planning model. The total demand is around 14,138,294 trips, including a warming up period from 5:00 to 6:00 AM.

DTALite's built-in procedure ODME is applied for the OD calibration. The ODME procedure performs a dynamic assignment to pursue the UE condition, then adjusts the OD table based on route choices under UE. For such a large network, the ODME requires 26 hours for 100 iterations (20 iterations for UE, 80 iterations for OD adjustment) under a 2-core, 24-CPU, 84-GB memory server. It takes numerous rounds of demand calibrations to match hourly counts. More details about ODME will be discussed in Chapter 4.

After adjusting demand parameters in traffic counts calibration, the research team applies historical speed data and adjusts supply side parameters for supply parameters calibration and validation. The regional integrated transportation information system (RITIS) commercial traffic dataset from analytics company INRIX, which includes link-based speed and travel time data, is collected and used for supply parameters validation. The links selected for supply parameters calibration and validation is marked with red points in Figure 3-7. After collecting the hourly average speed of these links for the afternoon peak period in April 2016, the research team compares real-world data with the DTA output and adjusts link capacity and link jam density to improve the accuracy.
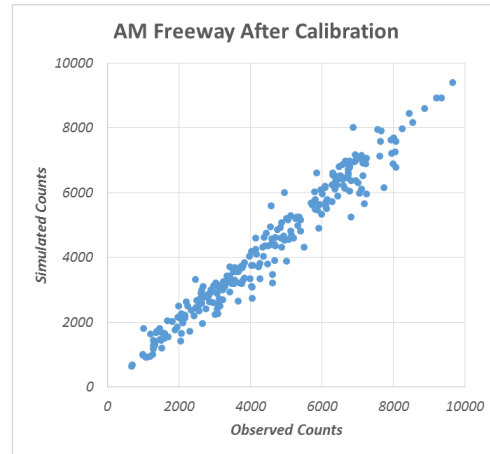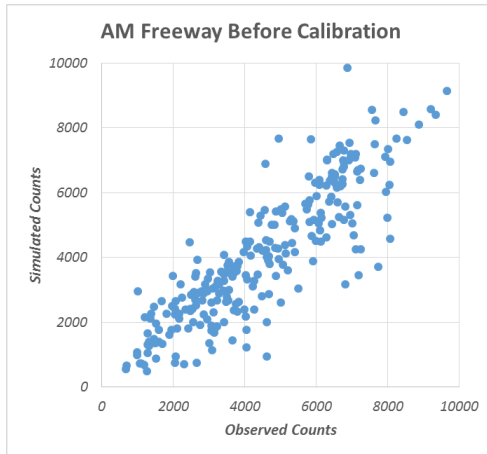
(a) AM peak freeway counts plot (22.4% to 9.2%)
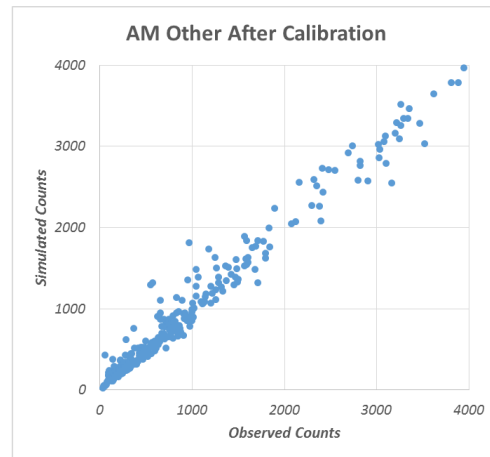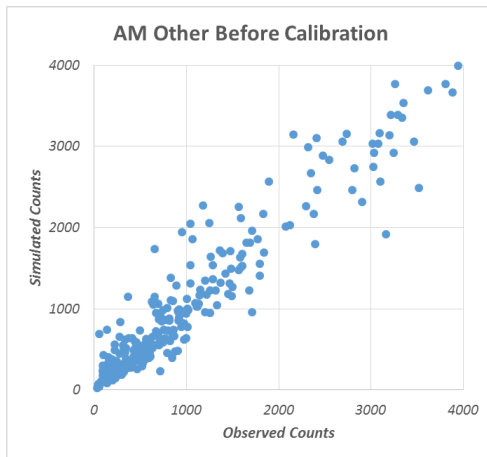
(b) AM peak other roadways counts plot (24.4% to 12.1%)
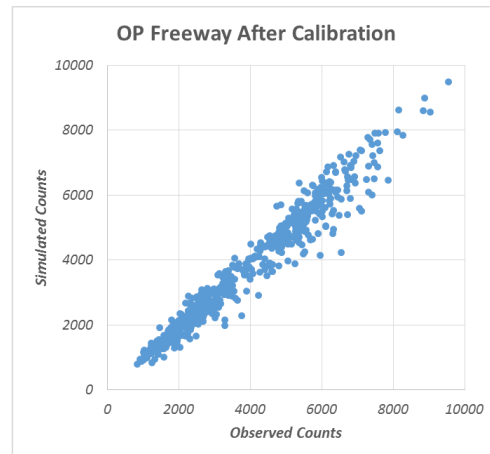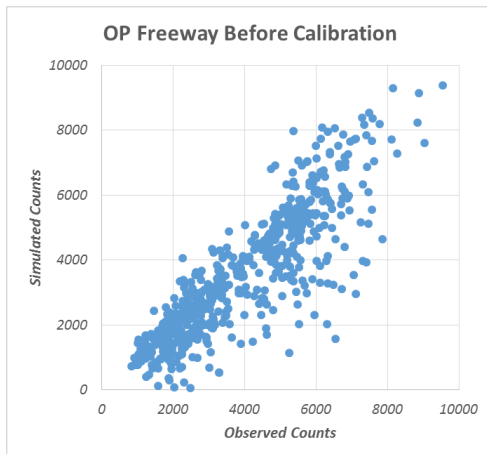
(c) Off peak freeway counts plot (24.8% to 9.9%)
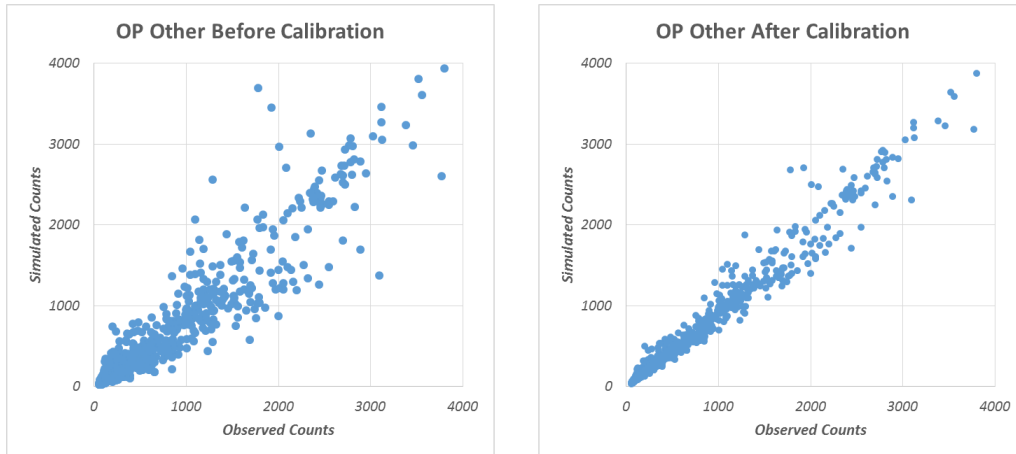
(d) Off peak other roadways counts plot (26.9% to 13.5%)
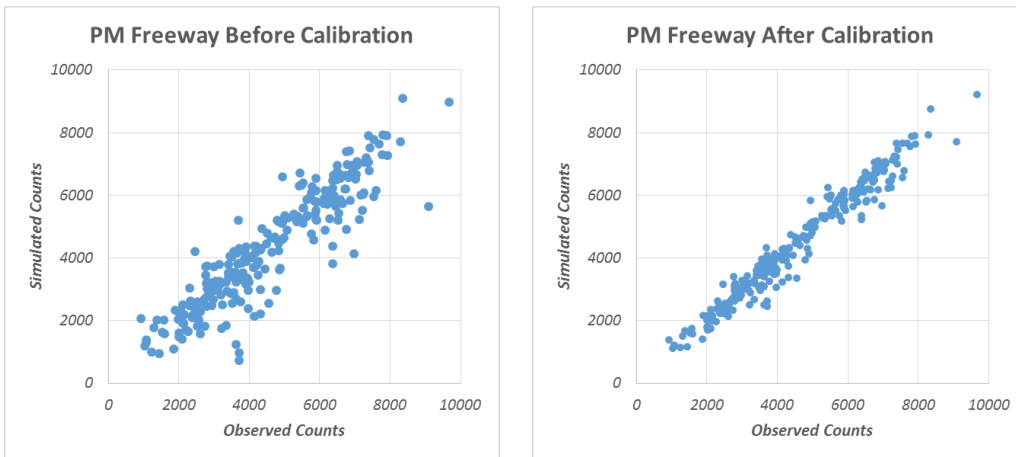
(e) PM peak freeway counts plot (18.8% to 9.4%)
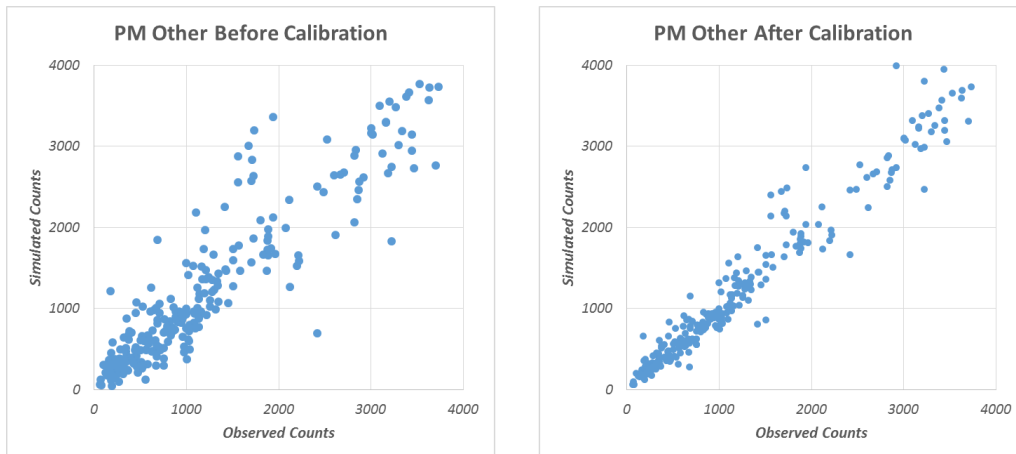
(f) PM peak other roadways counts plot (28.4% to 14.1%)

Figure 3- 8 Traffic counts calibration results of DTA simulation model

WMSE indicates that the normalized relative errors between simulated and observed traffic counts decrease significantly after the calibration, as shown in Figure 3-8. Additionally, the model confirms its validity by adjusting supply parameters based

on speed comparison. Figure 3-9 shows that the WMSE between the simulation speed

and historical average speed decrease from 19.7% to 11.4%.
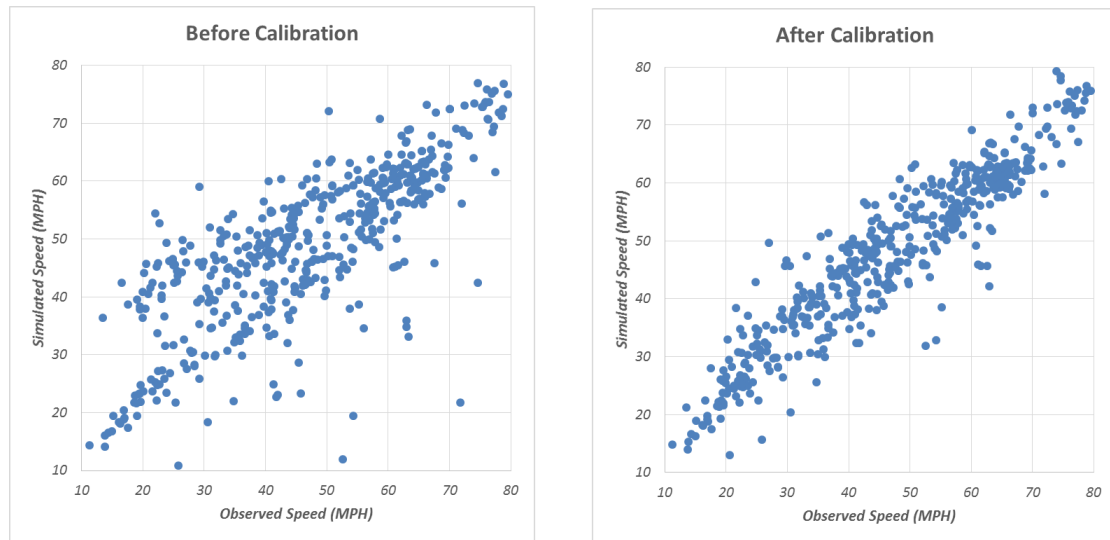
Figure 3- 9 Traffic speed calibration results of DTA simulation model

I have already finished on collecting historical traffic data for the test of a

gradient-based online calibration approach. Details on the methodology for online

calibration and the case study using this large-scale network will be introduced in

Chapter 4 and Chapter 5, respectively.

Section 3.7. Summary

This chapter integrates a mesoscopic DTA with the agent-based positive

travel behavior model as a system transportation network model (AgBM-DTA). In the

proposed framework, travelers no longer have perfect network knowledge to maximize

their travel utility. Instead, they are learning and searching for better choices to decrease

their costs due to delay by scheduling early and scheduling late. The integration with

the positive model enhances the behavior realism of DTA, resulting in the capability to

capture dynamic travel behavior pattern changes. This integration can be a valuable

approach for planning agencies to conduct studies on land developments, traffic-related

policies, and even a combination of the two. SBO is used to calibrate the AgBM-DTA.

A large-scale network model is developed and calibrated via historical data. The

application of this model for online calibration and decision support will be illustrated

in Chapter 5.

Chapter 4: System Model Online Calibration

Section 4.1. DTA online calibration versus offline calibration

In Chapter 2, the focuses of offline DTA calibrations and online calibrations have already been mentioned. On one hand, offline calibrations usually need a rich source of historical data to make sure the model can represent typical real-world scenarios (e.g. general weekday) as much as possible. The estimations of demand and supply parameters, which usually utilize optimization algorithms or heuristic algorithms, may take a long time to converge. On the other hand, online calibrations want to capture the dynamic change of traffic conditions. The pursuit of "optimal" parameters may not be applicable due to computational efficiency, especially for large-scale networks. Therefore, online calibrations are more like "parameter adjustment" processes rather than "parameter optimization" processes.

Many of the online DTA calibration approaches discussed in Chapter 2 consist of two processes: 1) offline calibration of one or different DTA scenarios as typical models; 2) using adaptive algorithms to adjust parameters via real-time data. For large-scale real-time applications, the computational speed is a critical issue for online calibrations. Concerning this, I propose a gradient-based online calibration procedure that quickly estimates the gradients for demand and supply parameter towards the gap between real-world observations and simulated measurements.

Section 4.2. A gradient-based fast online calibration approach

This section provides a brief introduction of the proposed online calibration approach. Like other approaches, the general idea of the proposed gradient-based online calibration approach is to utilize a number of offline models, which can represent different scenarios, as the basis to save computational time. These offline models are treated as base models, the route choice of the base models will be used to estimate the gradients of demand and supply parameters towards the calibration objective (the gap between the simulated and observed traffic measures). Based on the deviation between the offline simulation performance and the real-time traffic data, one will estimate the gradients and adjust demand and supply parameters online.
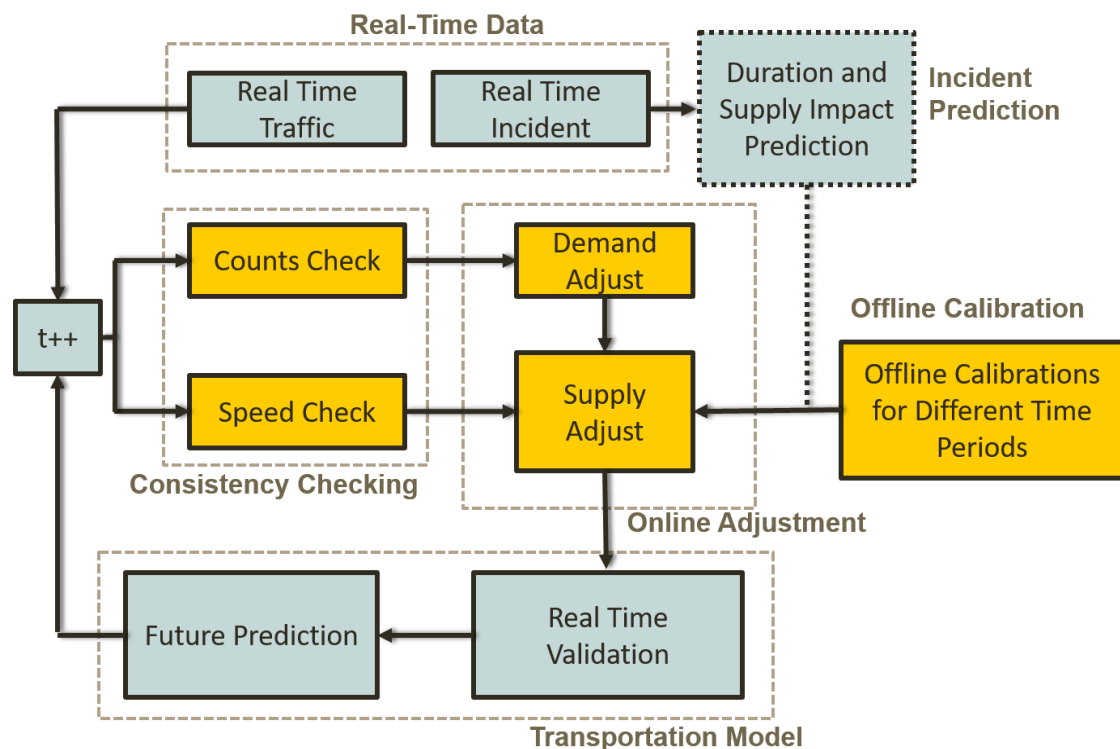
Figure 4- 1 Gradient-based online calibration approach

The flowchart of the proposed online calibration approach is shown in Figure

4-1. Before conducting the online approach, a rich source of historical traffic data is required for offline calibrations to seek typical model parameters for different times of the day. These offline model parameters (e.g. OD matrices and supply side parameters) are regarded as the basis for online adjustments. DTALite's demand calibration module ODME is utilized for this process. More detailed about the offline calibration will be covered in Section 4.3. In order to achieve a fast and reasonable online parameter adjustment, I develop a network performance checking method that uses some state vectors to describe the real-time condition of the transportation system compared with the historical average condition.

In demand parameter adjustment, I first check the difference between real-time traffic counts and historical average counts, which will be presented by the deviation of some state vectors. The DTA demand parameters will be adjusted based on both the deviation of these state vectors and some pre-defined adjustment rules. I propose a two-level travel demand performance checking method. The performance of transportation system model is described by two indicators: pattern indicator and sensor indicator (or sensor vector). Pattern indicator illustrates the overall difference between the historical counts and the real-time observed counts for the entire network; while, sensor indicator explains the consistency between historical average and real-time counts data for individuals or groups of counts sensors in the transportation modeling system. The correlations of the counts sensors are considered to describe the traffic conditions. The sensor indicator is a vector with dimensions fewer than the original number of counts sensors, so that some sensors are grouped together as one dimension

in the sensors indicator vector. Details on how to group sensors to decrease measurement dimension are illustrated in Section 4.4. Once sensors' counts are mapped to a sensor indicator vector, one can compare the difference (deviation) between the real-time sensor indicator vector and the historical average sensor indicator vector. The historical average vector for the modeling time period is associated with an offline model with calibrated OD matrices, supply parameters and route choices. Thus, based on the deviation, one knows which dimensions in the sensor indicator (groups of sensors) are different from the historical average values. Their associated OD pairs can be identified from the route assignment of the offline model. Therefore, OD volume adjustments can be made. More details on this online OD matrices adjustment method will also be discussed in Section 4.4.

Supply parameters are adjusted right after demand adjustment. The gradients of supply parameters towards the gap between simulated and observed speeds will be estimated based on the queueing diagram estimated from the demand calibrated model. The basic idea is to adjust free flow speeds if the link is uncongested, while adjusting the capacity if the link is congested. I should admit that this is a strong but reasonable assumption, because capacity will not affect simulated speeds once there is no queue on the link; and free flow speed will not make a difference once the link is congested. More details on this online supply parameter adjustment method will be illustrated in Section 4.5. DTA will be conducted again for speeds validation and near-future traffic condition predictions. In the future, a real-time incident information module will be used to detect and predict incidents that may cause huge capacity reduction in the

network. The incident information module will provide the transportation model with real-time incidents in terms of locations, begin time, and number of lanes blocked. Statistical or machine learning models will be used to predict the duration and impact of incidents to adjust link capacity and speed limit. This real-time incident data and duration prediction, which has not been completed at the current stage, will be fulfilled in future research. With the online calibrated model, one is able to conduct near-future traffic predictions for real-time traffic management or other ITS.

With the idea above, I decide to use one hour as the model time interval. For each hour of the day, an offline DTA model is calibrated as the basis. Real-time traffic counts data is collected and gathered hourly for consistence checking, then used for demand adjustment; speeds data is also gathered hourly for supply parameters adjustment. The proposed online parameter adjustment method is then applied to estimate real-time DTA parameters. The last hour's calibrated model can be used for network warm-up, and historical offline parameters for the next hour is used to predict near-future traffic. One can update the offline calibration dataset over time to incorporate the changes of travelers' driving behavior, the transportation network, and other variables in the long run.

Section 4.3. Offline calibration

The DTA package DTALite used in this dissertation has a built-in OD matrices estimation module called "ODME". ODME is a fast path-flow based optimization model which utilizes sensor counts, seed OD matrices, and DUE

assignment results to estimate a set of path flow volumes. Under the ODME module, DTALite will first run U DTA iterations (e.g. 20) to seek a path assignment under DUE, and then it will conduct anther V iterations (e.g. 40) to adjust OD tables based on the DUE path flows.

Similar with other offline OD calibration methods, ODME attempts to minimize the combined term of simulated errors and the difference between seed OD and target OD:

$$\min: \quad \delta \sum_{k} \sum_{h} \left(OD_{k,h} - OD_{k,h}^{Seed} \right)^2 + \sum_{i} \sum_{h} \left(x_{i,h} - x_{i,h}^{*} \right)^2 \qquad (4\text{-}1)$$

where δ is a predetermined coefficient, $x_{i,h}^{*}$ is the simulated link flow, $x_{i,h}$ is observed flow. An approximate gradient method, which utilizes a queue model to calculate link flow–density change due to incoming path flow change, is used to solve this optimization problem. The detailed mathematical formulation and solution algorithm can be referred to a recent paper by Lu et al. (2013).

After OD matrices calibration, historical average speeds data is used to calibrate the supply parameters (e.g. capacity and free flow speed). The methods for supply parameter adjustment is the same with the proposed online algorithm, which will be introduced in Section 4.5.

Section 4.4. Counts consistency checking and online demand adjustments

The counts consistency checking includes two parts: 1) pattern level checking; and, 2) sensor level checking. Here consistency refers to the differences between real-time traffic counts data and historical average counts data. The differences and offline

calibrated models will be used for online demand (OD matrices) adjustments. This section talks mainly about the consistency checking and online OD adjustment procedure.

4.4.1. Pattern level checking

Pattern level checking aims at detecting whether the entire system demand is higher or lower than the historical average demand. This compares real-time data with historical average data. Suppose $x_h = \left(x_{1,h}, x_{2,h}, ..., x_{p,h} \right)$ to be the vector of all the sensors' hourly counts at hour h, $x_h^H = \left(x_{1,h}^H, x_{2,h}^H, ..., x_{p,h}^H \right)$ is the historical average count vector. The pattern level checking attempts to obtain the ratio between the summations of the two vectors' elements.

$$dem_h = \frac{\sum_{i=1}^{p} x_{i,h}}{\sum_{i=1}^{p} x_{i,h}^H} \tag{4-2}$$

The demand factor dem_h will be used to online adjust the overall demand for hour h.

4.4.2. Principal component-based sensor level checking

A medium or large-scale DTA network usually needs numerous sensors for calibration. Compared with historical average counts, real-time sensors will tell traffic managers living traffic information. While, it is obvious to notice that the sensors' volumes are not independent from each other. That is, in some cases, two or more sensors are located along the same corridor with upstream and downstream relationships, resulting in positive correlations; in some cases, the sensors are on

parallel corridors that may be negatively correlated due to route choice. In order to use some measure to describe local traffic conditions, as well as incorporate the correlations among the sensors, the author uses Principal component analysis (PCA) to linearly map the sensor counts onto a lower dimension vector.

PCA is a statistical procedure that uses a linear orthogonal transformation to convert a dataset of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (PC). The number of PCs is less than or equal to the smaller of the number of original variables or the number of observations. This transformation is defined in such a way that the first PC has the largest possible variance (i.e. it accounts for as much of the variability in the data as possible), then each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors formulate an orthogonal basis set. Since PCA is sensitive to the relative scaling of the original dataset, one usually scales the data before conducting PCA. Suppose X is a n by p matrix, in which n denotes the number of data records and p denotes the original dimension of the data, P is a p by p transfer matrix, and Y is the result matrix, such that

$$Y=X \bullet P \tag{4-3}$$

PCA aims at seeking a P such that the entries of Y are independent with each other. That is, the covariance of Y, which is $S_Y = \frac{1}{n-1} Y^T \bullet Y$, is a diagonal matrix. Substituting equation (4-3) into S_Y, one has:

$$S_Y = \frac{1}{n-1} (XP)^T \bullet (XP) = \frac{1}{n-1} P^T (X^T X) P \tag{4-4}$$

Since $X^T X$ is a symmetric matrix, we know from linear algebra that it can be decomposed as:

$$X^T X = EDE^T \tag{4-5}$$

where D is a diagonal matrix and E is a matrix of eigenvectors of $X^T X$ arranged as columns. Therefore, substitute equation (4-5) into equation (4-4), one has:

$$S_Y = \frac{1}{n-1} P^T \left(EDE^T \right) P \tag{4-6}$$

Since E is orthogonal matrix, it is noticeable that $EE^T = I$. Therefore, if one takes P to be E, then

$$S_Y = \frac{1}{n-1} E^T \left(EDE^T \right) E = \frac{1}{n-1} D \tag{4-7}$$

is a diagonal matrix. Therefore, PCA can be achieved by making the transfer matrix to be the eigenvectors matrix of $X^T X$.

After PCA, the n by p sensor counts matrix has been linearly transferred to another n by p matrix (Y). This time, since the p dimensions of vector Y are independent with each other, one can rank their variances and pick the dimensions whose variances can represent the majority of the total variance to decrease dimensions. That is, one can do column exchange of P such that the diagonal values of D, which is denoted as d_i (i from 1 to p) is decreasing. In this dissertation, I select first q entries of Y such that $y_1, y_2, ..., y_q$ can represent at least 95% of the total variance of the data. I will utilize the new vector $y_1, y_2, ..., y_q$ as the sensor level vector to describe the network performance.

4.4.3. Gradient-based online demand adjustment approach

Based on the consistency check procedure, I proposes a fast OD matrices adjustment method. I would like to emphasize on "fast" because most of the work is supposed to be finished during the offline typical model calibration. Unlike space-state models (Antoniou et al. 2007), or transfer equation models (Tavana and Mahmassani, 2000), the proposed method does not need to iteratively estimate traffic states; Instead, I attempt to straightforwardly make online adjustments to save computational time.

The flowchart of the proposed online DTA demand parameter adjustments approach is shown in Figure 4-2. This OD matrices adjustment procedure request two assumptions:

1) the summation of sensor counts are positive related to the overall demand.

2) the route choice under offline model is similar to real-time scenarios during the same time period, unless there are non-recurrent congestions.

One critical issue in the two assumptions is that once the travel demand is much over normal level, the increasing of demand may have negative effect on sensor counts due to congestion. In order to fill up this weakness, I take into account real-time speed information into the supply parameter calibration and validation.
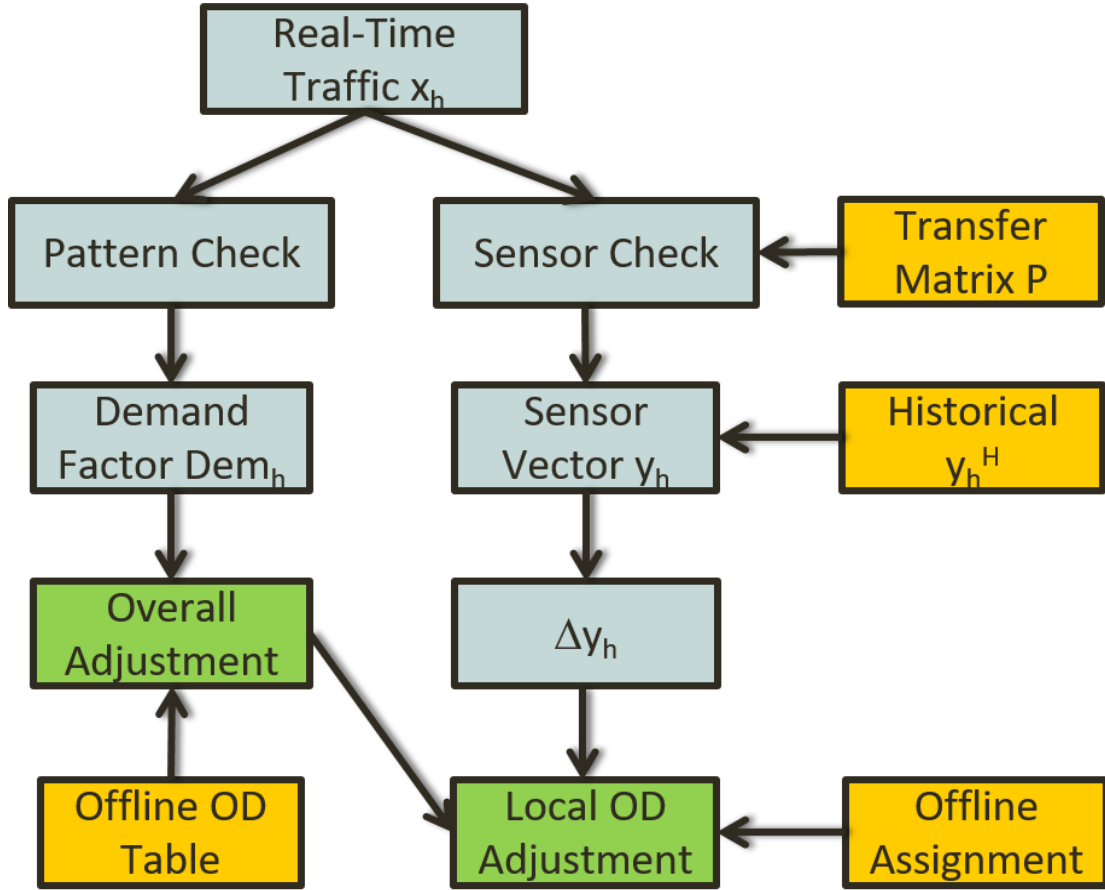
Figure 4- 2 Flowchart of online DTA OD matrices adjustments

From the consistency checking process discussed in previous sections, one is able to obtain two state indicators dem_h and y_h. dem_h reflects the overall real-time demand compared with historical level; while y_h provides local network condition. Let $OD_{k,h}^H$ denote the historical demand for OD pair k during hour h, α be a pre-defined overall adjustment coefficient. The overall demand is adjusted via:

$$OD_{k,h}^1 = OD_{k,h}^H \cdot \left[1 + (dem_h - 1) \cdot \alpha\right] \qquad (4\text{-}8)$$

where $OD_{k,h}^1$ denoted the temporary demand before local adjustments.

Suppose we only keep the first q dimensions of y_h in the PCA process such that these dimensions make up 95% of the total variance. Thus the transfer matrix is a p by q matrix and the y_h can be expressed as:

$$y_h = x_h \bullet P \tag{4-9}$$

Let y_h^H denote the historical average sensor vector, the difference can be written as:

$$\Delta y_h = y_h - y_h^H \tag{4-10}$$

Before online local adjustments, I need some preparations offline. Let P_i denote the *i*th row of matrix P, P_{ij} denote the *j*th element of that row. Then, it is reasonable to suppose P_{ij} can represent the contribution of sensor *i* on the *j*th PC in vector y_h. For each P_i I conduct a rank based on the absolute value of P_{ij} ($|P_{ij}|$); if for some j^*, $|P_{ij^*}|$ is the largest, I assume sensor *i* has the most significant impact on PC j^*. Based on the historical data, I first conduct PCA; then for the *p* by *q* transfer matrix, I attempt to find the significant PC for every sensor; if sensor *i* has the most significant contribution to PC *j*, then I put sensor *i* into set S_j, which contains all the significant sensors to PC *j*. This process is illustrated in Figure 4-3.

```
Begin
    For j = 1 to q
        set Sⱼ = {}
        For i = 1 to p
            If |Pᵢⱼ| is the max among |Pᵢ|
                Put i into Sⱼ
    Return all Sⱼ
End
```

Figure 4- 3 Find significant PC for every sensor

The purpose of defining S_j is to know which sensors will be involved if dimension *j* in Δy_h is significant. While, one still needs to know which OD pairs should be adjusted based on Δy_h and S_j. After I obtain S_j for all the *q* dimensions

(PCs), another offline preparation is to get the route choice and path assignment based on historical calibrated models. Let $R_{i,h}^{H}$ denote a set of OD pairs, such that all the OD pairs that passed sensor i during hour h are contained in the set. $R_{i,h}^{H}$ can be easily obtained by summarizing the trip trajectories from the output of the offline calibrated models. Let $C_{i,k,h}^{H}$ be the counts that OD pair k is contributing to sensor i during hour h from the offline model, and β is a predefined coefficient.

After getting P, $R_{i,h}^{H}$, $C_{i,k,h}^{H}$ and y_{h}^{H} during the offline process, the online local OD adjustments are preformed based on the following algorithm (Figure 4.4) and equations:

Begin
    For j = 1 to q
        Calculate TP$_j$
        For i in S$_j$
            Calculate TC$^H_{i,h}$
            For k in R$^H_{i,h}$
                Calculate OD2$_{k,h}$
End

Figure 4- 4 Local OD matrices adjustment algorithm

$$TP_{j}=\sum_{i\in S_{j}}\left|P_{i,j}\right|$$

(4-11)

$$TC_{i,h}^{H}=\sum_{k\in R_{i,h}^{H}}C_{i,k,h}^{H} \tag{4-12}$$

$$OD_{k,h}^{2}=OD_{k,h}^{1}+\Delta y_{h,j}\bullet\frac{P_{i,j}}{TP_{j}}\bullet\beta\bullet\frac{C_{i,k,h}^{H}}{TC_{i,t}^{H}} \tag{4-13}$$

Equation (4-11) to (4-13) means the local OD flow will be adjusted considering the contribution of the OD pairs to the sensor counts, as well as the contribution of the sensors to the PC dimension. After overall adjustments and local adjustments, the new demand will be loaded into the system model for traffic condition validation and prediction.

Section 4.5. Gradient-based online supply parameter adjustment approach

Currently, a lot of supply parameter calibration work has been done based on optimization approaches that attempts to search for an optimal combination of parameters to minimize the gap between real-world observations and simulation outputs. Although some researchers utilize state-space approaches to save computational time to make online calibration applicable, such filter-based methods are not perfect for mesoscopic model calibration (Ashok and Ben-Akiva 1993; Ashok 1996; Ashok and Ben-Akiva 2000; Antoniou et al. 2007). This is because any state-space models, transfer function models, or related filtering methods are trying to estimate the relation between supply parameters and simulation outputs. While, in mesoscopic DTA models, there is already predefined relation between the supply parameters and the link performance due to the macroscopic traffic model used in the DTA simulator. For instance, some simulators use BPR functions, some simulators use point queue models, and some simulators use more complex traffic models (Zhou and Taylor 2014). As a result, the estimated relation via state-space approaches can at most be the same accurate compared with the predefined models to describe the relation between supply

parameters and simulation outputs. With such concerns, the proposed online supply

parameter adjustment approach will not "estimate" the relation between model inputs

and outputs; instead, I propose an idea of estimating the gradient of supply parameter

adjustment directly from the queueing and discharging process.

I assume the free flow travel speed and capacity will not be constants but may

change with different factors, e.g. weather, time of day, incidents, etc. Link travel speed

will be affected by demand, free flow travel speed and capacity at the same time. Once

the demand is properly adjusted, the free-flow travel speed and capacity will be adjusted

during this online supply calibration process to ensure the validity of the system model

in terms of speed performance. I will estimate the simulated speed via the departure-

arrival diagram obtained from the simulation model; and make adjustments based on

the deviation between observed real-time link speeds and simulated speeds.

4.5.1. Queueing diagram under uncongested and congested scenarios

Once the demand adjustment has been done, the demand pattern for each link

is determined because in DTA models demand calibration usually refer to both time

dependent OD matrices and their corresponding time dependent route choices. Thus,

for one link with speed sensor, the departure-arrival diagram can be obtained from the

model with the calibrated demand.

Under uncongested case, i.e. there is no queue at all, the queueing diagram is

like Figure 4-5. In this figure, the horizontal axis denotes time, the vertical axis denotes

cumulative link counts, the solid blue curves denote cumulative arrival and departure

vehicles, the solid red line denotes the simulated travel time, which is also free-flow travel time (uncongested case). If the simulation is under estimating travel time (the real-world travel time is denoted by the dash red line), then one can adjust the free flow travel speed and obtain a new departure curve (dash gray line). In other words, I assume under uncongested condition, link speed (or travel time) is only affected by free-flow travel speed. Another way to calculate the average travel time is to use the total vehicle delay (the pink area, denoted as VT) divide the total number of arrivals between time h_1 and h_2 (denoted as A_{12}):

$$mt^* = \frac{VT}{A_{12}} \tag{4-14}$$

$$s^* = \frac{l}{mt^*} = \frac{l \cdot A_{12}}{VT} \tag{4-15}$$

where mt^* denotes the simulated link travel time (average travel time), s^* denotes the simulated link speed, l denotes the link length.
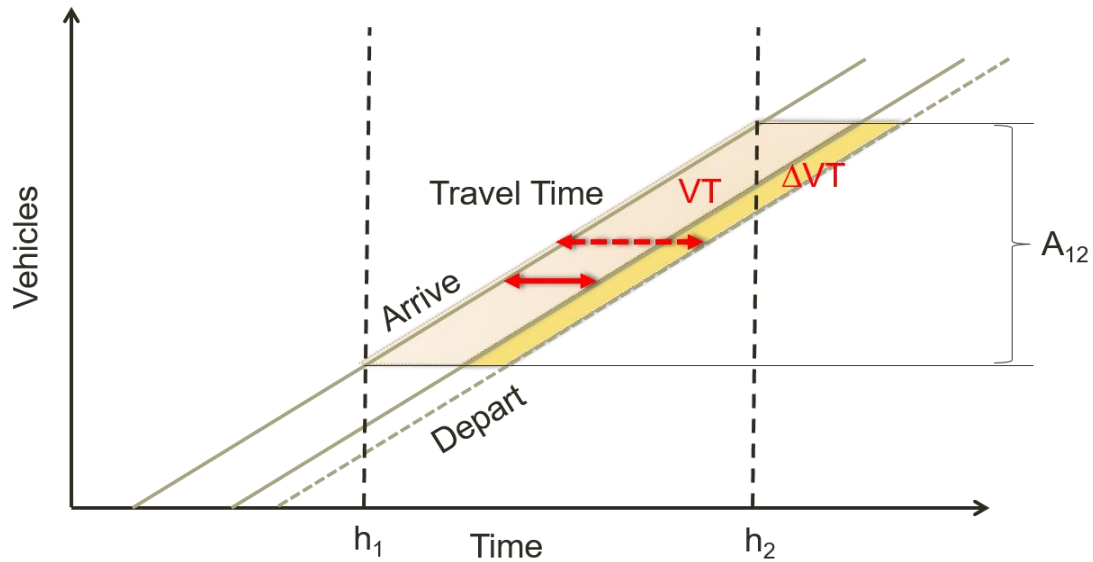
Figure 4- 5 Queueing diagram under uncongested condition

If the demand is over capacity, I assume the capacity needs to be adjusted for

online calibration. For instance, if the link travel speed is over estimated via the simulation, one needs to decrease the capacity. After decreasing link capacity, there will be a difference in terms of total vehicle travel time. This is shown in Figure 4-6, where the dash gray line denotes the departure curve after capacity change, and ΔVT denotes the change in terms of total vehicle delay. The deviation between detected travel speed and simulated speed provide us the gradient to decrease link capacity. The adjusted capacity can be determined via equation (4-14) and equation (4-15).
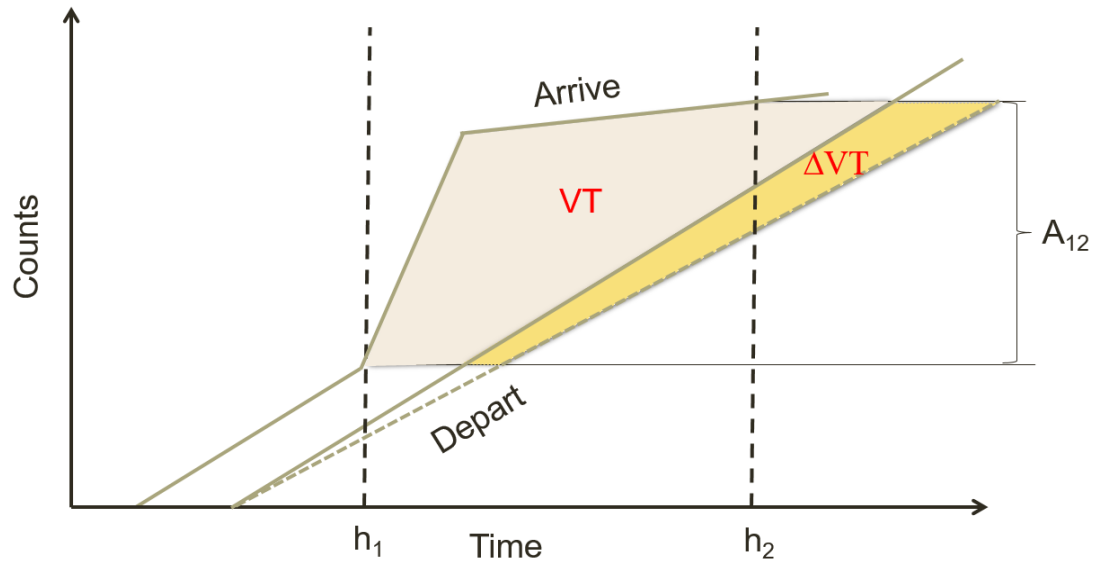
Figure 4- 6 Queueing diagram under congested condition

Given the general ideas above, detailed algorithm of the proposed online supply parameter adjustment approach will be introduced in next subsection.

4.5.2. Gradient-based online supply parameter adjustment approach

From previous subsections, I have discussed the way to estimate simulated

link speed (or travel time) based on queueing diagram for both uncongested and congested traffic condition. The proposed online supply parameter adjustment algorithm is developed on the behavior of queuing and discharging process. Before applying this algorithm, one should apply the demand adjustment algorithm to ensure the validity of the demand for the model. Then, the cumulative arrivals and departures based on the calibrated demand can be used for supply parameter adjustments.
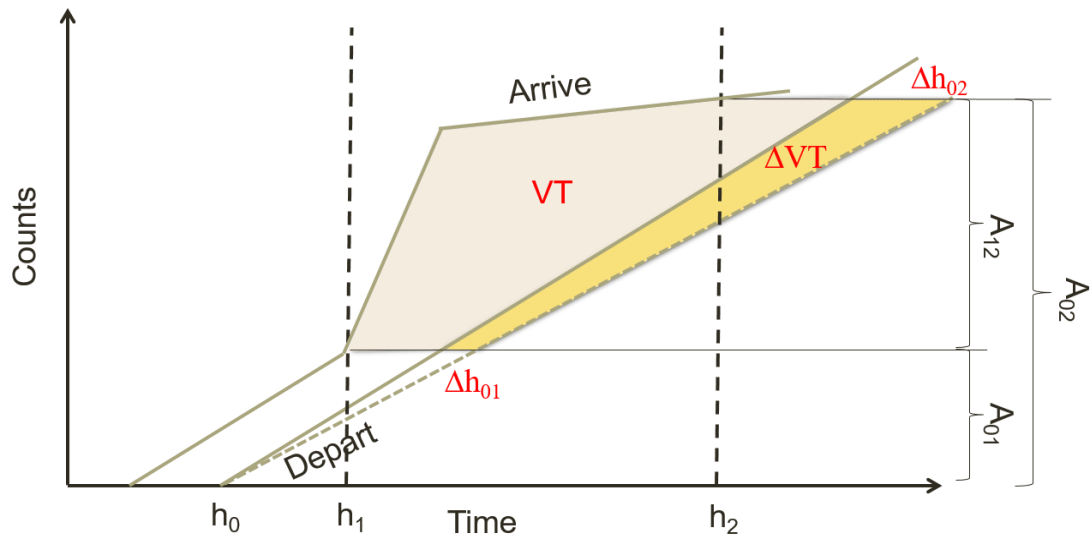
Figure 4- 7 Online supply adjustments under congested condition

In Figure 4-7, suppose h_0 denotes the beginning of the link discharging process, A_{01} / A_{02} denotes the total arrivals between the beginning of arrivals at h_1 / h_2, $\Delta h_1 / \Delta h_2$ measures the difference of time after the capacity adjustments, C^H / C^1 denotes the historical/adjusted capacity, s^* / s denotes the simulated/real-world link travel speed, ffs^H / ffs^1 denotes the historical/adjusted free flow travel speed. Notice that in this dissertation, I assume the historical parameters to be the same as

offline models' parameters. s^*, A_{01}, A_{02}, A_{12}, can be obtained from simulation outputs. Firstly, one needs to check whether the link is congested or not. If s^* is closed to the predefined ffs^H in the DTA simulator ($s^* \geq 0.95\,ffs^H$), one can say there is no queue and the link is uncongested; otherwise, the link is congested. The online supply parameter algorithm is as follows:

1) If the link is uncongested:

$$ffs^1 = s \tag{4-16}$$

2) If the link is congested, assume $C^1 = \eta C^H$, where η is the adjustment ratio. By the definition of capacity, one has:

$$\Delta h_{01} = \frac{A_{01}}{\eta C^H} - \frac{A_{01}}{C^H} \tag{4-17a}$$

$$\Delta h_{02} = \frac{A_{02}}{\eta C^H} - \frac{A_{02}}{C^H} \tag{4-17b}$$

$$\Delta VT = \frac{1}{2}\left(\Delta h_{01} + \Delta h_{02}\right)A_{12} = \frac{1}{2C^H}A_{12}\left(A_{01} + A_{02}\right)\left(\frac{1}{\eta} - 1\right) \tag{4-18}$$

Since one also knows s^* and s, another expression of ΔVT can be:

$$\Delta VT = \frac{l}{s}A_{12} - \frac{l}{s^*}A_{12} \tag{4-19}$$

Substitute equation (4-19) into equation (4-18), one has:

$$\frac{1}{\eta} - 1 = \frac{2C^H l\left(\frac{1}{s} - \frac{1}{s^*}\right)}{A_{01} + A_{02}} \tag{4-20}$$

One can notice that if the simulation under estimates link speed, then η is greater than 1 and the adjusted capacity is greater than original capacity from the offline model; on the contrary, if the simulation over estimates link speed, η is smaller than 1. Compared with state-space approaches, this algorithm tries to estimate the gradient

of supply parameter via an analytical way. The linear calculation can be fast and easy to implement for online DTA applications, even if the network is large.

One needs to know that this online supply parameter adjustment happens right after demand parameter adjustment, such that A_{01} and A_{02} are the estimated values after OD matrices adjustments.

Section 4.6. Integration with DTALite system model

The final step is to make this algorithm work smoothly with DTA simulators. In this dissertation, I select DTALite as the simulation engine because of its parallel computation feature for online applications. DTALite's agent-oriented feature makes it easier to achieve a fast online DTA modeling tool. Agent-oriented means it can read a list of travelers as its demand input. This list contains the origin, destination, departure time, and route choice of each traveler in the network. After preparing this list, DTALite can run one simulation-based DTA to model the traffic condition, with travelers departing and choosing routes exactly the same as the list. I utilize this agent list design of DTALite for the online calibration integration.

The integration flowchart is shown in Figure 4-8. From the offline calibrated models, one can obtain detailed travel information and the path assignment results for each traveler (contained in the output agent list). Based on the demand adjustment algorithm, I will directly modify the agent list for overall and local demand adjustment. That is, the agents associated with increasing OD pairs will be copied in the new agent list with the same travel pattern (i.e. departure time, trajectory); and the agents

associated with decreasing OD pairs will be deleted in the new agent list. With the new

agent list, one can summary the agents' trajectory such that the queueing diagram for

the links with real-time speed data can be estimated. Supply parameter adjustment then

takes place with respect to the departure-arrival diagram. The new agent list and supply

parameters are regarded as real-time inputs for the system transportation network model.

Another path fixed assignment will be run for both traffic counts and speeds validation,

as well as near future prediction. I need to admit that the path fixed assignment is too

strong to capture real-time travel behavior shifts. While, due to the limit of

computational power on large-scale networks, this is one feasible way to achieve real-

time applications. This path fixed assignment can be replaced by pursuing DUE when

more powerful DTA models or computers are ready in the market (Qu and Zhou 2017).
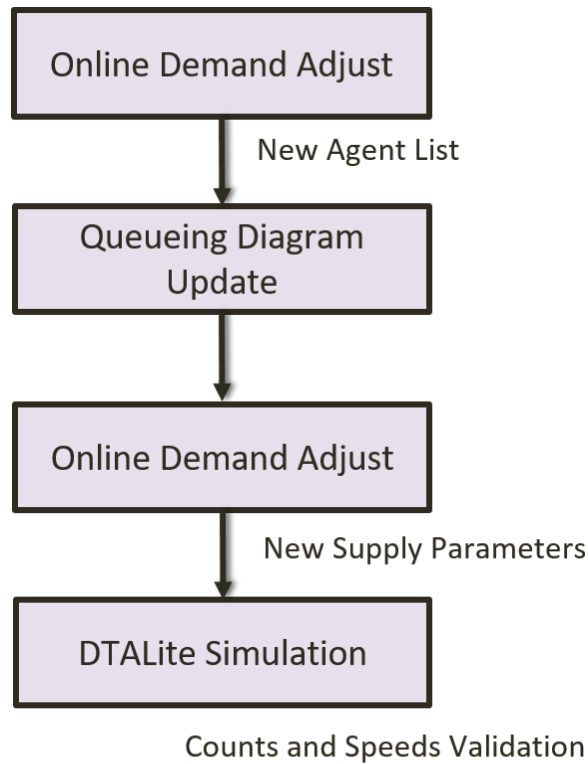
Figure 4- 8 Integrating online calibration method with DTALite

Section 4.7. Summary

This chapter attempts to propose a fast gradient-based online calibration approach. Based on historical data, hourly (or other time interval) models are calibrated offline. These offline models are treated as base models. Once real-time traffic measures are available, the deviation between the real-time data and historical data will be used to estimate the gradient of model parameters towards the calibration objective function. Both demand parameters (OD matrices) and supply parameters can be calibrated in the proposed approach. The gradients of demand parameters are estimated based on historical traffic assignment; and supply side gradients are calculated through queueing diagram.

The integration between the proposed approach and the system model (with

DTALite as engine) is discussed, too. One limitation is that with the current computational power, I need to run path fixed DTA to achieve real-time applications. This issue is expect to solve using more powerful computational technologies (Qu and Zhou 2017). Next chapter will focus on real-world applications via this online calibration approach.

Chapter 5: Applications and Real-Time Decision Supporting via System Model

Section 5.1. Introduction

This chapter talks mainly about the application of the proposed online calibration approach on a medium-scale real-world network and a large-scale real-world network. Supported by ARPA-E from the Department of Energy, a large-scale mesoscopic agent-based DTALite model has been developed for the majority of downtown areas of Washington. D.C. The geographical database of the network comes from HERE Chicago, which is a geodatabase and transportation company that provides true shape geographical data and transportation related applications. The GIS information has been converted to DTALite transportation network. The OD information comes from MWCOG directly, and has been transformed from static OD to time-dependent OD (the same method mentioned in Chapter 3) as seed ODs for calibration. The reason of using the true shape transportation network rather than MWCOG's planning network is that it can be more accurate for implementing real-time traffic management strategies, online guidance, the display of near future traffic conditions, and other applications. I use a subarea of the true shape DTA model for the medium-scale case study; while, for the large-scale case study, I still use the model developed in Section 3.6. This is because the computational speed is not high enough to model large-scale true shape networks which have a lot more links and nodes.

Section 5.2. Medium-scale general day case study

This section discusses the application of the proposed online calibration approach on a medium-scale real-world network. Since true shape network has much more nodes and links than planning networks (e.g. MWCOG network), its computational speed for real-time applications remains a critical issue. In order to test the proposed online calibration approach, I test a medium-scale true shape network in terms of the computational speed and accuracy. The network is shown in Figure 5-1, which covers the transferring area between I-95 and I-495 in north College Park, Maryland. It contains 7 TAZs, 278 links, and 242 nodes, 25 counts sensors, and 18 speeds sensors, which is directly obtained from the HERE Chicago network with MWCOG's demand via the DTALite's subarea cut procedure. All freeways, most major/minor arterials, and some local connectors/streets with true shapes are included in this network.
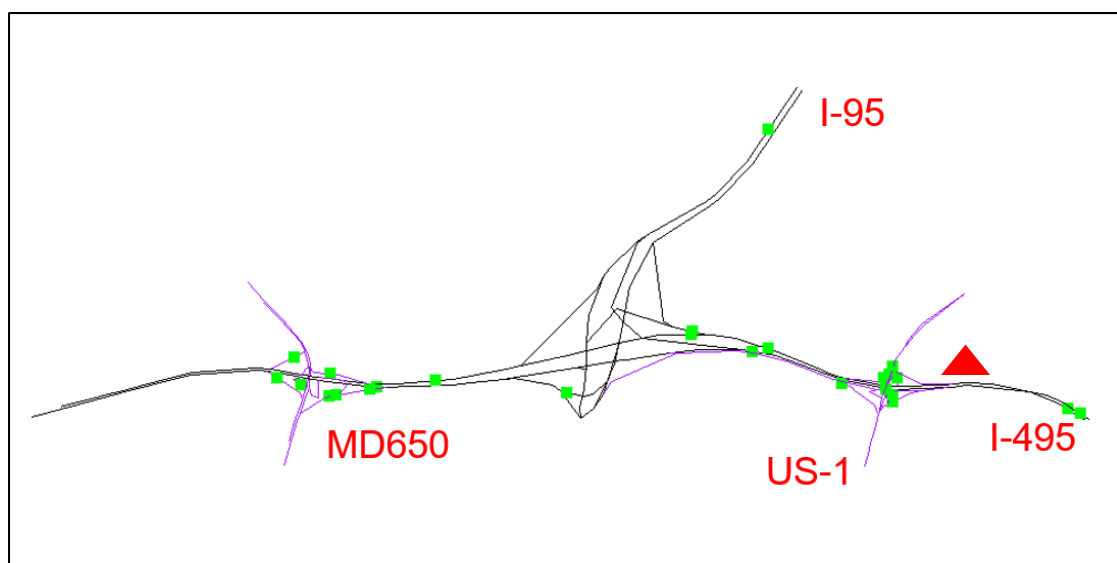
Figure 5- 1 Medium-scale true shape network

The first procedure is to obtain historical data for offline calibrations. The data used is from Center For Advanced Transportation Technology (CATT) lab's real-time traffic data application programming interface (API), which can provide sensor's real-time information (e.g. count, speed, and occupancy) on a 5-minute's basis. CATT lab has purchased RITIS data source that covers many sensor detectors on freeways operating by the state government. There are 41 sensors (green points in Figure 4-5) in the network, and the author collect two weeks' weekday data (from May 15th to May 19th, and from Jun 12nd to Jun 16th) as the historical data. The API's data, which is updated and requested for every 5 minutes, can also be used as real-time online data. These 5-minute interval counts are aggregated to one-hour interval to obtain hourly based offline models. It should be noted that in ideal cases, once there is enough time and data, one had better have offline models for each hour and for each day of week (Monday, Tuesday, etc.). In a demonstration purpose, I only consider all the five weekdays together as a "general weekday" model. I consider 7 hours' models, i.e. 4:00 to 5:00 AM, 5:00 to 6:00 AM, 6:00 to 7:00 AM, 7:00 to 8:00 AM, 8:00 to 9:00 AM, 9:00 to 10:00 AM, and 10:00 to 11:00 AM. Thus, free-flow condition, peak condition, and normal condition can all be tested. The average pattern of the summation of the 41 sensors' hourly counts are displayed in Figure 5-2, from which one can see the demand from 6:00 to 10:00 AM stays high due to commuting trips.
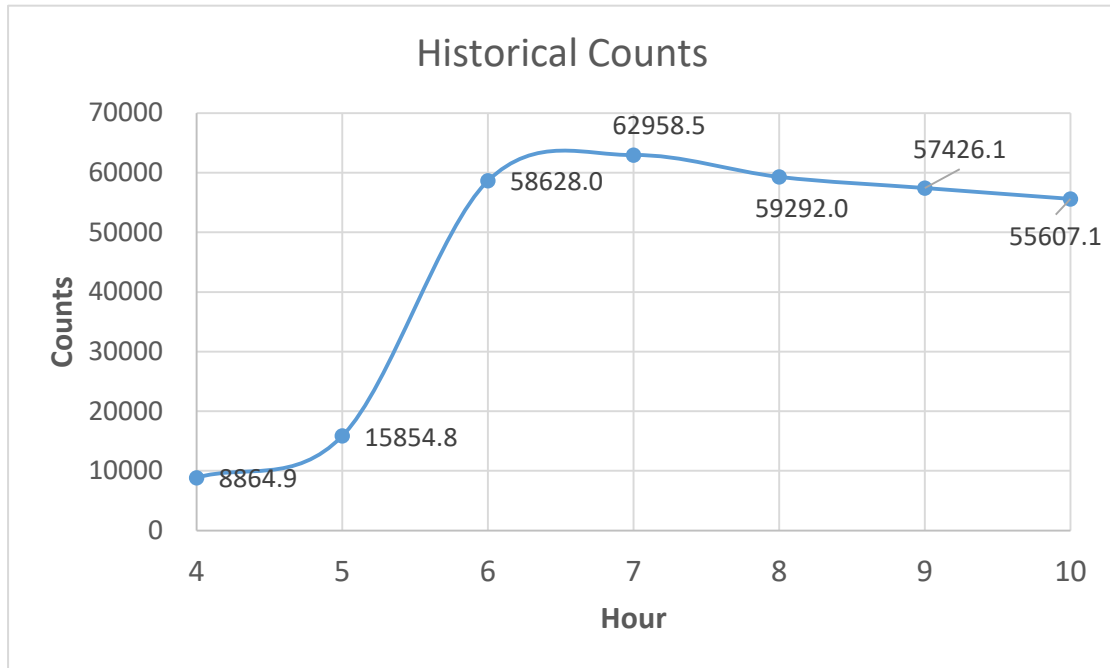
Figure 5- 2 Historical total counts pattern

Each hour's average historical counts and speeds are used to calibrate an offline typical model for that specific hour. DTALite's ODME module has been used for this offline demand calibration process, and the supply parameters are calibrated manually by similar approach compared with the proposed online adjustment algorithm (Section 4.5). Even though the network scale is not very big, it still takes around two weeks to finish the calibration of the 7 models. The observed counts and simulated counts for the calibrated off-line models for all the 7 hours are shown in Figure 4-7. WMSE, which is defined in Chapter 3, is used to indicate the performance of the offline demand calibration (Table 4-1). Similarly, the diagram and WMSE for observed speeds and simulated speeds are shown in Figure 5-3 and Table 5-1.
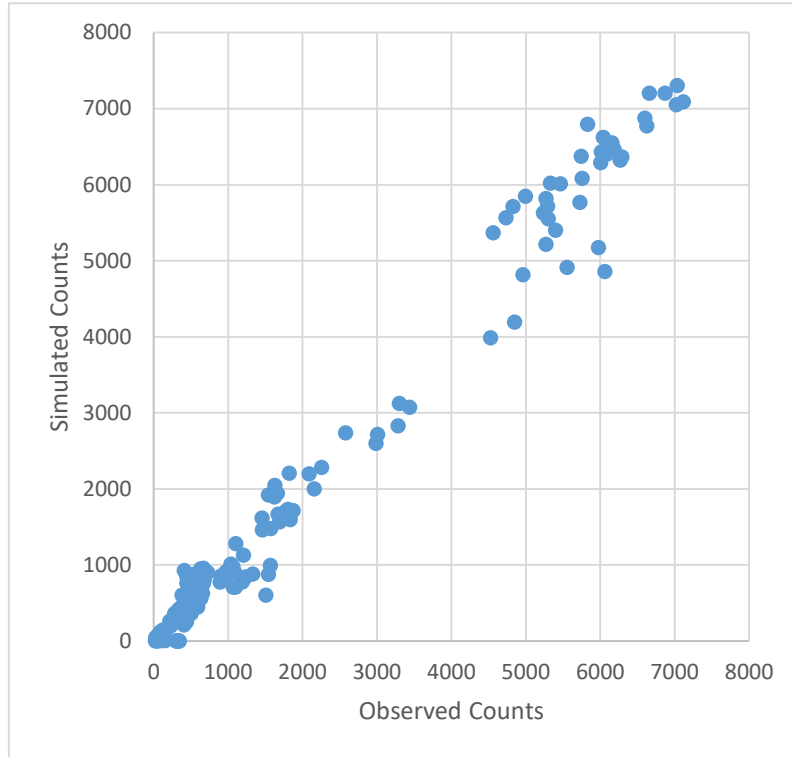
(a) Counts validation
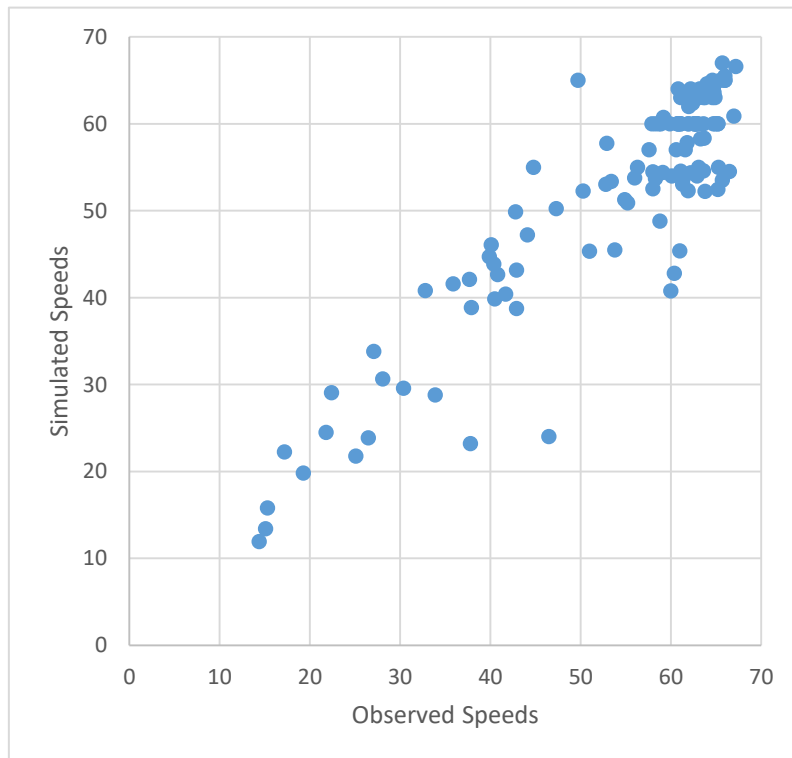
(b) Speeds validation

Figure 5- 3 Offline calibration performance for the medium-scale network

Table 5- 1 WMSE for offline calibration

Model	Counts WMSE	Speeds WMSE
4:00 to 5:00 AM	8.53	6.26
5:00 to 6:00 AM	9.53	8.73
6:00 to 7:00 AM	11.94	14.85
7:00 to 8:00 AM	11.02	17.54
8:00 to 9:00 AM	11.65	14.08
9:00 to 10:00 AM	9.33	13.34
10:00 to 11:00 AM	10.03	13.32

The 7 hours by 10 days' data (n equals to 70) has been used for PCA. The counts are scaled by the mean and standard deviation before PCA. After ranking the PCs by their variances, the author found the cumulative variance for first 14 PCs have exceeded 95% of the total variance (Figure 5-4). That is to say, based on the previous assumptions, I decide to use 14 PCs (y with 14 dimensions) to represent the network condition.

Figure 5- 4 Cumulative variance of PCA

The transfer matrix P is shown in Table 5-2. This matrix is used to obtain y_h^H based on the historical average sensor counts of different hours. The dark cells in Table 5-2 marks the largest absolute value for P_i, which means sensor i belongs to S_j.

Table 5- 2 Transfer matrix of PCA

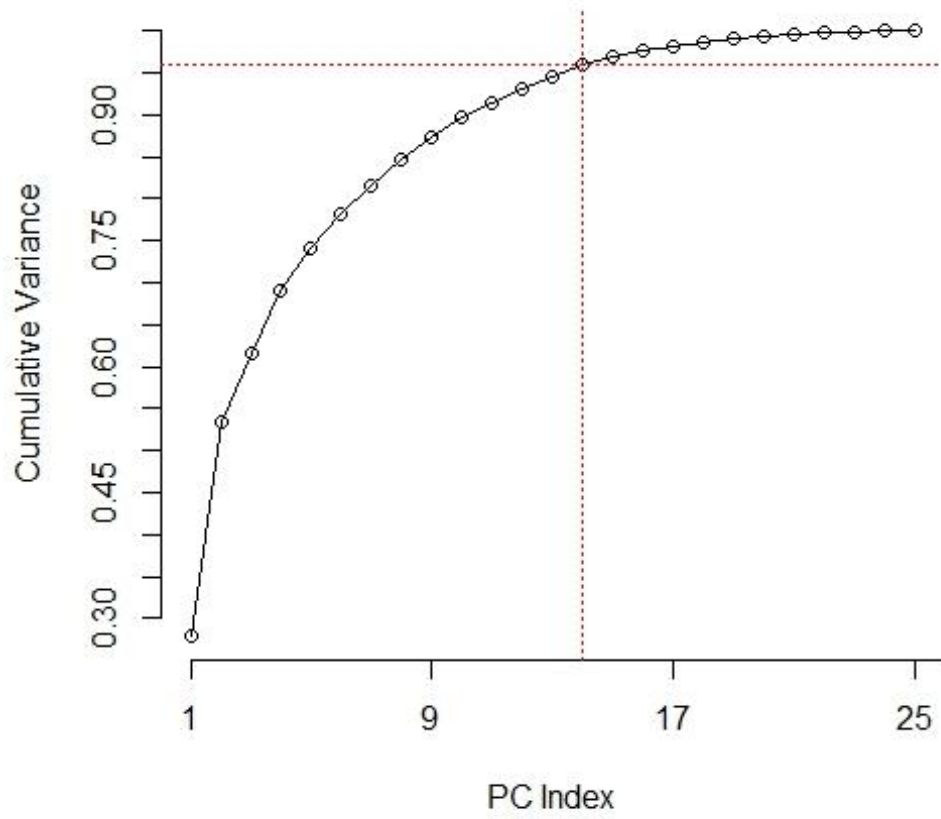| ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14
7053900	−0.05	0.14	0.34	−0.42	0.21	−0.10	0.14	−0.05	0.14	0.23	0.02	0.08	−0.29	0.56
7053901	0.27	0.00	0.03	−0.15	0.05	0.26	0.04	−0.38	−0.25	0.14	0.48	−0.35	0.26	0.10
7053902	0.06	0.12	−0.10	−0.22	−0.30	−0.63	0.22	0.15	−0.52	−0.10	0.12	0.08	0.05	−0.05
7053903	0.14	0.08	0.34	−0.29	−0.26	−0.05	0.10	0.33	0.46	−0.16	0.24	−0.15	0.39	−0.12
7054800	0.23	0.05	0.28	0.19	−0.22	−0.12	0.23	−0.13	0.01	−0.17	−0.47	−0.53	−0.25	0.10
7055200	0.17	−0.30	0.10	0.03	0.22	−0.05	0.11	0.02	−0.04	0.25	−0.05	0.11	−0.08	−0.28
7055201	−0.04	−0.35	−0.10	−0.09	0.07	−0.14	0.21	0.00	0.25	−0.01	−0.05	0.08	0.20	0.06
7056500	−0.14	−0.27	−0.13	−0.04	−0.22	−0.03	−0.42	0.09	−0.14	−0.14	0.11	−0.16	−0.02	0.41
7058000	0.34	−0.07	−0.15	−0.05	0.07	−0.04	0.11	−0.04	0.00	−0.07	−0.10	−0.05	0.00	−0.18
7059600	−0.28	−0.07	0.25	0.27	0.17	−0.22	0.01	0.12	−0.13	0.13	0.04	−0.11	0.02	−0.03
7059900	0.03	−0.36	0.03	0.06	−0.09	−0.03	−0.10	0.14	−0.08	0.21	0.11	−0.25	−0.10	0.16
7062800	0.25	0.09	−0.23	0.11	0.25	−0.30	−0.25	0.08	0.19	0.05	−0.21	0.05	0.27	0.26
7064200	−0.16	−0.29	0.00	−0.19	0.01	−0.05	0.21	−0.05	0.03	0.32	−0.09	−0.22	−0.04	−0.19
7064201	−0.12	−0.19	0.03	−0.23	0.42	0.21	0.07	0.18	−0.36	−0.39	−0.27	−0.21	0.32	0.04
7064800	0.10	0.09	0.32	−0.44	0.18	−0.05	−0.41	0.00	−0.18	0.03	−0.25	0.13	−0.06	−0.19
7065201	0.25	0.05	0.25	0.30	0.16	−0.23	−0.13	−0.03	−0.15	0.34	0.09	−0.02	0.29	−0.01
7066000	−0.04	−0.24	0.21	−0.04	−0.09	−0.25	−0.46	−0.43	0.19	−0.26	0.03	−0.07	−0.03	−0.24
7066300	−0.33	−0.14	0.01	−0.04	0.11	−0.10	0.10	0.05	0.12	0.02	0.07	−0.05	0.09	−0.12
7066301	0.27	−0.23	−0.08	0.02	−0.02	−0.05	0.11	−0.08	0.11	−0.01	−0.22	0.13	0.25	0.29
7066400	0.22	−0.18	−0.07	−0.10	−0.16	0.16	−0.22	0.55	0.02	0.26	−0.06	−0.12	−0.17	−0.08
7068600	0.23	−0.09	0.18	0.23	0.34	0.00	0.12	0.20	0.02	−0.43	0.35	0.04	−0.31	0.04
7069500	−0.33	0.06	0.14	0.19	0.12	−0.18	−0.03	0.12	0.07	0.02	0.04	−0.16	0.14	0.10
7078100	−0.04	−0.08	0.47	0.23	−0.33	0.32	0.04	0.06	−0.20	0.02	−0.18	0.36	0.25	0.09
7080500	0.18	−0.31	0.14	0.03	0.02	−0.08	0.01	0.05	−0.04	−0.13	0.19	0.28	−0.15	0.04
7080501	−0.06	−0.34	0.03	−0.10	−0.14	−0.05	0.16	−0.26	−0.03	−0.03	−0.02	0.23	0.00	0.11

With the offline hourly models, PCA and historical average data, I pick three

days (from July 5th to July 7th) for the case study of the online adjustment approach.

Four scenarios are compared:

1) offline: use the offline model directly without online calibration.

2) demand: based on the real-time counts, historical counts, and offline models, conduct

the on-line demand adjustment algorithm only for online calibration.

3) supply: based on the real-time speeds, and offline models, conduct the online supply

adjustment algorithm only for online calibration.

4) both: based on the real-time speeds, counts, historical counts, and offline models, conduct the online demand and supply adjustment algorithm for online calibration.

Table 5-3 shows the average counts WMSE and speeds WMSE for all hours of the three days' test. The best WMSE for each hourly model is marked with bold and underline. First, it shows that the counts WMSEs will be larger in the early morning hours. This is because in early morning, observed counts are small such that the denominators in equation (3-1) will be small as well. Therefore, changes in daily traffic pattern will lead to large WMSEs for early mornings. However, speeds WMSEs are small since it is almost free flow everywhere. When it comes to peak demand hours, counts WMSEs become better because the denominators in equation (3-1) are larger; however, speeds WMSEs will be worse due to variations in terms of times and locations of bottlenecks. This indicates that even though there is an offline typical model, the variation of traffic condition during peak hours still makes it difficult to obtain very accurate online models. Second, the "both" scenario is found better than the other scenarios in most of the cases, for both counts and speeds validation. The "demand" scenario will decrease counts MWSEs compared with the "offline" scenario, but it sometimes increases speeds MWSEs during peak demand hours. The "supply" scenario dose a good job in speeds validation, but it is no better for counts validation. These results indicate both demand and supply parameters need to be adjusted, especially during peak demand hours. On average, the proposed online calibration will decrease counts/speeds MWSE by 8.5%/12.7%, compared with the offline model. I choose fixed α and β (both 0.3) for online demand adjustment in this test. Sensitivity analysis

should be conducted in future research to understand whether overall adjustment or local adjustment play a more important role.

Table 5- 3 Average WMSE for different online calibration approaches

	Methods	4 to 5	5 to 6	6 to 7	7 to 8	8 to 9	9 to 10	10 to 11
Average Counts WMSE	offline	0.254	0.210	0.154	0.182	0.140	0.125	0.134
	demand	**0.223**	**0.182**	**0.144**	**0.171**	**0.137**	0.113	**0.121**
	supply	0.253	0.210	0.154	0.183	0.141	0.125	0.134
	both	**0.223**	**0.182**	**0.144**	**0.171**	0.138	**0.112**	**0.121**
Average Speeds WMSE	offline	0.064	0.100	0.165	0.175	0.176	0.178	0.168
	demand	0.064	0.100	0.170	0.181	0.204	0.176	0.173
	supply	**0.046**	**0.065**	**0.157**	0.170	**0.164**	**0.171**	0.162
	both	**0.046**	**0.065**	**0.157**	**0.163**	0.165	**0.171**	**0.161**

Table 5-4 summarizes the average running time of the four scenarios. DTA running time makes up the majority of the computational time. In the "both" scenario, supply parameter adjustment cannot be directly applied after demand adjustment because the queueing diagram will change after demand adjustment. Therefore, a queueing diagram updating procedure will be conducted to estimate the new queueing diagram after demand adjustments. However, in either the "demand" or "supply" scenario, it can directly adjust parameters based on the offline model, the gradients, and deviations. Even though the "both" scenario takes 45 seconds for online DTA modeling, it is still promising for online applications because the time factor (real-world time divided by simulation time) is around 240.

Table 5- 4 Computational time of different approaches

	Average Computational Time (sec)	DTA Iterations
offline	0	0
demand	39	1
supply	38	1
both	45	1

Section 5.3. Medium-scale non-recurrent day case study

With the medium-scale model developed in Section 5.2, I also check the performance of the proposed online calibration approach for non-recurrent transportation scenarios. INRIX, which is one of the largest traffic dataset in the U.S., has traffic incident data for most areas in the State of Maryland. By searching for incidents that happened within the study area during the morning time, I found there was a "vehicle on fire" incident reported on October 18, 2017. The incident happened from 9:02 AM to 9:48 AM at the segment marked with a red triangle in Figure 5-1. The timeline and number of lane closure are shown in Figure 5-5. During the incident, two right lanes of the freeway segment were closed (Figure 5-5).

Figure 5- 5 Timeline of the selected incident

INRIX also provides the traffic congestion map for the incident. Figure 5-6 illustrates that a bottleneck will occur because of the incident from around 9:00 AM, and the traffic will return to normal at around 10:00 AM. In order to test the online calibration approach for this incident, hourly traffic counts data and speeds data on October 18th 2017 have been downloaded from INRIX. The validation counts and speeds sensors are the same for the offline models in Section 5.1.
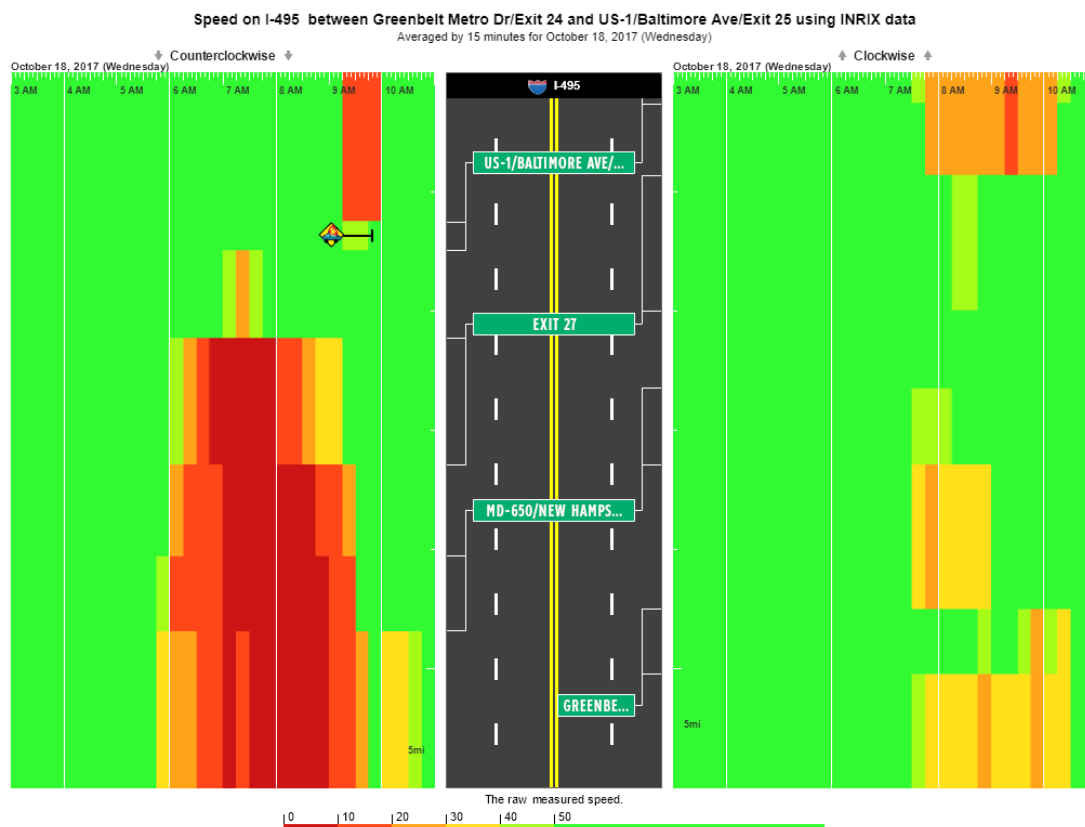
Figure 5- 6 Traffic congestion diagram for the incident

Table 5- 5 WMSE for different online calibration approaches for incident scenario

	Methods	4 to 5	5 to 6	6 to 7	7 to 8	8 to 9	9 to 10	10 to 11
Counts WMSE	offline	0.385	0.253	0.225	0.188	0.155	0.194	0.257
	demand	0.241	0.242	**0.215**	0.176	**0.153**	**0.178**	0.229
	supply	0.385	0.253	0.226	0.188	0.154	**0.178**	0.257
	both	**0.236**	**0.234**	0.218	**0.175**	**0.153**	0.179	**0.187**
Speeds WMSE	offline	0.086	0.102	0.193	0.195	0.178	0.256	0.219
	demand	0.086	0.101	0.181	0.192	0.189	0.296	0.265
	supply	0.066	**0.062**	0.175	**0.162**	0.164	0.226	0.214
	both	**0.065**	**0.062**	**0.169**	0.182	**0.159**	**0.218**	**0.196**

Table 5-5 shows the results for all the 7 hours for October 18th 2017 in terms of WMSE. Still, the four scenarios (i.e. offline model, demand calibration only, supply calibration only, and both demand and supply calibration) are compared. As expected, the WMSE for the online model after 9:00 AM is high because of the incident. The WMSE for counts is high before 7:00 AM because the observed counts are small, which

will make the relative error high. If one only applies demand online calibration, the WMSE for speeds after 9:00 AM will be even higher than without any online calibration (the offline model scenario); and the counts WMSE will increase compared with the offline model for 10:00 to 11:00 AM. This may be because that without considering incidents, the demand adjustment will overlook the impact of capacity decrease due to the incident. Similarly, the performance is poor for scenario "supply".

The results of this non-recurrent day case study indicates there is interaction between capacity drop and travel demand. Under non-recurrent days, both supply and demand parameters should be calibrated to reflect the capacity drop as well as its impact on drivers travel decisions.

Section 5.4. Large-scale network example
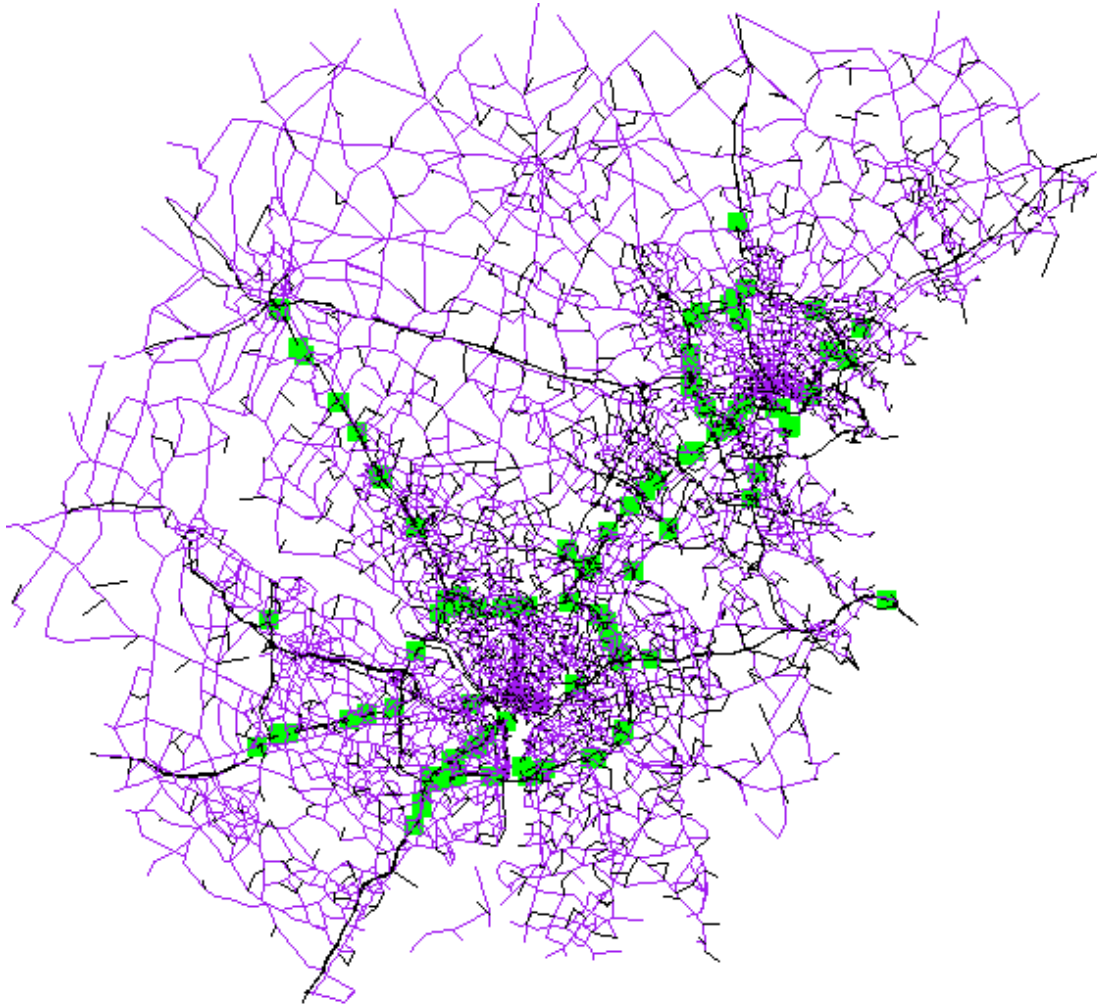
Figure 5- 7 Large-scale online system model

This section will show the on-line simulation model for the large-scale network in Section 3.6. The offline calibration has been discussed in Chapter 3, in which the counts data is obtained from Maryland SHA I-TMS. Since I-TMS does not provide real-time traffic data, I use RITIS data obtained from CATT lab's real-time traffic data API for both counts and speeds calibration. Since the data source has changed, a recalibration process is required before testing online calibration. I still consider 4:00 to 11:00 AM as the model horizon. 120 sensors from the CATT lab's API are selected (Figure 5-7), and each sensor provides both counts and speeds data. Traffic

data from December 20th 2017 to January 25th 2018 has been collected as historical data to recalibrate the offline models. The recalibration of the offline models takes around two weeks to finish. The WMSE for the offline calibration is shown in Table 5-6. Compared with the medium-scale network, the WMSE for both counts and speeds trend to be higher because it is more difficult to calibrate large-scale networks. I still choose

Table 5- 6 WMSE for offline calibration

Model	Counts WMSE	Speeds WMSE
4:00 to 5:00 AM	13.24	11.67
5:00 to 6:00 AM	12.90	11.53
6:00 to 7:00 AM	13.40	13.85
7:00 to 8:00 AM	12.62	16.42
8:00 to 9:00 AM	13.54	17.22
9:00 to 10:00 AM	13.90	15.86
10:00 to 11:00 AM	12.01	17.37

With these offline hourly models, PCA and historical average data, the I test three days (from January 5th to January 7th) for the case study of the online calibration approach. I also consider the four scenarios discussed in Section 5.2: offline, demand, supply, and both. Table 5-7 shows the average counts WMSE and speeds WMSE for all hours of the three days' test. The best WMSE for each hourly model is marked with bold and underline. Compared with Table 5-3, the improvement in terms of WMSE is less significant, especially during non-peak hours (4:00 to 6:00 AM, and 9:00 to 11:00 AM). This may be because large-scale networks have far more traffic sensors than smaller networks, so that the temporal and spatial correlations among the sensors are

more complex. As a result, online adjustments can hardly satisfy the matching of every

traffic sensor. This also reflects the difficulties of calibrating large-scale DTA models,

because the offline calibration can be a multiple solution problem; even though the

offline calibrated model matches real-world counts and speeds, it could not have

"correct" ground truth demand and supply model parameters to capture real-time

change of traffic conditions. The "both" still provides the lowest WMSEs in most of

the cases. Similar with the results in Section 5.2, the "demand" scenario will decrease

counts MWSEs compared with the "offline" scenario, but it sometimes increases speeds

MWSEs during peak demand hours; while, the "supply" scenario dose a good job in

speeds validation, but it is no better for counts validation. To summarize, the proposed

online calibration (the both scenario) will decrease counts/speeds MWSE by

14.0%/9.8%, compared with the offline model.

Table 5- 7 Average WMSE for different online calibration approaches

	Methods	4 to 5	5 to 6	6 to 7	7 to 8	8 to 9	9 to 10	10 to 11
Average Counts WMSE	offline	0.191	0.240	0.251	0.264	0.249	0.243	0.162
	demand	**0.188**	0.203	**0.188**	**0.184**	**0.196**	**0.239**	**0.157**
	supply	0.191	0.240	0.250	0.264	0.249	0.243	0.161
	both	0.189	**0.202**	**0.188**	**0.184**	**0.196**	**0.239**	**0.157**
Average Speeds WMSE	offline	0.124	0.119	0.156	0.230	0.274	0.225	0.150
	demand	0.124	0.119	0.158	0.232	0.273	0.225	0.150
	supply	0.123	**0.113**	0.151	**0.218**	0.223	**0.181**	**0.135**
	both	**0.122**	0.114	**0.150**	0.219	**0.222**	0.182	**0.135**

The average running time of the four scenarios of online calibrating this

large-scale DTA model is shown in Table 5-8. Compared with the medium-scale

network case, the additional time required for the "both" scenario is much lower. This

is because as the network size and demand increase, running DTA is dominating the total online simulation time. The time factor for this large-scale online system model is around 9.3.

Table 5- 8 Computational time of different approaches

	Average Computational Time (minutes)	DTA Iterations
offline	0	0
demand	18.4	1
supply	18.1	1
both	19.3	1

Section 5.5. Real-time decision support with the large-scale network model

5.5.1. Background

This section discusses the application of the proposed online transportation network model on real-time decision support. ICM and ATM are among the most effective approaches to corridor traffic operations and planning. These strategies, implemented individually or jointly, could provide additional effective capacity to peak-hour users, improve the detection and response to incidents and other adverse conditions, produce smoother traffic flows, and achieve more efficient use of multimodal corridor travel alternatives, and consequently enhance transportation system performance in safety, efficiency, reliability, and sustainability. Successful implementations ICM and ATM strategies can be found worldwide and in the U.S. Some famous strategies include:

1) ramp meters. Ramp meters are traffic signals placed at freeway entrance ramps to

control the flow rates at which vehicles enter the freeway mainline traffic stream. Almost all empirical evaluation studies have shown that ramp metering has been very effective in enhancing freeway speed, throughput, and safety. At times, ramp meters may cause certain drivers to wait long periods of time before granted freeway mainline access. This problem can be addressed in the design of metering algorithms.

2) managed lane. Managed lane facilities include but are not limited to high occupancy vehicle, high occupancy toll, transit only lanes, express lanes, and reversible lanes. They are widely seen in a pre-time manner with the exception of dynamically priced managed lanes. U.S. metro areas with successful implementation of managed lanes are too many to list.

Other important applications of ICM and ATM include junction control (e.g. applications in Washington State), speed harmonization (e.g. variable speed limit applications in New Jersey's NJ Turnpike), and corridor travel demand management (e.g. real-time ride sharing, early-warning traveler information system for flexible-schedule commuters). Developing appropriate modeling tools with the capability of analyzing traffic pattern, travel demand, and traveling/driving behavior along major corridors is the prerequisite to successfully testing, evaluating, and eventually implementing effective traffic operations and planning strategies. One important purpose of this dissertation is to apply the system model to model different ICM and ATM for large-scale network on the real-time basis.

5.5.2. Scenarios design of real-world large-scale application

Real-time evaluation of ATM and ICM strategies are still missing in current modeling systems. Utilizing the proposed online AgBM-DTA model, I test these traffic management strategies under real-time scenarios. The integration between AgBM-DTA and real-time traffic management decisions can be illustrated in Figure 5-8. The real-time data source provides online information to calibrate the system model; real time data can also be considered by traffic management agencies to design congestion mitigation strategies. These strategies will be simulated via the calibrated system model based on which the final decision can be made. Finally, the traffic management decisions will affect travelers' behavior and traffic conditions which will be reflected in the real-time data.
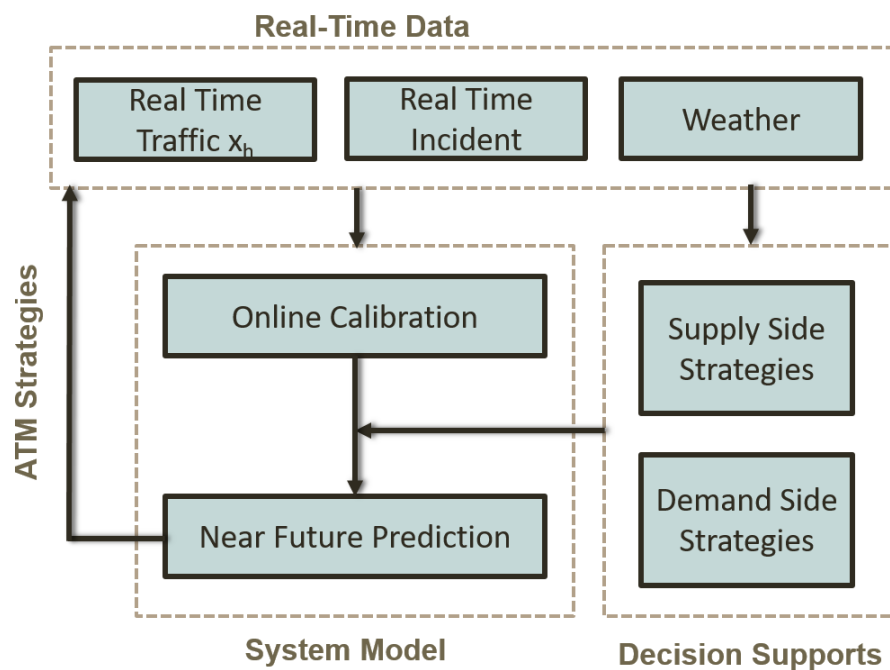
Figure 5- 8 System model integration for real-time decision supports

From both historical traffic data and the online large-scale system model presented in Section 5.4, we notice that I-495 westbound (WB) has recurrent congestion during morning commuting time, especially the freeway segments between I-95 and MD 185 (Figure 5-9). In order to provide decision supports to transportation operation agencies to mitigate traffic congestions, I have tested four ATM scenarios via the online system model:

1) base scenario. Calibrate the model online for 7:00 to 8:00 AM, and use the calibrated model to predict traffic condition for 8:00 to 9:00 AM.

2) ramp metering scenario. Calibrate the model online for 7:00 to 8:00 AM, implement ramp meter on the calibrated model, and predict traffic condition for 8:00 to 9:00 AM.

3) lane addition scenario. Calibrate the model online for 7:00 to 8:00 AM, consider to open hard shoulder of congested freeways to increase capacity. Simulate the traffic condition for 8:00 to 9:00 AM.

4) both scenario. Calibrate the model online for 7:00 to 8:00 AM, jointly consider ramp metering and hard shoulder running to predict traffic condition for 8:00 to 9:00 AM.
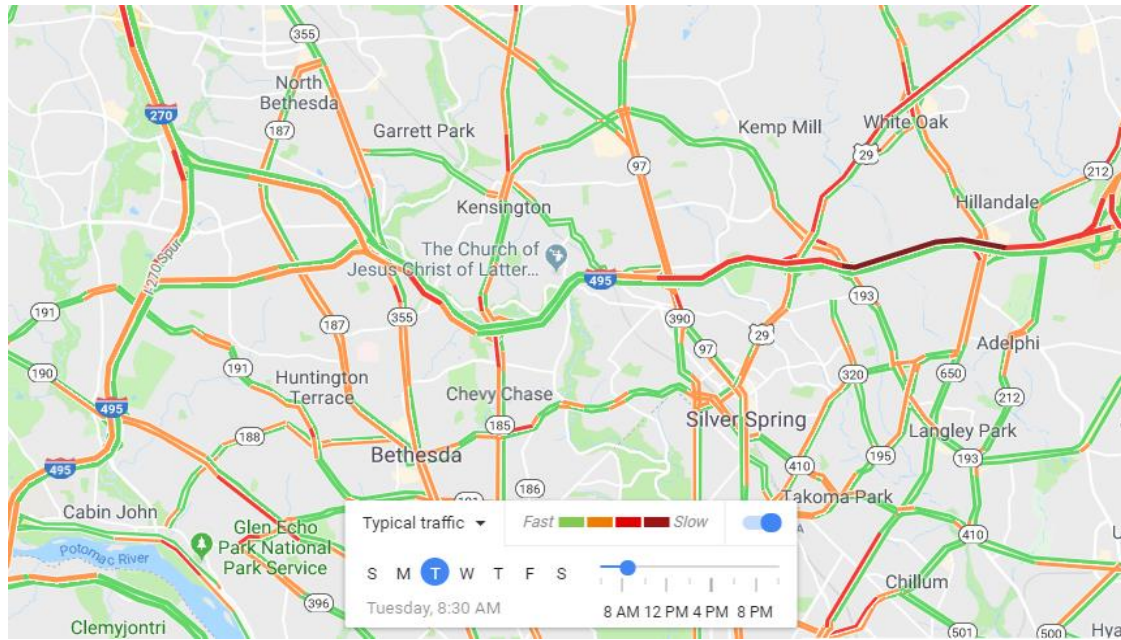
Figure 5- 9 I-495 westbound congestion pattern based on Google Map

The online calibration model for January 6th from 7:00 to 8:00 AM is used

for this real-time decision support application. I choose this specific modeling period

because the WMSEs for counts and speeds for 7:00 to 8:00 AM are 0.190 and 0.154,

respectively, which are lower compared with the other days and modeling periods.

Based on the historical offline models and the online calibration approach, I first use

the model to predict traffic conditions for 8:00 to 9:00 AM. Based on the prediction,

one is able to identify potential major bottlenecks. Figure 5-10 illustrates the predicted

traffic condition at 8:30 AM on January 6th, 2018 based on the 7:00 to 8:00 AM model.

It indicates the traffic on I-495 WB will get congested from MD 355 to MD 97. Based

on this prediction, I decide to test ATM on the congested segments. In "ramp metering"

scenario, the on-ramps to I-495 WB from US-29 to MD 285 will be metered. I use a

fixed meter rate for simplification, and the capacity for all the associated ramps are

decreased by 50% to model ramp metering. In "lane addition" scenario, I assume the

hard shoulder (left emergency lane) are open to improve traffic mobility on I-495 westbound. In order to model this lane addition scenario, I increase the number of lanes by 1 (from 4 to 5) for freeway mainline segments between the interchange with US-29 and the interchange with MD 355. The "both" scenario will be the combination of the ramp metering scenario and the hard shoulder scenario.
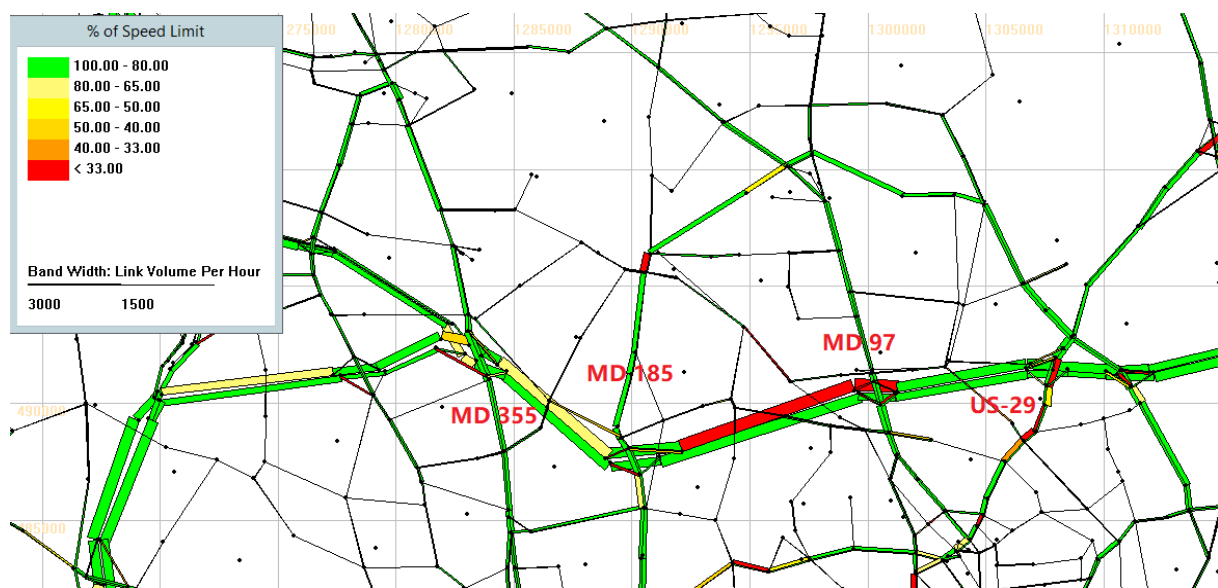
Figure 5- 10 Prediction based on calibrated model

5.5.3. Results of the online ATM application

The goal of this large-scale online ATM application is to provide transportation managers real-time decision supports on what strategies can be used to address near future traffic congestions. The benefit of using a DTA model is that it has a visualizable traffic network and can summarize both system level and corridor level traffic performance.

Table 5-9 illustrates the overall performance of different scenarios. Measures

such as system average travel time, average travel time index (mean travel time over free-flow travel time), average speed, and total delay are summarized. The red numbers is in bracket in Table 5-9 denote the total travel time saving (negative) or losing (positive) due to the implementation of the ATM. An interesting finding is that ramp metering will make the overall system more congested. With some 1.5 million demand, ramp metering will increase average travel time by 0.05 minutes, leading to additional 1180 hours of traffic delay. While, open the hard shoulder lane is shown to be an efficient way to mitigate congestion, such that the total delay will decrease by 260 hours. The combination of ramp metering and lane addition does not make any improvement due to the significant negative side effects of ramp metering.

Table 5- 9 Overall performance of different scenarios

Scenario	Avg. Travel Time (min)	Avg. Trip Time Index	Avg. Speed (mph)	Total Delay (K hr)
base	22.37	1.150	37.58	73.66
ramp metering	22.42	1.153	37.50	74.84
				(+1.18)
lane addition	22.36	1.150	37.59	73.40
				(-0.26)
both	22.42	1.152	37.50	74.71
				(+1.05)

In order to understand why ramp metering leads to more congestion, I plot both mainline corridor travel time (Figure 5-11) and ramp travel time (Figure 5-12). The travel time on freeway mainly has decreased significantly under all the three ATM strategies. This is reasonable because ramp metering will restrict on-ramp volume, lane addition will increase mainline capacity, and the combination of these two are further

improving the mobility of freeway mainline. While, since ramp metering is restricting on-ramp volume, it can bring in new bottlenecks at the entrance of freeway mainlines. The queue may spillback to major arterials and cause more congestion on local streets.
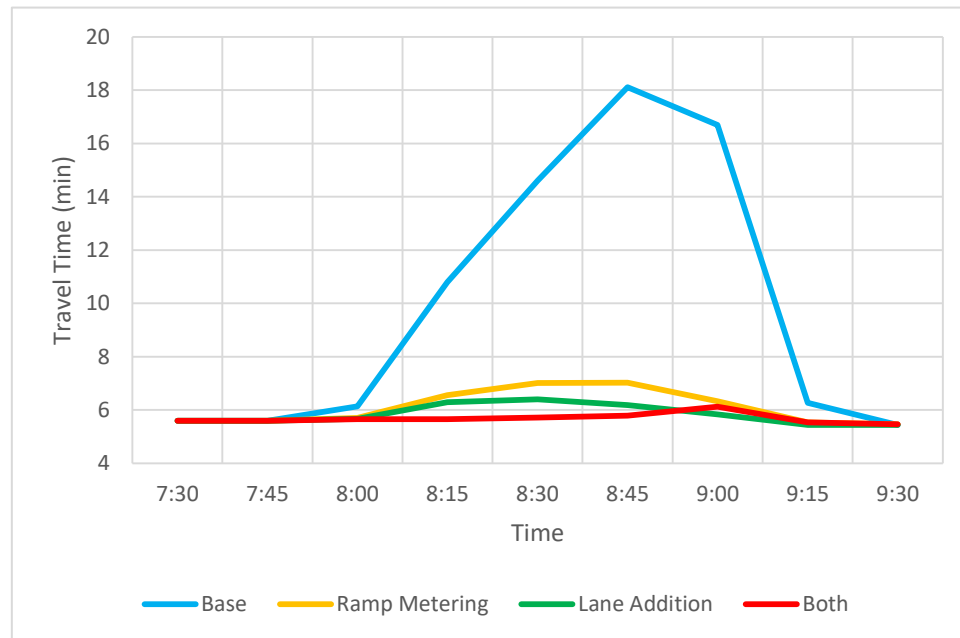
Figure 5- 11 I-495 WB travel time

In Figure 5-12, I give an example of ramp travel time under the four scenarios. I pick the on-ramp from MD 185 southbound (SB) to I-495 WB, and the other four on-ramps have similar travel time patterns. The travel time trends of "base" scenario and "lane addition" scenario are overlapped, so as to "ramp metering" scenario and "both" scenario. The fact that "ramp metering" scenario and "both" scenario will cause the same delay to on-ramp trips indicates there is a high demand from local streets to WB I-495 during morning peak hours. In such cases, ramp metering may not be as good as other ATMs because the remission of freeway congestion is based on the loss of travelers who would like to enter freeways at congested ramps.
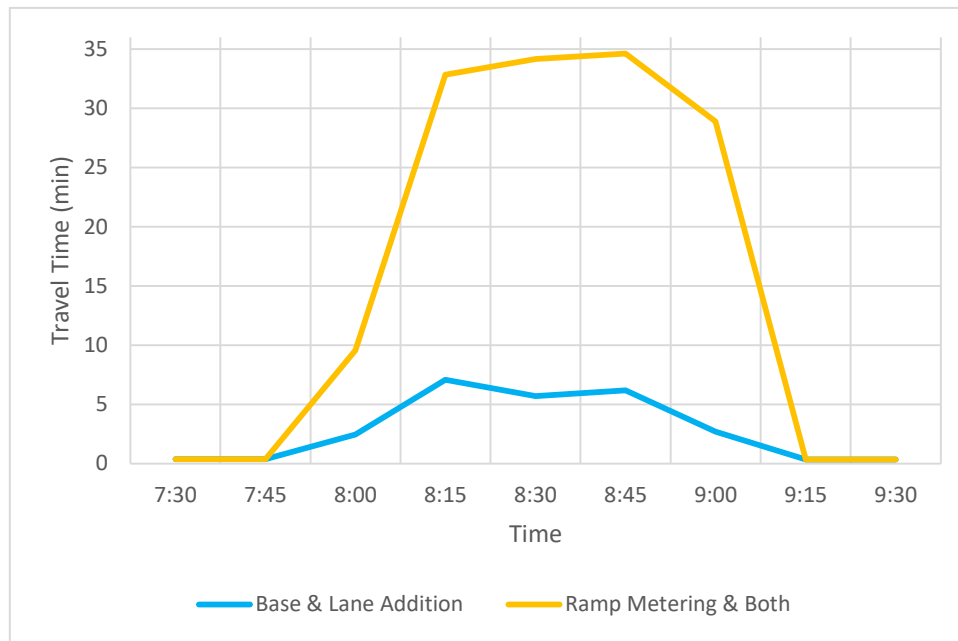
Figure 5- 12 On-ramp travel time (MD 185 SB to I-495 WB)

Based on the results of the online ATM evaluation, traffic managers can obtain valuable empirical views on what strategies can be applied for congestion mitigation. Since the time factor is around 10, real-time evaluation can be achieved by letting four computers to run these scenarios. I should admit that one weakness of this system model is that it has to conduct path fixed traffic simulation for different ATMs; therefore, the impact of ATMs on route choices are ignored. This issue can be solved given more advance traffic simulation methods or computational technologies.

Section 5.6. Summary

This chapter talks mainly about the application of the proposed online calibration approach. The online transportation network model has been tested on a medium-scale network with several freeway interchanges, the results indicates a time factor of 240 for real-time applications. I also test the performance under a non-

recurrent congested network. A large-scale network that covers the whole Washington D.C. area has also been tested. By calibrating hourly modeling horizon, the time factor can reach around 10, which is enough for real-time applications. A case study on using the online large-scale model to evaluate different ATM scenario has also been conducted, demonstrating the value of this modeling tool for real-time decision supports.

Chapter 6: Large-Scale Real World Application in Work Zone Scheduling

The developed system model is ready to be integrated with other models for transportation planning and operation applications. With the framework mentioned in Chapter 3, it will be interesting to apply the system model for more complex design and operation scenarios. This chapter aims at integrating the modeling system developed with another transportation planning related issue—work zone scheduling problem.

Section 6.1. Background of work zone scheduling research in Maryland

Road transportation has been a national travel mode for the United States: passenger transportation is dominated by a network of over 3.9 million miles of highways (DOT, 1997); around 60% of US freight is carried by trucks (DOT 2006). The growth on economic and people's living standard dependent highly on the performance of existing surface transportation systems. However, since the majority of national highway system has been completed for decades, a large fraction of today's road transportation infrastructure has reached the end of the design life. There are numerous direct negative effects due to the deterioration of roadways, e.g. damage to vehicle wear-and-tears, higher fuel consumption, more travel time, and travel safety. Therefore, state and federal transportation agencies have shifted their focus from constructing new roadways to maintain the existing highway infrastructure to maximize the performance of national highway systems. Highway maintenance and reconstruction activities are likely to increase in number, duration, and scope in the near future. The budget from Federal Highway Administration (FHWA) for highway

construction and maintenance increased from 34 million dollars to 48 million dollars from 2005 to 2015 (FHWA 2015). The budget merely included highway systems, and there were many local streets under reconstruction, too.

The maintenance of roadways usually need to close segments of lanes and shoulder lanes to form work zones. Drivers' travel experience can be greatly affected by work zones. The reduction of capacity because of lane closure may cause severe congestions. Since the lanes remaining open will be narrower due to the barriers, traffic safety can also be affected. For a single maintaining roadway, the design of characteristics of work zone, such as lane closure configuration, working schedule, and traffic control strategies can significantly reduce traffic congestion, improve safety, and increase maintenance efficiency. For long-term maintenance projects that will reconstruct a number of roadways, the maintenance order can greatly affect the overall projects efficiency and traffic impacts. Decision makers need to find an appropriate design that balances agency cost and user cost. Therefore, it is worthwhile to develop appropriate work zone analysis methods which can aid highway agencies in developing cost-effective highway maintenance or reconstruction plans.

The Maryland research team has been working on a work zone scheduling research project. This research project is performed under the Implementation Assistance Program (IAP) of the second Strategic Highway Research Program (SHRP2). The IAP is developed to help State and local departments of transportation, metropolitan planning organizations, and others interested in deploying products and solutions researched under SHRP2. The SHRP2 product pilot tested through this

initiative was the work zone impacts and strategies estimator (WISE) Software. The intent of the WISE software tool is to help agencies assess the optimal sequencing of renewal projects and help determine the cost-effectiveness of strategies to minimize, manage, and mitigate road user costs from safety or operational perspectives. A team comprised of the Maryland SHA and researchers from the UMD has been awarded this project under the SHRP2 IAP to conduct this WISE software tool implementation research. As the project manager from the UMD team, this work zone scheduling research is one of the major parts in the applications of my dissertation.

Work zones are a primary cause of unexpected delays and one of few types of traffic incidents that can actually be controlled by transportation agencies. The underlying notion of the SHRP2 research on WISE is that planning and scheduling of work zones can mitigate some of these delays, and also improve the travel reliability and safety of the overall transportation system. To help transportation agencies with this task, the WISE tool was created to support work zone planning and scheduling decisions at the regional level. The WISE tool currently relies on existing traffic data and DynusT (Chiu et al. 2011) to evaluate impact of different projects on traffic delays. The WISE tool is primarily intended for planning projects that impact traffic for at least a few weeks, while the expected benefits from its application are reduced negative mobility, safety, and economic impacts of highway renewal activities. In terms of methodological contributions, this project as allowed the UMD team to extend WISE capabilities by introducing travel behavior models that consider possible modal shift and peak spreading effects due to work zones. Also, the UMD team has successfully

demonstrated the feasibility of replacing DynusT with AgBM-DTA (proposed in Chapter 3). The most significant outcome of this research is the production of an application-ready, integrated transportation model and work zone schedule optimization tool that can help schedule work zone projects on a planning horizon of 1-5 years; and in the future, work zone operations coordination in real-time based on the ongoing development of this tool for real-time modeling applications.

Section 6.2. Work zone schedule and management with system model

As noted previously, the WISE software tool was developed to be a decision support system used by planners and engineers to evaluate the impact of work zones and determine strategies to reduce these impacts (Ismart et al. 2014). WISE can evaluate the impact of multiple projects that have been incorporated into the transportation improvement program and the network. The software package has a relatively user friendly graphical user interface (GUI) to link different modules together (e.g., WISE planning module, WISE operation module, and DynusT). The original WISE tool utilizes the mesoscopic traffic simulator DynusT to evaluate the impact of work zone projects and operations applications. However, Maryland SHA has been looking for ways to incorporate this work zone scheduling tool into Maryland transportation modeling system. The major contribution towards this project is to integrate the WISE work zone scheduling algorithm within AgBM-DTA to provide a more comprehensive tool for use by Maryland SHA.

6.2.1 Model integration

This section provides additional detail on the integration of WISE and SILK AgBM-DTA. Using WISE source code obtained for this research project, I have modified and extended this code to work with AgBM-DTA. The WISE work zone schedule optimization module utilizes a meta-heuristic algorithm to search for the optimal schedule to minimize total cost. The total project cost contains both agency cost from construction and user cost from delay. A Tabu search algorithm is applied to find the optimal sequence of projects with defined starting time and construction mode (daytime, nighttime, or both). In the original WISE software, WISE reads an initial traffic assignment result from DynusT, and then searches for an optimal schedule based on its meta-heuristics algorithm. For each project and each month, if construction is feasible (mode and start time), the work zone schedule optimization module estimates the traffic performance after travelers detour due to work zones. If the result reduces total cost, WISE schedules this project and updates the current solution. The algorithm stops if a predefined maximum iteration number is reached, or if in the most recent five continuous iterations, the algorithm does not find a solution with improved objective function value. Once the final schedule is obtained, a user can run DynusT again for more accurate evaluation and operational use.
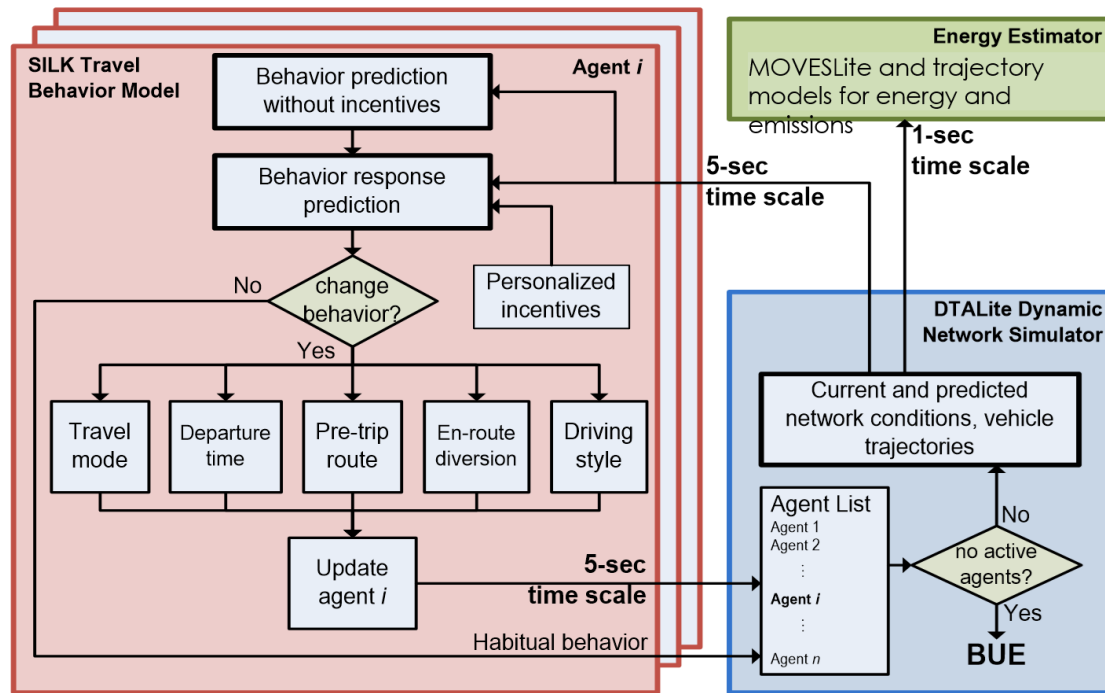
Figure 6- 1 AgBM-DTA integration flowchart

Agent-based travel behavioral models have the capability of mimicking and simulating the travel behavior changes of each user in the system. In order to enhance the capability of traditional travel demand modeling, this modeling tool has been efficiently interfaced with a dynamic traffic assignment model, such as DTALite. Once integrated with a traffic simulator, the system can be complete, given that all traffic conditions in the transportation network can be imitated by the simulator. This motivates the proposed integration of agent-based models and the DTA simulator, as illustrated by the flowchart above. There are two levels of integration (Figure 6-1):

1) between-day integration: on one particular simulation day, agents are able to acquire information from previous days and accumulate knowledge about the transportation system. For instance, when an autonomous vehicle is introduced to a household in a future year, members of the household will respond and rearrange their trips. Seniors

and juveniles who previously relied on non-auto modes now may consider riding in the vehicle. Working adults may need to readjust departure time to accommodate the foreseeable increase in vehicle usage. These changes to each agent are modeled and outcomes are fed into DTALite to simulate dynamic traffic conditions, based on which agents will adapt their behaviors again.

2) within-day integration: within the same day, information is conveyed between AgBM and DTALite. Real-time information on congestion and different non-recurrent incidents has been made available to a certain percentage of agents, which reflects the fact that advanced traffic information system (ATIS) subscribers and Google/INRIX users have access to timely estimates of traffic congestion. This type of information exchange would trigger dynamic behavior adaptation. En-route diversion is a likely reaction and is incorporated in this integration. Future studies may also internalize dynamic modal shifts (park-and-ride options along major freeways, ridesharing, etc.).

The between-day integration allows the system model to capture travelers' long-term travel behavior shifts due to the capacity reduction from work zone projects; within-day integration can be used to evaluate different operation strategies for work zone managements. In the dissertation, my contribution of the AgBM-DTA is to integrate the between-day departure time choice model with DTALite (as discussed in Chapter 3). The rest AgBM-DTA integration work was done by my colleagues, and my job in this work zone scheduling work is to integrate WISE scheduling algorithm into this AgBM-DTA system model.
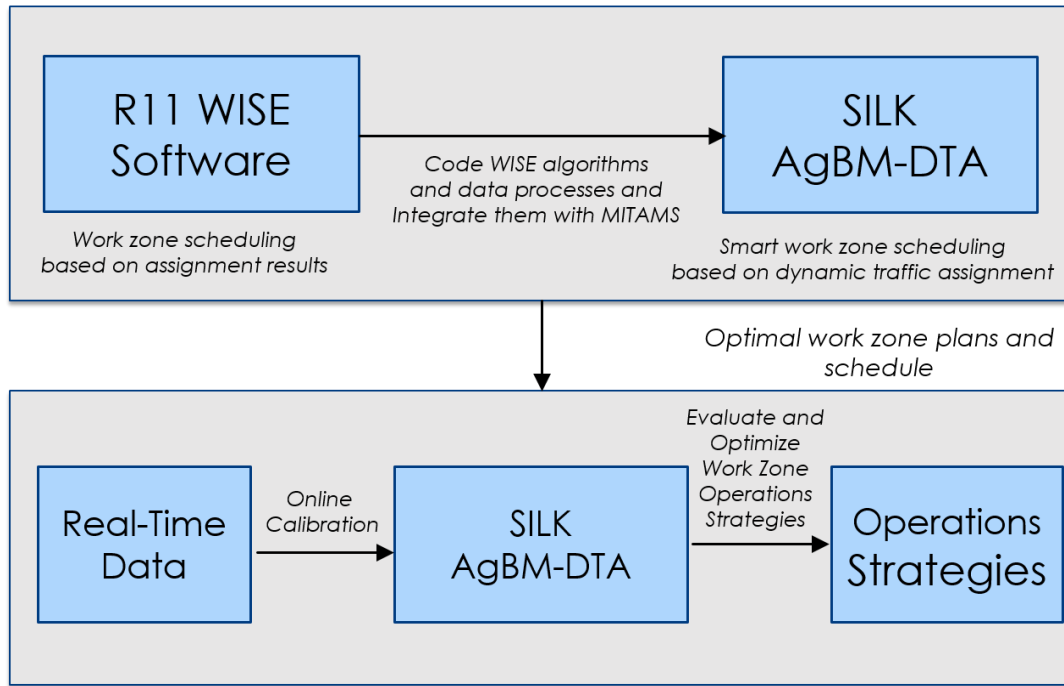
Figure 6- 2 Integrating WISE with SILK AgBM-DTA

The flowchart in Figure 6-2 shows the integration between WISE and SILK AgBM-DTA model. The upper level denotes the functionality for work zone schedule optimization; the lower level denotes the work zone operation applications. In the work zone schedule optimization horizon, the original WISE algorithm is coded to interface with DTALite. That is, the data structure of work zone projects, planning characteristics, and strategies are coded within DTALite's format. The outputs of AgBM-DTA can be directly imported to the WISE algorithm for schedule optimization. After the upper level modeling, the optimal schedule will be coded into a real-time system model under development. With real-time detected data and online calibration technics, AgBM-DTA is able to simulate and evaluate the performance of different operation strategies (e.g., variable speed limit, ramp metering, etc.) to improve traffic conditions under work zone scenarios. Based on the scope of work, the upper level integration has been finished by the UMD team; the lower level, with the help of the proposed online calibration

approach in Chapter 4, is also feasible in future work.
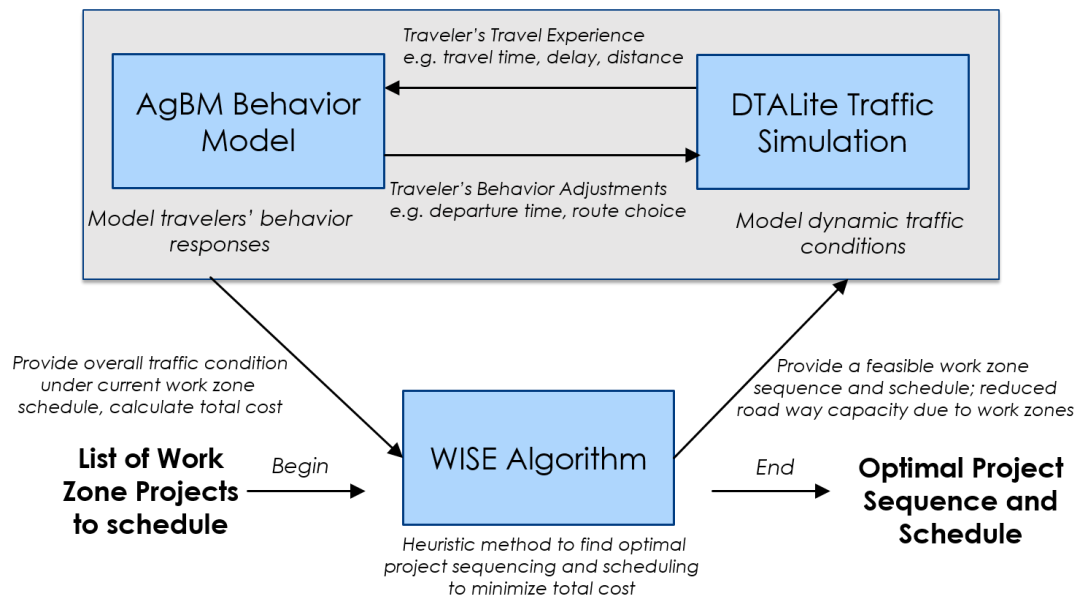
Figure 6- 3 Flowchart of work zone schedule optimization

Details about the upper level integration are illustrated in Figure 6-3. First, the AgBM behavior model is integrated with DTALite to simulate between-day and within-day travelers' behavior shifts and the corresponding traffic conditions. This AgBM-DTA system model will replace the previous WISE DynusT for traffic performance estimation. The work zone schedule optimization begins with a list of work zone projects with their schedule constraints. Then, the AgBM-DTA system model will be run for a general day base scenario without any work zone projects. WISE will read the assignment results and use its own heuristic algorithm to estimate the cost of different combinations of schedules. Once WISE has finished running, the optimal schedule will be obtained, and AgBM-DTA is conducted again to accurately estimate the user costs under the WISE optimal work zone schedule.

6.2.2. Inputs & outputs for the integrated model

The AgBM-DTA model attempts to simulate the interaction between travelers and traffic infrastructures. It has two types of inputs: demand-side inputs and supply side inputs. Demand side inputs are the information of travelers to be modeled in the system. For agent-based modeling, we need the trip origin, destination, departure time, preferred arrival time, and social demographical information. DTALite can convert OD matrices to an agent list with trip origin, destination, and departure time through traffic assignment. Travelers' social demographical information and preferred arrival time can be generated through a population generation algorithm developed from previous work, such that, to integrate with WISE, the research team only needs to obtain reliable OD matrices as demand-side input. Supply-side inputs are the transportation network information (e.g. nodes, links, and zones). Node information covers the coordinates and signal control types; link information comprises link attributes such as capacity, free flow speed, jam density, etc.; zone information is used to load demand from OD matrices to the network.

The outputs of AgBM-DTA are comprehensive, spanning from network level to link level, and then to individual traveler level. Since DTALite is a mesoscopic traffic simulator, it records detailed information for each traveler (e.g., origin, destination, departure time, demand type, arrival time, trajectory). For link-level outputs, AgBM-DTA can restore time-dependent link performance in terms of volume, speed, and density, which can be used to evaluate the impact of work zone projects on certain roadways. Network level outputs provide the total delay, average travel time, and total

number of travelers. Network level model outputs is useful to evaluate overall traffic performance on work zone scenarios.

The WISE work zone schedule module requires four kinds of inputs: planning inputs, project inputs, strategy inputs, and traffic assignment inputs. Planning inputs are used to specify the entire time horizon for scheduling all the work zone projects. Each month needs a preference (i.e., whether to allow construction for that month) and demand factor. Project inputs involve the information of all the work zone projects to be scheduled. Information includes location (link ID, from/to nodes), construction cost, construction duration, earliest start month, latest finish month, and capacity/speed limit reduction. WISE allows users to specify some predefined work zone operation strategies to either save construction time or decrease travel demand. Strategy inputs are used to provide such operation strategies to WISE. Finally, before running the scheduling algorithm, WISE needs to read traffic assignment results from DTALite as the without work zone scenario. Link volume and speed information are used in WISE to estimate the additional user cost due to work zone construction. The output of WISE is simple, which provides the optimized starting month for each work zone project.

Section 6.3. Integrated model application

6.3.1. Future planned work zone projects in the consolidated transportation program (CTP)

Information regarding future planned work zone projects in the Consolidated

Transportation Program (CTP) was obtained from Maryland SHA. The location of the projects are shown in Figure 6-4. The circles show the location of minor projects and the triangles show the location of major ones. There are 226 minor projects and 83 major projects in the CTP. In our case study 10 of these projects have been considered for scheduling optimization.
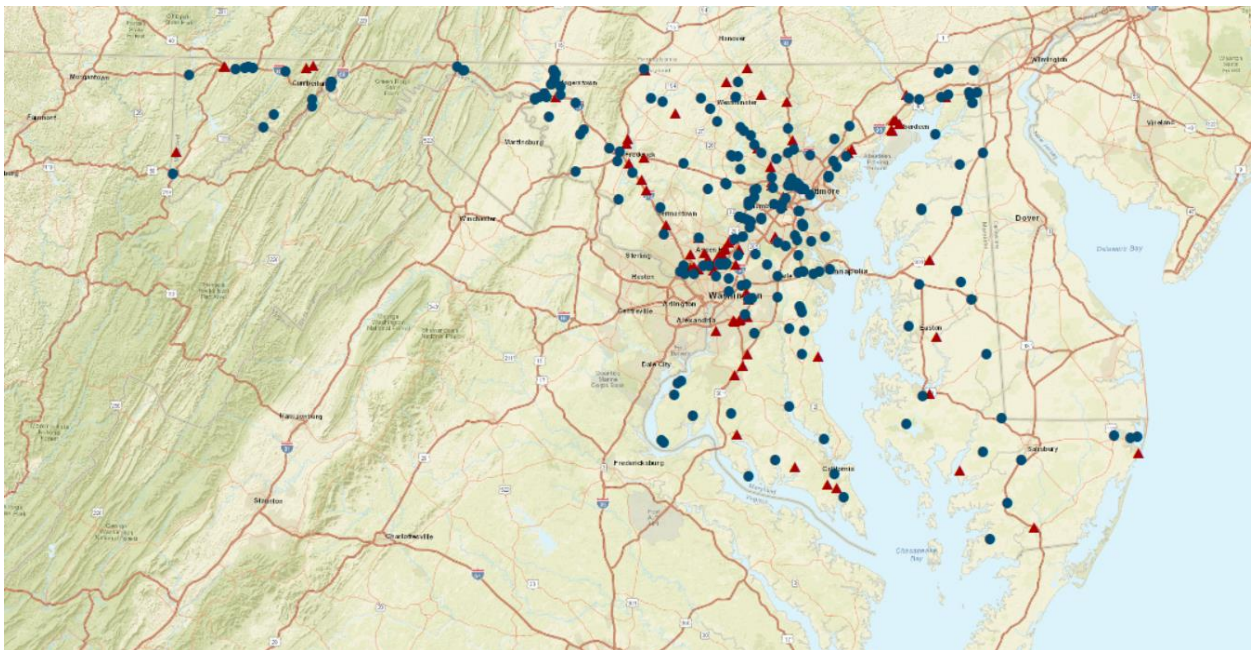
Figure 6- 4 Location of major and minor projects in CTP

6.3.2. Work zone schedule application

In this large-scale application, the system model presented in Section 3.5 is integrated with WISE work zone scheduling algorithm. 10 major work zone projects planned in the near future are considered. These projects are obtained and selected from the CTP dataset, the locations of which are displayed in Figure 6-5. Table 2 summarizes descriptions of these projects. Since there is no detailed information on the number of lane closures or speed limits for these projects, the UMD team assumes the capacity

and speed limit will drop 50% during construction periods. The construction duration

of each project is obtained in the CTP dataset and shown in Table 2.

Figure 6- 5 Locations of the projects in the real-world application

Table 6- 1 Future work zone projects in study area

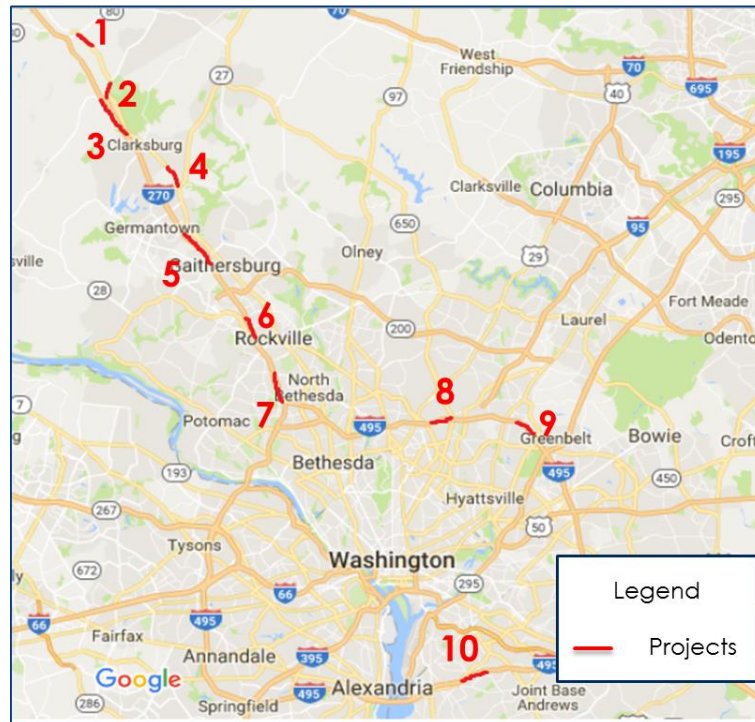| ID	Route	Description	Duration (Months)
1	MD 355	Replace bridge 10086 over Bennett Creek.	12
2	MD 355	Replace bridge 15053 over Little Bennett Creek.	12
3	I-270	Resurface/rehabilitate	24
4	MD 355	Intersection capacity improvements	6
		Construct a new I-270 interchange at Watkins Mill Road.	36
		Bicycle and pedestrian improvements will be included where	
5	I-270	appropriate.	
6	I-270	Traffic management	24
7	I-270	Safety/spot improvement	24
8	I-495	Replace bridge 15136 over I-495.	36
	I-495	Construct a full interchange along I-95/I-495 at the Greenbelt	6
9		Metro Station.	
10	I-495	Phase 2 Access improvements from MD 5 (Branch Avenue)	12
		and I-95/I-495 to the Branch Avenue Metro Station.	

The planning horizon for the real-world application is from the beginning of 2017 to the end of 2020, during which we assume no construction can be undertaken in January or December. For comparison purposes, two scenarios are considered: 1) a naïve case that randomly decides the starting time for each project and; 2) a WISE case that utilizes WISE to obtain the optimal schedule. In the naïve case, the performance of the schedule will be evaluated via AgBM-DTA directly; in the WISE case, it follows the flowchart in Figure 6-3 to reach WISE optimal first, and then utilizes AgBM-DTA to estimate the traffic performance.
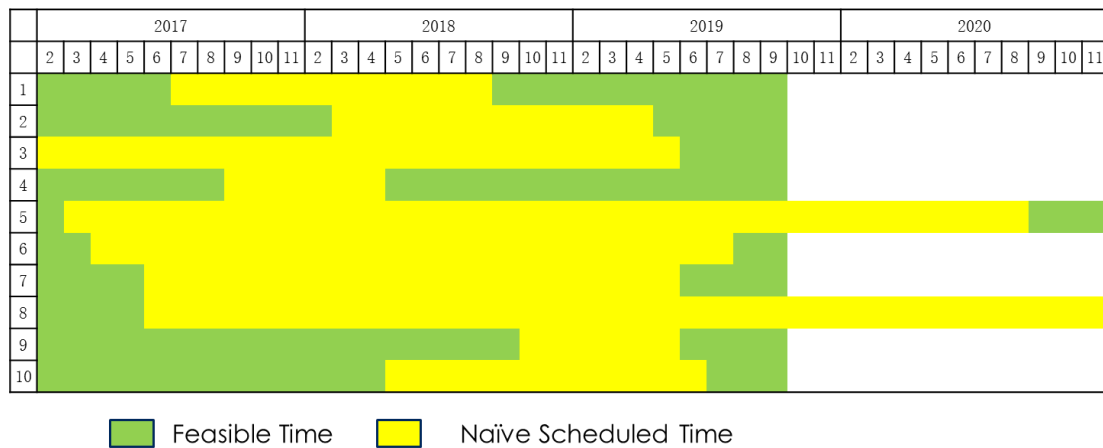
(a) Naïve schedule for the 10 projects
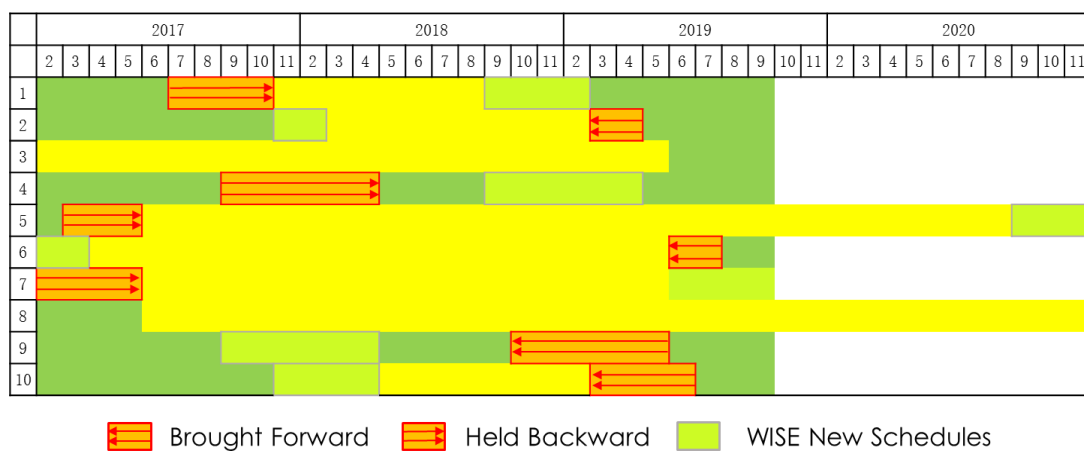
(b) WISE optimal schedule for the 10 projects

Figure 6- 6 Naïve schedule and WISE optimal schedule

It takes approximately one hour to run the WISE optimization module; the

AgBM-DTA model takes longer to complete because it must run 40 "monthly" traffic

conditions with different combinations of work zone projects for both the naïve case

and the WISE case. The final schedules of the two cases are shown in Figure 10. In

Figure 6-6 (a), the yellow cell means the month is scheduled for construction for the

corresponding project for the naïve case, and the green cell means the month is feasible

to schedule for the corresponding project. Figure 6-6 (b) shows the WISE case schedule,

and the difference from the naïve case.

The traffic performance after running AgBM-DTA is shown in Table 6-2.

Comparing the WISE case with the naïve case, work zone cost totals have savings of approximately $ 4.0 million during the 40-month work zone period. One can also observe that WISE gives a heuristic schedule where the single work zone delay is minimized, while the dependencies among projects are not perfectly considered (e.g., project 1,2 and project 5, 7, 8). The application shows WISE is helpful in decreasing the total delay cost.

Table 6- 2 Traffic performance summary

	Naïve	WISE	Change
Average Speed (mph)	33.723	33.725	+0.002
Total Delay (M Hours)	1,069.09	1,068.82	-0.279
Total Delay Cost (M $)	16,036.37	16,032.33	-4.03

Section 6.4 Conclusions and recommendations

6.4.1 Current work summary

The research team has extended the original WISE algorithm to work with SILK AgBM-DTA model to enhance work zone planning and operation applications in Maryland. At the current stage, the research team has integrated the original WISE work zone schedule optimization algorithm into AgBM-DTA. The integration is done using python code to: read and write work zone schedule inputs; read DTA model outputs;

and call the WISE optimization algorithm to provide optimal schedules as inputs for the DTA model for planning and operation applications. One small network and one large-scale, real-world DTA model have been tested to show the capability of the integrated tool. The real-world application is based on planned work zone projects obtained and selected from the CTP dataset. The DTA model has been well calibrated and validated via field detected traffic data. The results indicate WISE can provide a cost-saving schedule compared with a naïve scenario.

6.4.2 Improvements and future work

There are still limitations that can be addressed by future improvements. The WISE algorithm, based on the Tabu search algorithm, can be improved with advanced SBO methods in future research to further decrease the total work zone-related cost. SBO is an optimization method that can work well with problems whose objective functions can only be calculated via simulations (Chen et al. 2015; Chen et al. 2016).

Moreover, the current SILK AgBM-DTA is an off-line model that simulates travelers' behaviors and traffic conditions for typical scenarios. That is, the integrated WISE tool is for off line scheduling, but cannot be used to support work zone operations and real-time decision-making. The UMD research team is currently working on the online version of SILK AgBM-DTA. Once complete, the online AgBM-DTA can be applied to support work zone traffic operations. The integrated tool will not only be useful for planning applications, but also applicable   to support real-time simulation for work zone traffic and demand management, ATM strategies, and travel demand

management and ATIS guidance. It's also possible that an on-line AgBM-DTA simulation tool could be used as a decision support engine for real-time integrated corridor management strategies (as was identified in the recently completed I-95 Integrated Corridor Management Concept of Operations).

Another possible improvement relates to performance measures. The current performance measures are from the original WISE SHRP2 project, and may not be specific to Maryland. The research team will consider integrating Maryland DOT/SHA work zone performance measures with WISE.

6.4.3. Tool implementation

The research team has integrated and demonstrated the WISE work zone schedule tool for applications in the state of Maryland. To implement this tool, the research team proposes to package all the modules into SILK AgBM-DTA as an easy-to-use software application. Currently, DTALite already has a GUI which can display and edit a traffic network, conduct traffic simulation, visualize simulation outputs, and define incidents/tolls/work zones, etc. The research team is working on building an AgBM modeling approach into the GUI. That is, the integrated AgBM-DTA can be easily run by clicking some GUI buttons. In future research, the UMD team will also build the WISE work zone optimization module into the DTALite GUI.
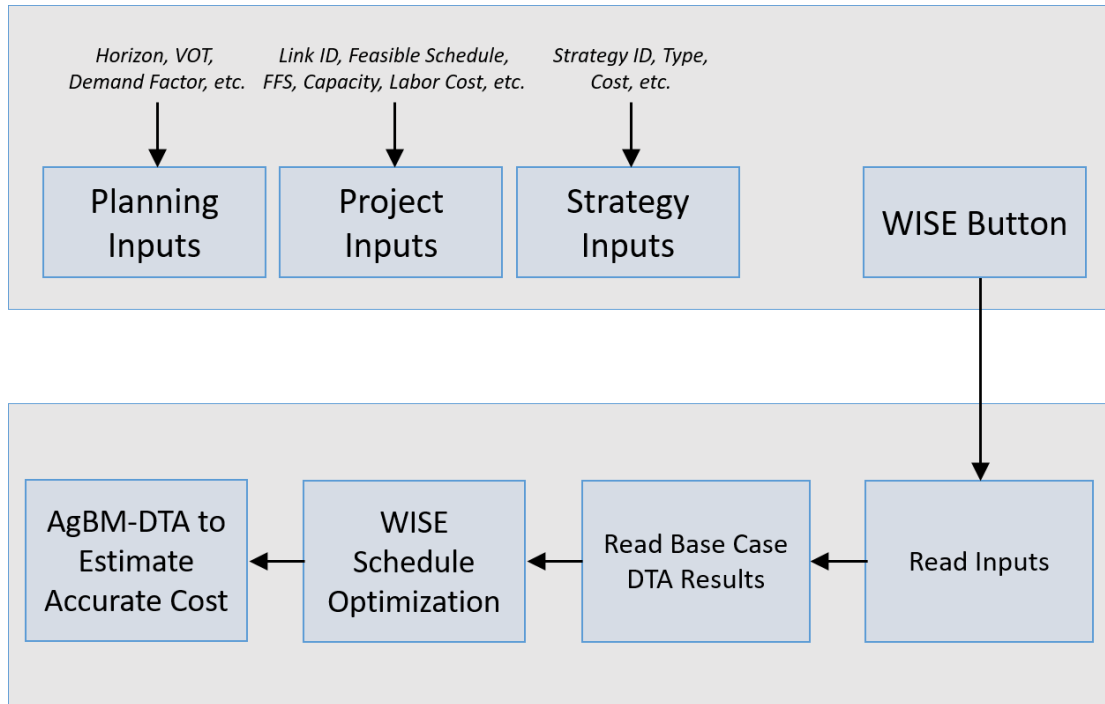
Figure 6- 7 Implementation of the integrated tool

This idea recommends adding a group of buttons in the DTALite GUI for WISE (Figure 6-7). Three input buttons are designed: "planning inputs", "project inputs", and "strategy inputs". Once the "planning inputs" button is clicked, users are able to specify the planning inputs such as scheduling horizon, value of time, and demand factors. The "project inputs" is used to type in work zone project information, in terms of earliest begin time, latest end time, strategies, duration, and construction cost. With the help of DTALite GUI, users can define the link ID, capacity reduction, and speed limit reduction of the work zone projects. The "strategy inputs" button can help the users define work zone construction strategies for WISE. After preparing these inputs, users can just click the "WISE" button to conduct work zone scheduling. The built-in program will start by reading WISE inputs, followed by the base case DTA results, and then run the WISE scheduling module to optimize work zone schedules.

Finally, the optimal schedule will be taken into AgBM-DTA to estimate accurate work zone cost as well as to conduct operational applications.

Section 6.5 Summary

Transportation network modeling can be used for both planning and operation applications. This chapter aims at showing the capability of the developed AgBM-DTA model on a specific planning application—work zone. Work zones are a primary cause of unexpected delays and one of the few types of traffic incidents that can actually be controlled by transportation agencies. The underlying notion of the research on WISE is that planning and scheduling of work zones can mitigate some of these delays, and also improve the travel reliability and safety of the overall transportation system.

The most significant outcome of this research is the production of an application-ready, integrated transportation model and work zone schedule optimization tool that can help schedule work zone projects on a planning horizon of 1-5 years; and in the future, work zone operations coordination in real-time based on the ongoing development of this tool for real-time modeling applications. This research effort benefits heavily through the leveraging of ongoing work to develop Maryland agent-based mesoscopic transportation modeling system.

In a large-scale, real-world application of the integrated transportation and work zone schedule optimization tool (AgBM-DTA with WISE) developed under this project, 10 major work zone projects from the Maryland CTP have been considered.

Integrating AgBM-DTA with WISE to optimize the scheduling of these 10 major projects resulted in a total user delay cost savings of $ 4 million over a 40-month construction period.

Chapter 7: Summary and Conclusions

Section 7.1 Work summary and conclusions

This dissertation play with a complex transportation modeling system. The first part attempts to integrate DTA with the SILK agent-based positive travel behavior model as a system model that is capable to evaluate transportation planning and traffic management related scenarios. In the proposed framework, travelers no longer have perfect network knowledge to maximize their travel utility. Instead, they are learning and searching for better choices to decline their costs due to delay, schedule early, and schedule late. The integration with the positive model enhances the behavior realism of DTA, resulting in the capability to capture dynamic travel behavior pattern changes. I also discuss the calibration of behavior parameters for the integrated system model via SBO. A county level scale system model has been developed to study the traffic impact of cumulative traffic analysis. Since the traffic impact study has already been covered in my master thesis (Zhu 2014), in this dissertation I mainly focus on the calibration of the integrated AgBM-DTA model.
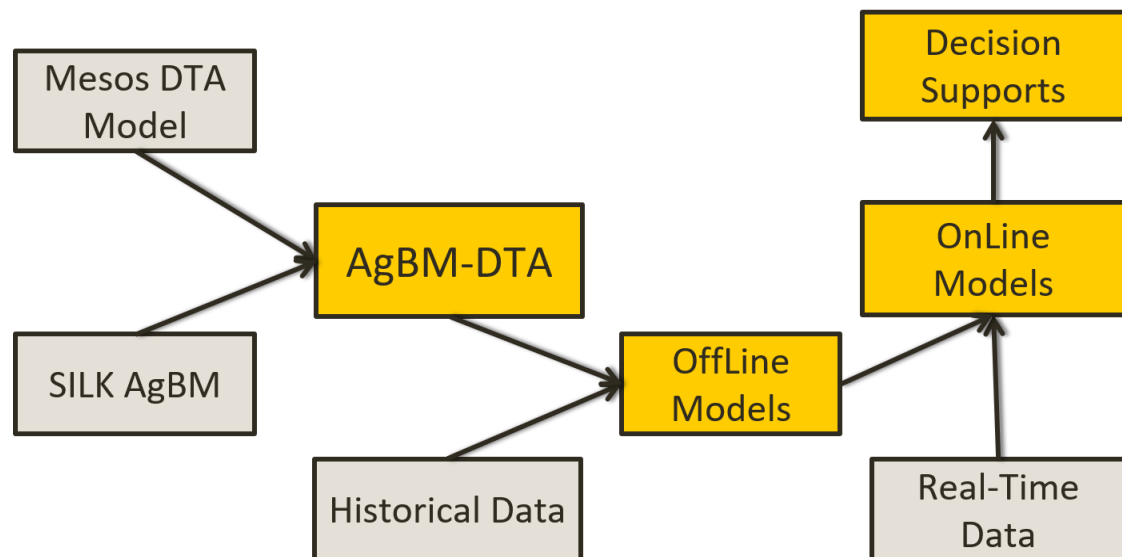
Figure 7- 1 Summary of the online modeling approach

In the second part, a gradient-based fast online calibration procedure has been proposed to enable the system model for real-time decision supports. This research is one of the earliest attempts to introduce both agent-based modeling and online network modeling for large-scale networks. In order to achieve real-time simulation, I use historical data to obtain a series of offline models; then I use the offline model's path assignment and queueing diagram to estimate the gradient of DTA (both demand and supply) parameters towards the gap between simulation outputs and real-time traffic observations. For demonstrative purposes, a true shape medium-scale simulation model has been developed and tested. I design and conduct case studies to evaluate the performance of the proposed online calibration approach under both recurrent and non-recurrent conditions. The results indicate the proposed model is fast for online transportation modeling. In addition, I also test the online modeling approach on a large-scale network that covers the whole Washington D.C. area. After reviewing different ATM strategies, I conduct an online case study to show the value of this

proposed tool for real-time decision supports.

In the last part, I integrated the system model into an existing work zone scheduling optimization module. The purpose of this integration is to show the capability of the system model on planning applications. As an open source tool, it can be integrated with different modules for a more variety of applications. The whole framework of the proposed transportation network modeling system is illustrated in Figure 7-1.

The contribution of this dissertation includes: 1) integrate an agent-based travel behavior model into DTA models to enhance the behavior realism; 2) propose a gradient based fast online calibration procedure that contains a principal component based consistency checking process, a linear adjustment process, and an optimization-oriented parameter estimation process; 3) demonstrate the practical value of the proposed system model on both planning side and real-time operation side. The practical value of this tool can be as much as the theoretical value of this dissertation because the implementation of the proposed tool is straight forward.

To conclude, I have learnt the entire process to integrate different model components as a whole system. It is feasible to apply this system model for offline or online applications. The time factor for online applications is over hundreds for medium-scale networks, and around 10 for large-scale models. Future work is required to consider enroute travel behavior shifts for ATM evaluations.

Section 7.2. Future Work

There are still limitations in the system model. The AgBM part may request more ground truth data for further calibration. In the online calibration part, future research could include the real-time incident detection and prediction procedure into the online framework for non-recurrent scenario calibration. For now, one can only detect non-recurrent scenarios by outliers in terms of real-time counts and speeds data. Moreover, different α and β need to be tested for sensitivity analysis to understand whether overall adjustment or local adjustment play a more important role. Future research could also include parallel computation for both DTA and online parameter adjustments.

Bibliography

Anthony, D. (2004). Still Stuck in Traffic: Coping with Peak-Hour Traffic Congestion. Brookings Institution Press.

Antoniou, C., Ben-Akiva, M., & Koutsopoulos, H. N. (2007). Nonlinear Kalman filtering algorithms for on-line calibration of dynamic traffic assignment models. IEEE Transactions on Intelligent Transportation Systems, 8(4), 661-670.

Area, D. M. (1955). Traffic Study. Detroit, MI.

Ashok, K. (1996). Estimation and Prediction of time-dependent Origin-Destination Flows. PhD thesis, Massachusetts Institute of Technology.

Ashok, K. and Ben-Akiva, M. (1993). Dynamic O-D matrix estimation and prediction for real-time traffic management systems. In Daganzo, C., editor, Transportation and Traffic Theory, pages 465-484. Elsevier Science Publishing.

Ashok, K. and Ben-Akiva, M. (2000). Alternative approaches for real-time estimation and prediction of time-dependent origin-destination flows. Transportation Science, 34(1):21 36.

Balakrishna, R., H. N. Koutsopoulos, and M. Ben-Akiva. Calibration and Validation of Dynamic Traffic Assignment Systems. Transportation and Traffic Theory: Flow, Dynamics and Human Interaction. In *Proceedings of the 16th International Symposium on Transportation and Traffic Theory,* University of Maryland, College Park, 2005, pp. 407-426.

Balakrishna, R., H. N. Koutsopoulos, and M. Ben-Akiva. Simultaneous Off-line Demand and Supply Calibration of Dynamic Traffic Assignment Systems. *Presented at 85th Annual Meeting of the Transportation Research Board*, Washington, D.C., 2006.

Balakrishna, R., M. Ben-Akiva, and H. Koutsopoulos. Offline Calibration of Dynamic Traffic Assignment: Simultaneous Demand-And-Supply Estimation. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2003, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 50-58.

Ben-Akiva, M.E. (1974). Structure of passenger travel demand models. *Transportation Research Record: Journal of the Transportation Research Board*, 526, 26-42.

Ben-Akiva, M. Deveopment of Dynamic Traffic Assignment System for Planning Purposes: DynaMIT User's Guide, 2002.

Behrisch, M., L. Bieker, J. Erdmann, and D. Krajzewicz. SUMO–Simulation Of Urban Mobility: An Overview. In *Proceedings of SIMUL 2011, the Third International Conference on Advances in System Simulation*. ThinkMind, 2011.

Ben-Akiva, M. E., Koutsopoulos, H. N., Mishalani, R. G., & Yang, Q. (1997). Simulation laboratory for evaluating dynamic traffic management systems. Journal of Transportation Engineering, 123(4), 283-289.

Caliper. *TransCAD User's Guide*. USA: Caliper Cooperation, 2005.

Caliper. TransModeler Traffic Simulation Software – Version 2.5 User's Guide. 2009.

Cantelmo, G., Cipriani, E., Gemma, A., & Nigro, M. (2014). An adaptive bi-level gradient procedure for the estimation of dynamic traffic demand. *IEEE Transactions on Intelligent Transportation System*s, 15(3), 1348-1361.

Cantelmo, G., Viti, F., Cipriani, E., & Nigro, M. (2017). A Utility-based Dynamic Demand Estimation Model that Explicitly Accounts for Activity Scheduling and Duration. Transportation Research Procedia, 23, 440-459.

Carpenter, S., Jones, P.M. (1983). *Recent Advances in Travel Demand Analysis*. Gower,

Aldershort, England.

Cascetta, E., and S. Nguyen. A Unified Framework for Estimating or Updating Origin/Destination Matrices from Traffic Counts. *Transportation Research Part B: Methodological*, Vol. 22, No. 6, 1988, pp. 437-455.

Cascetta, E., D. Inaudi, and G. Marquis. Dynamic Estimators of Origin– Destination Matrices Using Traffic Counts. Transportation Science, Vol. 27, No. 4, 1993, pp. 363-373.

Chen, B., & Cheng, H. H. (2010). A review of the applications of agent technology in traffic and transportation systems. IEEE Transactions on Intelligent Transportation Systems, 11(2), 485-497.

Chen, C., Gong, H., & Paaswell, R. (2008). Role of the built environment on mode choice decisions: additional evidence on the impact of density. Transportation, 35(3), 285-299.

Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. Transportation research part C: emerging technologies, 68, 285-299.

Chen, C. H., & Schonfeld, P. (2004). Work zone optimization for two-lane highway maintenance projects. Transportation Research Record: Journal of the Transportation Research Board, (1877), 95-105.

Chen, C. H., Schonfeld, P., & Paracha, J. (2005). Work zone optimization for two-lane highway resurfacing projects with an alternate route. Transportation Research Record: Journal of the Transportation Research Board, (1911), 51-66.

Chen, C. H. P., & Schonfeld, P. M. (2007). Pavement resurfacing intervals in work zone optimization. In Transportation Research Board 86th Annual Meeting (No. 07-

1625).

Chen X., Zhu Z., He X. & Zhang L. (2015) Surrogate-based optimization for solving mixed integer network design problem. *Transportation Research Record*, 2497, 124-134.

Chien, S., & Schonfeld, P. (2001). Optimal work zone lengths for four-lane highways. Journal of Transportation Engineering, 127(2), 124-131.

Chitturi, M., Shaw, J., Campbell, J., & Noyce, D. (2014). Validation of Origin-Destination Data from Bluetooth Reidentification and Aerial Observation. *Transportation Research Record: Journal of the Transportation Research Board*, (2430), 116-123.

Chu, L., H. X. Liu, J. -S. Oh, and W. W. Recker. Calibration Procedure for Microscopic Traffic Simulation. *Presented at 83rd Annual Meeting of the Transportation Research Board*, Washington, D.C., 2004.

Daganzo, C. F. (1977). On the traffic assignment problem with flow dependent costs—I. Transportation Research, 11(6), 433-437.

Daganzo, C.F. (1979). *Multinomial Probit: The Theory and Its Applications to Demand Forecasting*. Academic Press, New York.

Doan, D. L., A. Ziliaskopoulos, and H. Mahmassani. On-Line Monitoring System for Real-Time Traffic Management Applications. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1678, Transportation Research Board National Research Council, Washington, D.C., 1998, pp. 142-149.

Doniec, A., Mandiau, R., Piechowiak, S., & Espié, S. (2008). A behavioral multi-agent model for road traffic simulation. Engineering Applications of Artificial

Intelligence, 21(8), 1443-1454.

Dorigo, M., Optimization, learning and natural algorithms, Politecnico di Milano, 1992

DOT, U. (1997). National transportation statistics. Bureau of Transportation Statistics, Washington, DC.

DOT, U. (2006). Freight in America. Bureau of Transportation Statistics. Washington, DC.

Schonfeld, P., Yang, N., & Tien, S. L. (2006). Optimization of Work Zone Decisions through Simulation (No. MD-06-SP508B4C).

France, J., & Ghorbani, A. A. (2003, October). A multiagent system for optimizing urban traffic. In Intelligent Agent Technology, 2003. IAT 2003. IEEE/WIC International Conference on (pp. 411-414). IEEE.

FHWA (2015). FHWA FY 2015 Budget Estimations. https://www.transportation.gov/sites/dot.gov/files/docs/FHWA-FY2015-Budget-Estimates.pdf

Gartner, N., and C. Stamatiadis. Integration of dynamic traffic assignment with real-time traffic adaptive control system. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1644, Transportation Research Board of the National Academies, Washington, D.C., 1998, pp. 150-156.

Gentile, G., L. Meschini, and N. Papola. Spillback Congestion in Dynamic Traffic Assignment: A Macroscopic Flow Model with Time-Varying Bottlenecks. Transportation Research Part B: Methodological, Vol. 41, No. 10, 2007, pp. 1114-1138.

Grégoire, P. L., Desjardins, C., Laumônier, J., & Chaib-draa, B. (2007, September). Urban traffic control based on learning agents. In Intelligent Transportation

Systems Conference, 2007. ITSC 2007. IEEE (pp. 916-921). IEEE.

Hao, J.Y., Hatzopoulou, M., Miller, E.J. (2010). Integrating an activity-based travel demand model with dynamic traffic assignment and emission models. *Transportation Research Record: Journal of the Transportation Research Board*, 2176, 1-13.

Hashemi, H., Abdelghany, K. F., & Abdelghany, A. F. (2017). A multi-agent learning approach for online calibration and consistency checking of real-time traffic network management systems. Transportmetrica B: Transport Dynamics, 5(3), 369-389.

Hawas, Y. E. Calibrating Simulation Models for Advanced Traveler Information Systems/Advanced Traffic Management Systems Applications. *Journal of Transportation Engineering*, Vol. 128, No. 1, 2002, pp. 80-88.

He, R., Miaou, S., Ran, B., and Lan, C. (1999). Developing an on-line calibration process for an analytical dynamic traffic assignment model. In Presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C.

He, X. Simulation-Based Optimization of Transportation Systems: Theory, Surrogate Models, and Applications. PhD Dissertation, University of Maryland, College Park, MD, 2014.

Hoch, I. (1959). Forecasting Economic Activity for the Chicago Region: Final Report. Chicago Area Transportation Study.

Huynh, N., Mahmassani, H., and Tavana, H. (2002). Adaptive speed estimation using trans fer function models for real-time dynamic traffic assignment operation. In Proceedings of the 81st Annual Meeting of the Transportation Research Board.

Inose, H., Fujisaki, H., & Hamada, T. (1967). Theory of road-traffic control based on

macroscopic traffic model. Electronics Letters, 3(8), 385-386.

Ison, S. (1996). Pricing road space: back to the future? The Cambridge experience. Transport Reviews, 16(2), 109-126.

Janson, B.N. (1991). Dynamic traffic assignment for urban networks. *Transportation Research Part B*, 25(2), 143-161.

Jiang, X., & Adeli, H. (2003). Freeway work zone traffic delay and cost optimization model. Journal of Transportation Engineering, 129(3), 230-241.

Kaysi, I. (1993). An integrated approach to vehicle routing and congestion prediction for real-time driver guidance (Doctoral dissertation, Massachusetts Institute of Technology Cambridge, Mass).

Kitamura, R., Chen, C., Pendyala, R. M., & Narayanan, R. (2000). Micro-simulation of daily activity-travel patterns for travel demand forecasting. Transportation, 27(1), 25-51.

Koppelman, F.S., Pas, E.I. (1980). Travel-choice behavior: models of perceptions, feelings, preference and choice. *Transportation Research Record: Journal of the Transportation Research Board*, 765, 26-33.

Kosonen, I. (2003). Multi-agent fuzzy signal control based on real-time simulation. Transportation Research Part C: Emerging Technologies, 11(5), 389-403.

Mahmassani, H. S., H. Sbayti, and X. Zhou. Dynasmart-p Version 1.0 User's Guide. Maryland Transportation Initiative, College Park, Maryland, 2004, p. 137.

Mammar, S., Messmer, A., Jensen, P., Papageorgiou, M., Haj-Salem, H., & Jensen, L. (1996). Automatic control of variable message signs in Aalborg. Transportation Research Part C: Emerging Technologies, 4(3), 131-150.

Levinson, H. S., Zimmerman, S., Clinger, J., & Rutherford, G. S. (2002). Bus rapid transit: An overview. Journal of Public Transportation, 5(2), 1.

Li, R., & Shi, Q. (2003, October). Study on integration of urban traffic control and route guidance based on multi-agent technology. In Intelligent Transportation Systems, 2003. Proceedings. 2003 IEEE (Vol. 2, pp. 1740-1744). IEEE.

Lin, D.-Y., Eluru, N., Waller, S.T., Bhat, C.R. (2008). Integration of activity-based modeling and dynamic traffic assignment. *Transportation Research Record: Journal of the Transportation Research Board*, 2076, 52-61.

Logi, F., & Ritchie, S. G. (2002). A multi-agent architecture for cooperative inter-jurisdictional traffic congestion management. Transportation Research Part C: Emerging Technologies, 10(5), 507-527.

Lowrie, P. R. (1982). The Sydney coordinated adaptive traffic system-principles, methodology, algorithms. In International Conference on Road Traffic Signalling, 1982, London, United Kingdom (No. 207).

Lu, C. C., Zhou, X., & Zhang, K. (2013). Dynamic origin–destination demand flow estimation under congested traffic conditions. Transportation Research Part C: Emerging Technologies, 34, 16-37.

Martinelli and Xu 1996 Delay estimation and optimal length for four-lane divided freeway workzone Journal of transportation engineering 122(2).

May Jr, A. D., & Harmut, E. M. (1967). Non-integer car-following models. Highway Research Record, (199).

McCoy, P., & Mennenga, D. (1998). Optimum length of single-lane closures in work zones on rural four-lane freeways. Transportation Research Record: Journal of the Transportation Research Board, (1650), 55-61.

McCoy, P. T., Pang, L. M., & Post, E. R. (1980). Optimum length of two-lane two-way no-passing traffic operation in construction and maintenance zones on rural four-lane divided highways. Transportation research record, 773, 20-24.

Meng, Q., & Weng, J. (2013). Optimal subwork zone length and project start time for short-term daytime work zones from the contractor's perspective.Transportation Research Part C: Emerging Technologies, 29, 72-83.

Merchant, D. K., & Nemhauser, G. L. (1978). A model and an algorithm for the dynamic traffic assignment problems. Transportation science, 12(3), 183-199.

Mokhtarian, P. L., & Chen, C. (2004). TTB or not TTB, that is the question: a review and analysis of the empirical literature on travel time (and money) budgets. Transportation Research Part A: Policy and Practice, 38(9), 643-675.

Omrani, R., and L. Kattan. (2012). Demand and Supply Calibration of Dynamic Traffic Assignment Models: Past Efforts and Future Challenges. In Transportation Research Record: Journal of the Transportation Research Board, No. 2283, Transportation Research Board of the National Academies, Washington, D.C., pp. 100-112.

Panwai, S., & Dia, H. (2007). Neural agent car-following models. IEEE Transactions on Intelligent Transportation Systems, 8(1), 60-70.

Papageorgiou, M., Hadj-Salem, H., & Blosseville, J. M. (1991). ALINEA: A local feedback control law for on-ramp metering. Transportation Research Record, 1320(1), 58-67.

PTV. VISSIM User Manual (Version 4.51). 2008.

PTV. VISUM 11 User Manual. Karlsruhe, Germany, 2009.

Qin, X., & Mahmassani, H. (2004). Adaptive calibration of dynamic speed-density

relations for online network traffic estimation and prediction applications. Transportation Research Record: Journal of the Transportation Research Board, (1876), 82-89.

Qu, Y., & Zhou, X. (2017). Large-scale dynamic transportation network simulation: A space-time-event parallel computing approach. *Transportation Research Part C: Emerging Technologies*, *75*, 1-16.

Roozemond, D. A. (2001). Using intelligent agents for pro-active, real-time urban intersection control. European Journal of Operational Research, 131(2), 293-301.

Rossetti, R. J., Bampi, S., Liu, R., Van Vliet, D., & Cybis, H. B. B. (2000). An agent-based framework for the assessment of drivers' decision-making. In Intelligent Transportation Systems, 2000. Proceedings. 2000 IEEE (pp. 387-392). IEEE.

Rothkrantz, L. J. (2009, June). Dynamic routing using the network of car drivers. In Proceedings of the 2009 Euro American Conference on Telematics and Information Systems: New Opportunities to increase Digital Citizenship (p. 11). ACM.

Schelling, T. C. (1971). Dynamic models of segregation. Journal of mathematical sociology, 1(2), 143-186.

Schneider, M. (1959). Gravity models and trip distribution theory. Papers in Regional Science, 5(1), 51-56.

Schonfeld, P., & Chen, C. H. (2006). Work Zone Optimization for Multiple-Lane Highway Resurfacing Projects with Time Constraints and Alternate Route. In Transportation Research Board 85th Annual Meeting (No. 06-1812).

Schonfeld, P., & Chien, S. (1999). Optimal work zone lengths for two-lane highways. Journal of Transportation Engineering, 125(1), 21-29.

Schonfeld, P. M., & Tien, S. L. (2006). Duration Versus Cost Trade-offs in Work-Zone Optimization. In Transportation Research Board 85th Annual Meeting (No. 06-1789).

Schrank, D., Eisele, B., Lomax, T., (2012). The Urban Mobility Report. Texas Transportation Institute, College Station, TX.

Schrank, D., Eisele, B., Lomax, T., & Bak, J. (2015). 2015 urban mobility scorecard.

Shi, X., Xu, J., Xu, Y., & Song, J. (2005, September). A simulation study on agent-network based route guidance system. In Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE (pp. 248-253). IEEE.

Tang, Y., & Chien, S. (2008). Scheduling work zones for highway maintenance projects: Considering a discrète time-cost relation.Transportation Research Record: Journal of the Transportation Research Board, (2055), 21-30.

Tang, Y., & Chien, S. (2009). Optimization of Work Zone Schedule Considering Time-Varying Traffic Diversion. Transportation Research Board.

Tavana, H. and Mahmassani, H. (2000). Estimation and application of dynamic speed density relations by using transfer function models. Transportation Research Record 1710, pages 47-57.

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., C. Antoniou, Y. Wen, and M. Ben-Akiva. Calibration of Dynamic Traffic Assignment Models with Point-To-Point Traffic Surveillance. In Transportation Research Record: Journal of the Transportation Research Board, No. 2090, Transportation Research Board of the National Academies, Washington, D.C.,

2009, pp. 1-9.

Wahle, J., Bazzan, A. L. C., Klügl, F., & Schreckenberg, M. (2002). The impact of real-time information in a two-route scenario using agent-based simulation. Transportation Research Part C: Emerging Technologies, 10(5), 399-417.

Wang, Y., & Papageorgiou, M. (2005). Real-time freeway traffic state estimation based on extended Kalman filter: a general approach. Transportation Research Part B: Methodological, 39(2), 141-167.

Wardrop, J. G. (1952). Road paper. Some theoretical aspects of road traffic research. Proceedings of the institution of civil engineers, 1(3), 325-362.

Weyns, D., Holvoet, T., & Helleboogh, A. (2007, September). Anticipatory vehicle routing using delegate multi-agent systems. In Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE (pp. 87-93). IEEE.

Wu, J.H., Chan, Y., Florian, M. (1998). The continuous dynamic network loading problem: a mathematical formulation and solution method. *Transportation Research Part B*, 32(3), 173-187.

Xiong, C., & Zhang, L. (2013). Positive model of departure time choice under road pricing and uncertainty. Transportation Research Record: Journal of the Transportation Research Board, (2345), 117-125.

Yang, Z. S., Chen, X., Tang, Y. S., & Sun, J. P. (2005, August). Intelligent cooperation control of urban traffic networks. In Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on (Vol. 3, pp. 1482-1486). IEEE.

Yang, J., Liu, Y., Qin, P., & Liu, A. A. (2014). A review of Beijing′s vehicle registration lottery: Short-term effects on vehicle growth and fuel consumption. Energy Policy, 75, 157-166.

Yang, N., Schonfeld, P., & Kang, M. W. (2009). A hybrid methodology for freeway work-zone optimization with time constraints. Public Works Management & Policy, 13(3), 253-264.

Yang, Q., H. Koutsopoulos, & M. Ben-Akiva. Simulation Laboratory for Evaluating Dynamic Traffic Management Systems. In Transportation Research Record: Journal of the Transportation Research Board, No. 1710, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 122-130.

Zambonelli, F., & Parunak, H. V. D. (2002, September). Signs of a revolution in computer science and software engineering. In International Workshop on Engineering Societies in the Agents World (pp. 13-28). Springer Berlin Heidelberg.Zhang, L. (2007). Developing a positive approach to travel demand forecasting: Theory and model. *Traffic and Transportation Theory*, 17, 136-150.

Zhang, L. (2007). Developing a positive approach to travel demand forecasting: Theory and model. *Traffic and Transportation Theory*, 17, 136-150.

Zhou, X., & Mahmassani, H. S. (2005). Recursive approaches for online consistency checking and OD demand updating for real-time dynamic traffic assignment operation. *Transportation Research Record*, 1923, 218-226.

Zhou, X., & Mahmassani, H. S. (2007). A structural state space model for real-time traffic origin–destination demand estimation and prediction in a day-to-day learning framework. *Transportation Research Part B: Methodological*, 41(8), 823-840.

Zhou, X., H. S. Mahmassani, and K. Zhang. (2008). Dynamic Micro-Assignment Modeling Approach for Integrated Multimodal Urban Corridor Management. *Transportation Research Part C: Emerging Technologies*, Vol. 16, No. 2, pp. 167-186.

Zhou, X., and J. Taylor. DTALite: A Queue-Based Mesoscopic Traffic Simulator for Fast Model Evaluation and Calibration. Cogent Engineering, Vol. 1, No. 1, 2014, p. 961345.

Zhu, Z. (2014). Traffic analysis on cumulative land development and transportation related policy scenarios (Doctoral dissertation, University of Maryland, College Park).

Zhu, Z., C. Xiong, X. Chen, and L. Zhang. Integrating Dynamic Traffic Assignment and Agent-Based Travel Behavior Models for Cumulative Land Development Impact Analysis. *Presented at the 7th International Symposium on Travel Demand Management*, Tucson, Arizona, 2015.

Zhu, Z., Xiong, C., Chen, X., He, X., & Zhang, L. (2015). Agent-Based Microsimulation Approach for Design and Evaluation of Flexible Work Schedules. Transportation Research Record: Journal of the Transportation Research Board, (2537), 167-176.