Exploration of Algorithms for Calibration and Optimization of Transportation Networks

by

Yundi Zhang

B.Eng. Civil Engineering
Hong Kong University of Science and Technology, 2014

SUBMITTED TO THE DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN TRANSPORTATION
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2017

Signature of Author: Signature redacted
Department of Civil and Environmental Engineering
May 19, 2017

Certified by: Signature redacted
Moshe E. Ben-Akiva
Edmund K. Turner Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by: Signature redacted
Jesse Kroll
Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

Exploration of Algorithms for Calibration and Optimization of Transportation Networks

by

Yundi Zhang

Submitted to the Department of Civil and Environmental Engineering
on May 19, 2017 in Partial Fulfillment of the Requirements for
the Degree of Master of Science in Transportation

ABSTRACT

Calibration of a DTA model is needed before applying it in a real-world road network. Calibration is often formulated as an optimization problem. This problem is difficult because of the large number of parameters, computational burden of the simulator, and stochasticity of the simulation process.

Different algorithms to address the calibration problem are reviewed, and the WSPSA algorithm is showed best performance in offline calibration of a DTA model. This algorithm is an extension to the SPSA algorithm, by incorporating a weight matrix in the gradient calculation process which takes into account the heterogeneous correlation between measurements and parameters.

However, existing literature on the WSPSA algorithm did not fully develop the model to accommodate all types of calibration problems. Specifically, a comprehensive framework to generate the weight matrix is not developed to account for traffic speed data and non-OD parameters. Besides, tests indicate that the convergence rate highly depends on the settings of the algorithm.

In this thesis, extensions and improvements are made to different aspects of the WSPSA algorithm, including generation of the weight matrix, and updating the parameters. The proposed WSPSA algorithm demonstrates better performance in the synthetic test.

Following the synthetic tests, a case study case study is conducted in a real-world network, where the proposed algorithm is applied to calibrate a microscopic traffic simulator. Sensor count and speed data from real-world observations are used to simultaneously calibrate OD demand and behavior parameters. The calibration yields satisfactorily accurate result, which proves superior performance of the proposed WSPSA algorithm.

Thesis Supervisor: Moshe E. Ben-Akiva

Title: Edmund K. Turner Professor of Civil and Environmental Engineering

Acknowledgement

At the moment when I have just completed this thesis, I feel excited and grateful. Nearly two years ago, I obtained the opportunity to work as a student research assistant in the Intelligent Transportation Systems (ITS) Lab at MIT. From then on, I started my research on traffic network calibration and toll optimization under the DTA context. Faced with this challenging topic, I dug deeper step by step following instructive suggestions from my advisors and colleagues. Now I am glad to have finally yielded some deliverable, and I shall acknowledge this work to all those who offered inspirations and supports to me.

First I would like to express my sincerest gratitude to my research advisor Professor Moshe Ben-Akiva. I would like to thank him for leading me to such an engaging research topic and for helping me shape the big picture. I also thank him for being understanding and supportive, without which I could not have been able to freely explore so many possibilities in addressing the problem. In addition, he has always been enthusiastic about research innovations – it is a steady energy that encourages me to keep focused on overcoming obstacles and looking forward.

Second I will express my thanks to Doctor Bilge Atasoy. As my post-doctoral supervisor, she has invariably been giving me valuable guidance and suggestions on research in details. I am especially grateful that she has been discussing about research updates with me on a daily basis for the past several months. The meetings greatly accelerated my research progress and timely communications helped me stay on the right track. I would like to acknowledge the efforts of Doctor Bilge Atasoy, a most learned and intelligent supervisor with a high sense of responsibility.

Next, I must thank Cintra, our collaborator on this toll optimization project. I would like to thank our collaborators Ricardo Sanchez, Thu Hoang, Andres De Los Rios, Megan Rhodes, John Brady, and Wei He, for the continuous help and valuable feedback throughout the project. They also provided us with multiple types of real-world data, without which I would not be able to apply methodologies to practice to conduct the case study.

Following that, I want to thank other members of the ITS Lab, especially Doctor Ravi Seshadri, who led me into the world of traffic simulation and calibration, as well as Haizheng Zhang, Shi Wang, Samarth Gupta, who have been working together with me on this topic. They are splendid teammates since not only have they refined methodologies and platforms I have applied, but also they provided insights that enriched my knowledge base. I am honored to be a member of such an awesome research team.

Last but not the least, I want to give thanks to my family and friends. They have been supporting me both materially and spiritually. I feel blessed for being surrounded by caring and liberal parents, whom I wholeheartedly respect for giving me not only general guidance and advices, but also freedom. I am also thankful to my friends, the most intelligent peers who can always make my life vigorous and colorful.

This thesis is yet a new start of my research. I am now pursuing the PhD degree in transportation. I can foresee that there are many unknown unknowns, some of which are bitter while the rest are sure to be rewarding. I will start my new trip of exploration with curiosity and courage. Again from the bottom of my heart, I express my gratitude to everyone having offered help to me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

With increasing mobility demand, traffic congestions have been a persistent problem in many places. Congestions not only reduce travel efficiency and traffic safety for travelers, but also waste fuels and deteriorate the environment. In view of this, various congestion management schemes have been researched and implemented in recent years.

1.1 Congestion Management Scheme

Congestion Management Scheme aims to improve transportation system performance and reduce the adverse impact of traffic congestions by either altering traffic demand of changing transportation supply.

Among congestion management schemes, road pricing (i.e., tolling) is a traditional strategy, which may aim to generate revenue to recover road construction and maintenance cost thus incentivize improvement of transportation infrastructures, as well as managing congestion by altering travel behaviors.

Traditionally, flat-rate tolling has been successful in generating revenue, but was not effective in changing travelers' behaviors. In most cases, congestions are not caused by excessive overall demand, but by peaked flow at peak hours due to uneven distribution of travel demand. Time-of-day tolling strategy could address this to some extent, by altering travelers' choice of departure times, and this tolling strategy is more effective in reducing congestion during peak hours.

Toll managed lane (Figure 1.1) is a congestion management strategy where certain lanes of a freeway (often called Express Lanes) are open to vehicles at a charge, and driving on those lanes is faster than their counterpart general purpose lanes. In this case, a more advanced tolling strategy, dynamic tolling (Figure 1.2), would be necessary, because ensuring free-flow condition on managed lanes is a key constraint in managed lane operations, while static or time-of-day tolling are not capable of doing this. Dynamic tolling strategy dynamically adjust toll rates to address changing traffic conditions. When a reactive tolling strategy is in place, toll rates increase when flow on managed lanes are high, and vice versa, so that the desired level-of-service on managed lanes could be maintained. In practice, this is typically done with a look-up table together with professional judgement.

Figure 1.1 Toll managed lanes

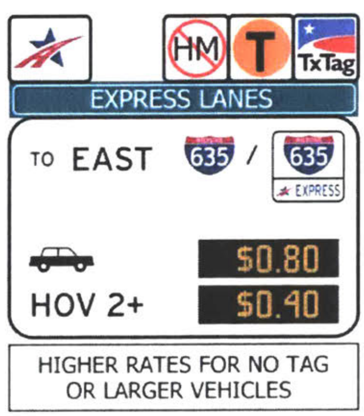

Figure 1.2 Dynamic pricing on toll managed lanes

In contrast to the reactive control strategies, optimization-based tolling strategies can consider different types of objectives, while maintaining the constraints. For example, a managed lane operating company may want to maximize its toll revenue, while ensuring the travel speed on managed lanes is above a certain value. Compared to reactive dynamic tolling, optimization-based tolling strategies can explicitly model the revenue-maximization objective as an optimization problem.

Application of the optimization-based tolling strategy often requires a Dynamic Traffic Assignment (DTA) Model that runs in real-time to evaluate the impact of changes in toll rates, Real-time decisions on change of toll rates will be made according to the projected traffic conditions and toll revenue in the DTA model. Besides dynamic tolling, research and application of other congestion management schemes may also utilize DTA models.

1.2 Dynamic Traffic Assignment Model

Dynamic Traffic Assignment is a model that dynamically assign traffic to a road network. Compared to a static model for planning purpose, DTA model captures real time traffic conditions and the evolvement of traffic congestions by modeling time-dependent traffic demand and network status.

DTA models are divided into analytical models and simulation-based models. Analytical models are highly efficient in terms of computation, but the highly simplified demand and supply models usually cannot capture the complex behaviors of travelers and interaction between demand and supply. Simulation based models use more realistic models to simulate travelers and their interaction with the road network, which usually consumes much more computational resources than analytical models. However, a mesoscopic simulation-based DTA model simulates traffic in a higher level instead of modeling individual vehicles, thus it is capable of modeling a much larger network within certain computational constraints, compared to a microscopic traffic simulation model.

1.3 Thesis Motivation

With the help of a DTA model, the optimization-based tolling strategy can rely on a simulation-based dynamic toll optimization scheme (Wang, 2016). This would be a solution for a more effective way of tolling which uses a DTA model to proactively predict traffic conditions in real-time, under different scenarios of future toll rates. Based on the predictions and associated objective values, this scheme searches for the future toll rates that optimize the objective.

The success of this framework relies on an efficient and accurate algorithm to search for the optimal toll rates, as well as a DTA model that is well-calibrated to accurately simulate the real network.

Calibration is an iterative process of finding the best input parameters to the model that makes the outputs match real-world measurements. For a DTA model, input parameters consists of demand parameters (i.e., time-dependent origin-destination demand), supply parameters (free-flow speed, capacity, etc. of each road segment) and driving-behavior parameters (value of time, acceleration and deceleration rates, lane-changing parameters, etc.), and outputs often include sensor count and speed readings, travel time, etc. In other words, calibration of the DTA model is process of finding the best OD demand and other parameters to the model that results in a set of sensor readings matching real-world sensor measurements. Accurate calibration of the simulators ensures the simulators represent the real world accurately, and is essential for the validity of any management schemes developed.

Besides, testing of the toll optimization scheme requires a closed-loop framework, where a microscopic traffic simulator is used in place of the real network. To ensure the validity of the closed-loop testing framework, calibration of both the DTA model and the microscopic simulator is needed.

It is often straightforward to formulate calibration into an optimization problem, which is to minimize the deviation between simulated and real-world traffic conditions by changing the input parameters to the model. On the other hand, congestion management schemes often involve some optimization problems, e.g., to maximize social welfare, to minimize total travel time, etc. Therefore, better congestion management schemes need to benefit from appropriate algorithms for the calibration and optimization of transportation networks.

This thesis focuses on the algorithms for calibration of the transportation network simulators.

Since the calibration is formulated as an optimization problem, and improvements are made to account for the characteristics of a transportation network, the algorithms developed and explored will be transferrable to calibrating other transportation network simulators and optimizing transportation networks. Besides, this calibration work needs to understand the effect of tolling on behavior of drivers, which is also critical for the follow-up work on toll optimization.

1.4 Thesis Outline

Literatures about calibration are reviewed in the first section of Chapter 2. The calibration is formulated as an optimization problem that minimizes the deviation between true and simulated traffic conditions, and literatures on such optimization algorithms are reviewed in the second section of Chapter 2.

Chapter 3 starts with in-detailed description of the models in the traffic simulators, as well as technical details of the closed-loop testing framework and the formulation and solution algorithms of the calibration problem. Improvements and modifications to the calibration algorithm are then introduced, which are proved to improve the efficiency of calibration process and accuracy of the calibration results.

The improved calibration algorithm is tested by calibrating a microscopic traffic simulator with data from a real-world road network. The results, together with empirical considerations and network-specific issues are presented in Chapter 4.

This case study is part of the toll optimization scheme. Future research will be discussed in Chapter 5, which includes the toll optimization scheme, as well as further improvements and other algorithms for calibration.

Chapter 2

Literature Review

This chapter reviews literatures relevant to the thesis topic. Firstly, literatures on offline calibration of dynamic traffic assignment models are introduced. Then different algorithms for calibration and optimization are reviewed.

2.1 Framework for Calibration of DTA Model

Calibrations of a DTA model is essential to ensure it accurately mimic the real-world traffic conditions. A DTA model often consists of two sets of models: demand and supply models. DTA models are divided into macroscopic, mesoscopic and microscopic models, which uses demand and supply models with different degree of fineness. For example, a macroscopic supply model uses speed-flow relationship to describe traffic conditions on road segments, while a microscopic model simulates the movement of individual vehicles, which includes driver's acceleration and braking behaviors and lane changing behaviors. Depending on the their degree of fineness, demand models may include OD demand, travel behavior and driving behavior parameters, while supply models may include road segment characteristics like free flow speed and capacity.

2.1.1 Calibration of OD Demand

Abundant researches have been conducted to address OD demand calibration. One direction of OD flow estimation studies is static OD estimation, which assumes static OD flows. However under DTA context, dynamic OD flows are required as input to reflect time-dependent traffic conditions. This research direction has been widely explored. Balakrishna et al. (2005) and Toledo et al. (2014) summarized research advancements on dynamic OD flow estimations.

2.1.2 Calibration of Behavior Models

Behavior models capture individual driver's travel and driving behavior in the microscopic level. Behavior models includes route choice, car-following (acceleration, deceleration and desired speed), lane-changing (gap acceptance, merging, yielding and look-ahead) and compliance (response to arterial signals, ramp meters and toll plazas) These models are described by a large number of parameters, the calibration of which is challenging yet important for reflecting realistic traffic conditions with high fidelity.

Route choice model is the core component of behavior models. Ben-Akiva and Lerman (1985) systematically introduce discrete choice models. The discrete choice model depicts individual preference using a utility function which includes attributes of the alternative route and characteristics of the individual. Considering heterogeneity among individuals, some parameters can be assumed to be distributed in the population. For example, one can argue that the value of time (VOT) factor is user-specific so that it should follow a distribution (e.g., lognormal distribution). Then for a given OD pair in a network, choice set generation techniques are first performed to generate alternative routes. Within the choice set, the driver prefers the route with the largest utility. However, utilities are perceived by users and thus deviate from systematic utilities. Random utility theory models the difference between perceived utility and systematic utility by an error term. Different distributions of the error term leads to different models, among which Path Size Logit model (Ben-Akiva and Bierlaire, 1999) are frequently applied in route choice applications. Parameters in route choice models often includes the mean and standard deviation of the assumed value-of-time distribution. Calibration of these parameters ensures the route choice model accurately captures the aggregate choices of drivers.

Other important behavior models are car following and lane changing models. Kim and Rilett (2004) applied genetic algorithms (GA) to calibrate driving behavior parameters.

2.1.3 Calibration of Supply Parameters

For DTA models that utilize a macroscopic supply simulator, link-based supply models that describe the relationships between speed-density and speed-flow are fundamental supply models.

Greenshields (1935) is the first researcher who systematically depicts the speed-flow relationship. But for link-based models, practically the number of speed-density function parameters can be huge. To address this problem, Van Aerde and Rahka (1995) proposed an approach to limit the scale of problem through classifying segments into groups based on their characteristics. Then for each group, the authors estimated group-specific supply models. This method showed compromised accuracy in predicting traffic of the entire network since the localized fits cannot reflect the entire network conditions. To solve the problem, Kundre (2002) proposed a three-stage procedure to calibrate the supply models. The three stages estimated supply parameters on levels of single segment, subnetwork, and entire network, which present increasing complexity. They simulated a large network and showed that simultaneous perturbation stochastic approximation (SPSA) algorithm provided results comparable to previous research, while reducing computation time. Following this work, Park et al. (2006) estimated speed-density relationships using a DTA model adopting methods proposed by Van Aerde and Rakha (1995). They concluded that adjustments on the calibration results are still needed to obtain reliable traffic prediction.

2.1.4 Joint Calibration of Demand and Supply Parameters

When applying a DTA model on real-world road network with unknown parameter values, joint calibration of demand and supply parameters are necessary. This process can be done either iteratively or simultaneously.

Darda (2002) proposed a framework for iteratively calibrate OD and other parameters. Parameters, including car following and lane changing models parameters in a microscopic simulator, were calibrated using Box-Complex algorithms given fixed demand. He further suggested calibration of OD demand can also be iterative when network size is large.

Gupta (2005) presented an iterative calibration framework to jointly calibrate OD, behavior and supply parameters with a mesoscopic DTA model. The OD demand was calibrated by the GLS method, while behavior and supply parameters are estimated with specific empirical methods.

It is generally agreed that iterative calibration cannot capture the complex interaction among different models of the simulation process, because calibration of each type of parameters are under the assumption of other parameters being known.

To address this shortcoming, simultaneous calibration was proposed and applied by Balakrishna (2006) to calibrate the mesoscopic DTA model. OD demand and other parameters are calibrated simultaneously with three algorithm, including Box-complex, SNOBFIT and SPSA. The case study indicated that SPSA outperforms the others in terms of fitting sensor count and speed, and simultaneous calibration is better than the iterative method in both efficiency and accuracy.

2.1.5 Summary

Since traffic simulation is performed with the interaction of demand and supply models, calibration of both models are necessary. Abundant researches have been conducted to address separate calibration of demand, behavior models, and supply models. However, when applying a DTA model to a real-world road network, simultaneous calibration of all its component models are necessary, and separate calibration of individual models may be both inefficient and suboptimal, due to the inability to capture the interactions between different models.

Although there is no consensus on performance indicator for evaluating the estimated parameters, a common practice is to use the normalized root mean square error (RMSN) of sensor measurements. In this study both traffic counts and speed are included in the measurement. The inclusion of speed enables the measurement to better reflect traffic conditions, and increases the reliability of calibration result, so that calibrated model can mimic the real world in terms of both speed and flow.

2.2 Algorithms for Calibration and Optimization

DTA model calibration is often formulated as an optimization problem which minimized some sort of deviation between the simulated and actual measurements. Therefore, an optimization algorithm is desired for solving such problems. Besides, the evolvement of traffic network state can also be formulated as a non-linear state-space model, and then the Extended Kalman Filter (EKF) method can be used to solve the calibration problem.

For offline simultaneous calibration of OD demand and other parameters, the Weighted Simultaneous Perturbation Stochastic Approximation (WSPSA) was found to be both accurate and computationally affordable.

2.2.1 Generalized Least Square Method

Generalized Least Square (GLS) is an algorithm for solving minimization problems with an object function in the form of sum of squared errors. Since calibration is often formulated as minimizing the sum of squared errors between simulated and actual measurements, it is straightforward to apply the GLS algorithm. However, to obtain the GLS estimator of the parameters, it requires estimation of a transformation matrix to linearize the relationship between measurements and parameters (i.e., measurement function). For OD demand calibration, assignment matrix can be used for this purpose, thus GLS is widely applied in OD estimation.

Toledo et al. (2004) proposed a framework to iteratively calibrate OD demand and behavior parameters of a microscopic traffic simulator, where the OD is estimated by GLS and other parameters are estimated by Box-complex algorithm. Gupta (2005) also proposed an iterative framework to calibrate a mesoscopic DTA model, where OD demand is estimated by GLS, and other types of parameters are estimated with specific methods.

Many of these research uses assignment matrix to linearize the measurement function. Use of assignment matrix assumes flow to be proportional to OD demand. However, in congested networks such assumption may not hold as the increase in OD demand may lead to more serious congestion which reduces the flow on parts of the network.

2.2.2 Extended Kalman Filter Method

The Kalman Filter Method formulates evolvement of network state as a state-space model, where the state of a road network refers to the demand and supply parameters of the network at a specific time.

The state-space model captures the evolvement of state through a transition equation, and uses a measurement equation to represent the relationship between measurements and the state (parameters). In case the measurement equation is linear, the Kalman filter method can estimate true state based on the true measurement and a priori state in one step. Otherwise the measurement function has to be linearized through a matrix, and extended Kalman filer (EKF) is applied for state estimation. The process may need to be iterative because the linearization matrix may change as estimated state vector changes.

Antoniou et al. (2004, 2007) applied EKF in online calibration of a mesoscopic DTA model. The measurement function is linearized by its local gradient, which is estimated with numerical method. Two gradient estimation methods, finite difference (FD) and simultaneous perturbation (SP) are tested and compared, and the proposed SP method outperforms the FD method in terms of computation time, without much loss of accuracy. Their difference in accuracy are further reduced with several iterations of the EKF.

2.2.3 Simultaneous Perturbation Stochastic Approximation (SPSA)

Simultaneous Perturbation Stochastic Approximation (SPSA) was proposed by Spall (1992) to efficiently solve a stochastic optimization problem. It uses SP, instead of FD, to compute an approximate gradient between objective function and parameters with O(1) computation effort, and uses gradient descent method to iteratively improve the objective.

SPSA, when applied in DTA model calibration, is a method that does not require linearization of the measurement function.

Balakrishna et al. (2007) applied SPSA in offline calibration of a DTA model. Demand and supply parameters are calibrated simultaneously, and the results suggested good performance of the algorithm.

2.2.4 Weighted SPSA

Despite recognizing SPSA as the most suitable algorithm for simultaneous calibration of OD and other parameters, Lu (2013) doubted whether SPSA could maintain similar convergence rate and calibration accuracy when the network has a larger scale and parameter vector has a much larger dimension. Tests affirmed his doubt, and it was found that, when performing gradient estimation with SPSA, a lot of irrelevant measurements would affect the gradient for a particular parameter. In light of this, he proposed the Weighted SPSA (WSPSA) algorithm, where a weight matrix is incorporated in gradient estimation. The weight matrix measures the correlation between measurements and parameters, and accurate estimation of the weight matrix is key to the success of WSPSA matrix.

Lu (2013) used assignment matrix as the weight matrix between sensor count measurements and OD demand, and an identity matrix as weight matrix between sensor count measurements and behavior parameters. The case study demonstrated plausible performance of WSPSA algorithm.

However, this study did not incorporate sensor speed measurements into the objective function. Inclusion of supply parameters may also require extension of the weight matrix.

Chapter 3

Methodology

This chapter presents the methodologies in different aspects of this study, including two examples of traffic simulators, the closed-loop testing framework, as well as the algorithms for calibration and the improvements and extensions made to the WSPSA algorithm.

3.1 Traffic Simulators

When studying or applying congestion management strategies, traffic simulators, including DTA models and microscopic traffic simulators are tools to simulate traffic conditions under different scenarios, and they are important for decision makers to understand any potential impacts of managements strategies. Any type of traffic simulator needs to be calibrated in order to represent the conditions on the real transportation network. This section presents two simulators which are used for our framework, and gives an overview of the embedded models and assumptions in these simulators, and how they simulate the real network.

3.1.1 DynaMIT

DynaMIT (DYnamic Network Assignment for the Management of Information to Travelers) is a mesoscopic DTA model developed in the Intelligent Transportation Systems (ITS) Lab of Massachusetts Institute of Technology (MIT). It is a simulation-based DTA that estimates and predicts traffic conditions, and when fully developed, generates traffic management strategies and provides consistent guidance information to travelers (Ben-Akiva, 2010).

DynaMIT consists of several modules (Figure 3.1), including state estimation, state prediction, strategy generation, etc. A state is a vector consisting demand and supply parameters. State estimation is the real-time process of incorporating an initial estimated state, historical data and real-time surveillance data to achieve a more reliable estimation of the current state. Online calibration is essential in the state estimation process in order to get an accurate estimate of the current state. It is the process of iteratively using measurements to "correct" an estimated state. In other words, we adjust the estimated state vector to make resulting traffic conditions converge to actual surveillance data. Online calibration is automatic when DynaMIT is running with input of real-time traffic surveillance information. However, when calibrating DynaMIT offline, this feature is intentionally turned off.

State prediction module predicts future states based on current state, taking into consideration any historical information, strategies (e.g., future toll rates) to be deployed and travelers' response to guidance information. The iterative process ensures that information to travelers will be consistent with their actual experience. Any real time control strategies such as toll optimization will be integrated into the strategy generation module of DynaMIT.

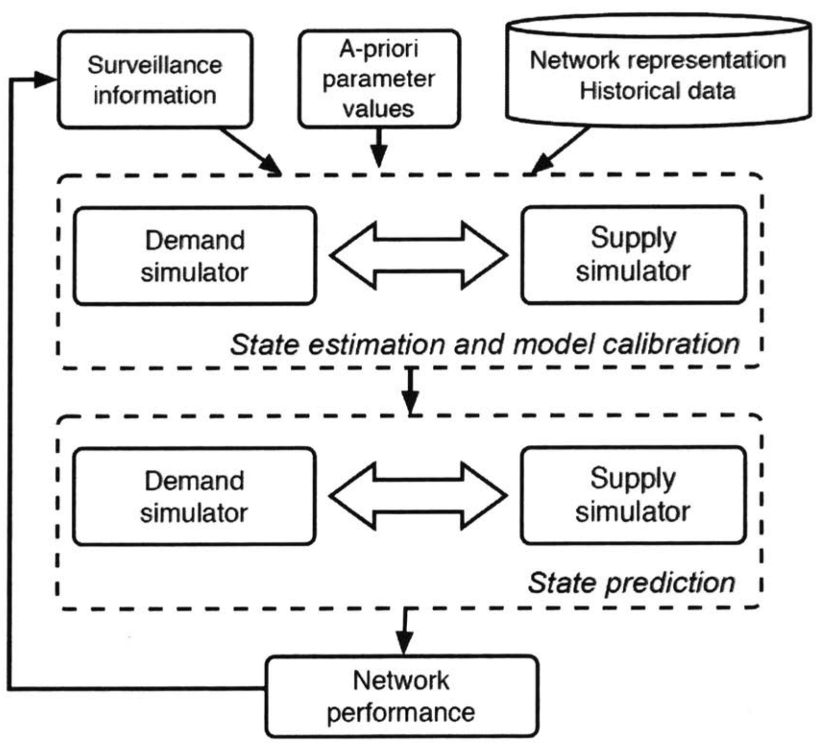

Figure 3.1 Modules in DynaMIT

3.1.2 MITSIM

MITSIM is a microscopic traffic simulator developed in the ITS Lab of MIT. MITSIM incorporates road topography, time-dependent OD demand, and other parameters, including driving behavior parameters and choice model parameters, and explicitly models drivers' route choice using discrete choice models, based on generalized cost which consists of travel time and tolls, and then simulates individual vehicle's movement and generates simulated sensor readings and other surveillance information.

MITSIM simulates individual vehicles, and uses a set of microscopic models to simulate the route choice decisions and driving behaviors of the drivers, which includes car following, lane changing, etc.

The route choice model in MITSIM uses discrete choice analysis to generate probabilistic route choice decisions of individual drivers. There are two types of models: link-based or path-based route choice models. In the path-based model, a set of pre-defined paths forms the choice set of a driver with specific origin and destination. The choice among paths are modeled with path-size discrete choice model, which takes into consideration the similarities between paths that are partially overlapping. A driver's probability of choosing a certain Path *i* is then

$$\mathrm{P(i)} = \frac{e^{V_i + \ln PS_i}}{\sum_{j \in C} e^{V_j + \ln PS_j}}$$

where V_i and V_j are systematic utilities of routes i and j, and PS_i and PS_j are the corresponding path sizes which capture drivers' perception of partially overlapping routes.

The systematic utility function of Path *i* can be represented in a simplified way as

$$V_i = \beta_{time} TT_i + toll_i$$

where β_{time} is the value of time (VOT) of a particular driver, TT_i is the drivers' knowledge of travel time on Path *i*, and $toll_i$ is the driver's knowledge of toll rates at the decision point.

Settings in MITSIM determine drivers' knowledge of travel time and toll rates. To better mimic the real-world, drivers are assumed to obtain travel time information from mobile navigation applications, which provide the current traffic conditions (i.e., travel time) on downstream links. As for toll rates, drivers are assumed to know the real-time toll rates only when they are close to the gantry. When a downstream gantry is distant from the route choice decision point, the driver uses the historical toll rates (at that time of day) to make decisions.

With these designs, MITSIM is capable of evaluating the impacts of various traffic management strategies, including changes of toll rates.

3.2 Closed-loop Testing Framework and its Calibration

The Simulation-based Dynamic Toll Optimization Scheme (Figure 3.2) is an example of utilizing a DTA model to automatically make congestion management decisions. It is designed in a rolling-horizon manner so that the estimation-optimization-prediction process repeats every 5 minutes (or other pre-defined length). For each 5-minute simulation interval, DynaMIT estimate current network state with real-world traffic surveillance data, then run iterations to find the optimized toll rates, and provide corresponding network state predictions for the next fifteen minutes (or other pre-defined length). The optimized toll rates are then implemented in the real world, and surveillance data are obtained for the next repetition of the process. (Wang, 2016)

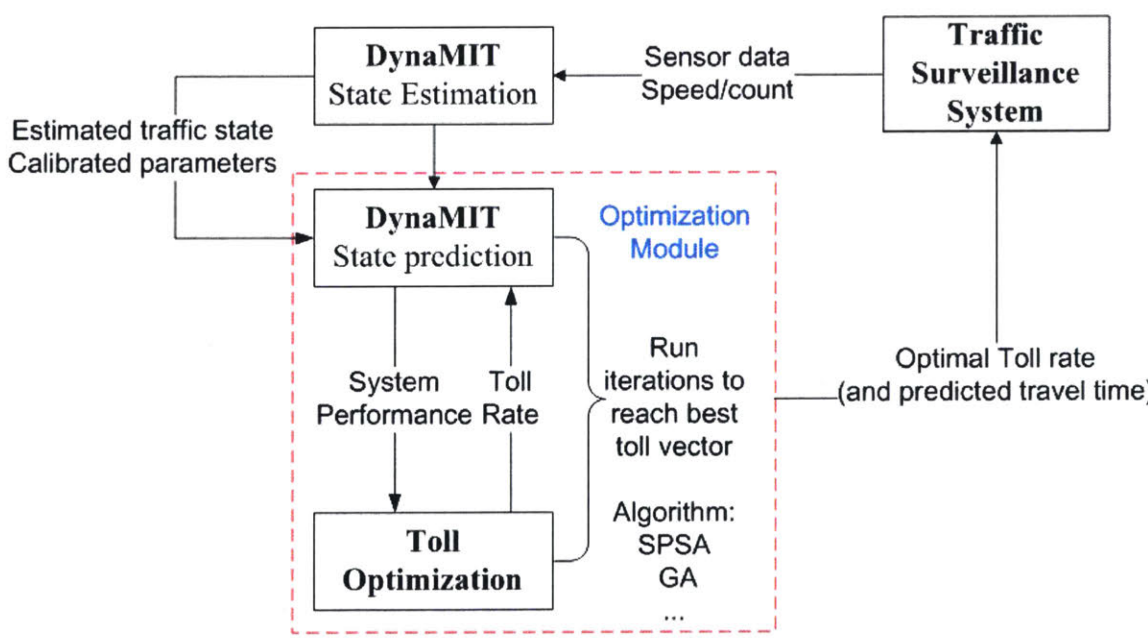

Figure 3.2 Simulation-based dynamic toll optimization scheme (Wang, 2016)

3.2.1 Closed-loop Testing Framework

Before the toll optimization framework can be implemented in the real world, the validity and performance of the models and algorithms developed need to be tested with another simulator. Based on the toll optimization framework, the closed-loop testing framework is developed by using MITSIM to replace the real world (Figure 3.3), and the performance of the toll optimization scheme could be tested and evaluated in MITSIM. In this testing framework, instead of implementing the optimized toll rates in the real world, they are implemented in MITSIM. In stead of obtaining sensor data from real world traffic surveillance system, DynaMIT obtains sensor data from MITSIM.

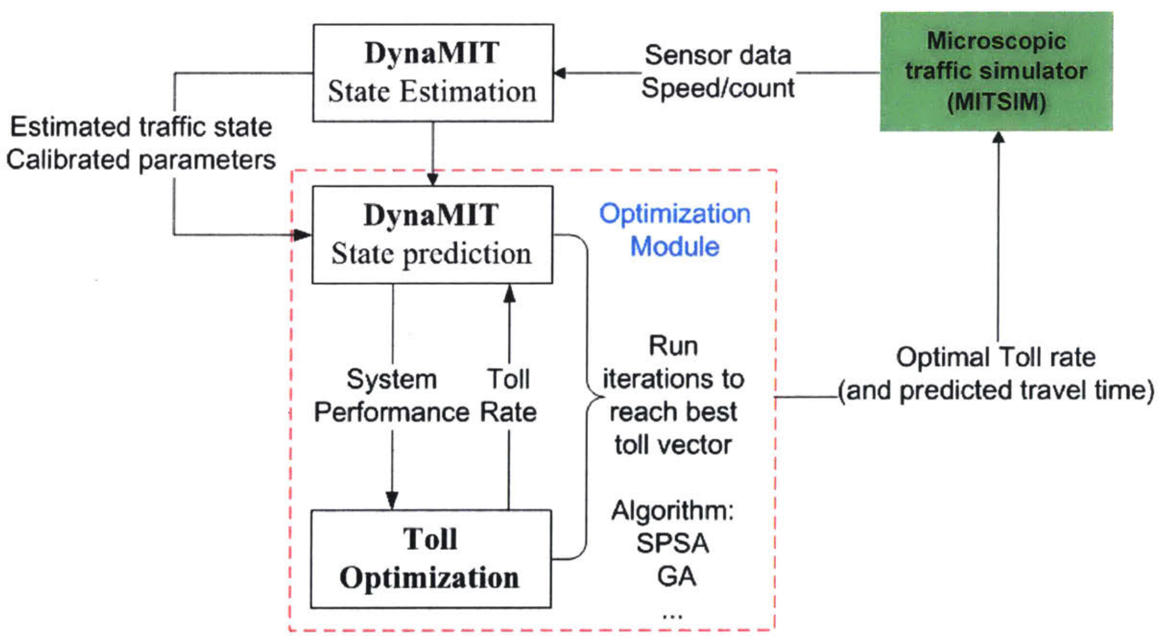

Figure 3.3 Closed-loop testing framework (Wang, 2016)

The closed-loop testing framework requires that the microscopic traffic simulator represents the real world accurately, i.e., drivers in the simulator behaves in the same way as those in the real world, and demand-supply interactions occur in the same way, etc. This can be achieved by calibrating MITSIM towards the real world.

3.2.2 Calibration of MITSIM

Calibration of a transportation network simulator is difficult due of its stochasticity and unobserved factors that are not captured by the simulator, as well as the randomness in real-world traffic conditions. Besides, the calibration algorithm needs to be highly efficient considering the heavy computational burden of the simulator.

As a microscopic traffic simulator, MITSIM takes demand, supply, and behavior parameters as input, and generates simulated measurements including sensor readings and travel time (Figure 3.4). Calibration of MITSIM is the inverse of this process, which searches for the best parameters that results in known measurements.

Figure 3.4 MITSIM simulation process

Parameters to be calibrated are selected depending on the needs of the specific application. For the closed-loop testing framework of toll optimization scheme, since toll rates are always explicitly known, they are not calibrated. A subset of behavior parameters are calibrated since they are most relevant to the problem. OD demands are also calibrated.

Simultaneous calibration of demand and other parameters is challenging because of their different magnitudes and complex interactions. In addition, incorporating multiple types of data (sensor count and speed) further complicates the problem.

3.2.3 Calibration of DynaMIT

In the closed-loop testing framework, calibration of DynaMIT towards MITSIM is also necessary so that the realized traffic conditions in MITSIM match the predicted traffic conditions in DynaMIT, and this ensures the validity of the optimization result.

Instead of online calibration which takes real-time input data, this calibration is offline. It is the process of adjusting the input parameters to DynaMIT so that the simulated measurements match the simulated measurements from MITSIM. The process is similar with MITSIM, except for the differences in the set of input parameters to be calibrated. Since DynaMIT is a mesoscopic DTA model, it doesn't have driving behavior models like car-following and lane-changing models. However, it requires a set of link-specific supply parameters which need to be calibrated, including free-flow speed, capacity, etc.

3.3 Formulation and Solution Algorithms for the Calibration Problem

3.3.1 Formulation as a Least-square Optimization Problem

The offline calibration problem is formulated as a minimization problem, with an objective function in the generalized least square (GLS) form, which measures the divergence between simulated and real-world traffic conditions:

$$\hat{x} = \mathrm{argmin}_x \, z$$

$$z = \left(y - h(x)\right)^T R^{-1}\left(y - h(x)\right) + (x - x^a)^T W^{-1}(x - x^a)$$

where

x is the parameters,

x^a is the initial estimate of the parameters,

$\hat{x}$ is the estimated parameters,

y is the real-world traffic conditions measurements, e.g., sensor readings,

$h()$ is the relationship between parameters and traffic condition measurements,

R is the variance-covariance matrix of traffic measurements,

W is the variance-covariance matrix of the parameters.

Since the variance-covariance matrices are often unknown, and the initial estimate is often not credible, the objective could be simplified as a normalized root mean square error (RMSN) form:

$$z = \sqrt{\sum_{j=1}^{M} (y_j - h_j(x))^2 / M} \Big/ \bar{y}$$

where

j is the index of measurement

M is the number of measurements

Gradient descent is a standard algorithm for this kind of optimization problem, where gradient can be estimated by finite difference in each iteration. However, finite difference requires a large number of h(x) evaluations, which is not practical due to the heavy computational burden of a simulator.

Simultaneous Perturbation Stochastic Approximation (SPSA) is an algorithm suitable for this problem due to its highly efficient gradient estimation approach (Spall, 1992).

3.3.2 The SPSA Algorithm

The SPSA algorithm is an iterative gradient descent method to solve a stochastic optimization problem. With an initial vector of parameters *x*, the algorithm perturbs the vector in two opposite directions to obtain x^+ and x^-. Then the simulator runs twice with the two vectors of perturbed parameters and obtain their corresponding objective function values z^+ and z^-. Then a gradient is calculated, and the vector of parameters is updated against the gradient direction by a certain step size:

$$x_i^+ = x_i + d_i$$

$$x_i^- = x_i - d_i$$

$$z = \sqrt{\sum_{j=1}^{M} (y_j - h_j(x))^2 / M} \Big/ \bar{y}$$

$$g_i(x) = \frac{z^+ - z^-}{2d_i}$$

$$x_{k+1} = x_k - a_k g_k$$

where

$d_i = \pm 1$ with equal probabilities

k is the index of iteration

a is the update size factor, a parameter of the algorithm

g_k is the gradient of *z* over *x*, calculated in the *k*-th iteration

i is the index of parameter

d_i is the perturbation size of parameter *i*

(Note that subscribe *k* is omitted expect for the last equation)

The SPSA is less accurate than finite difference due to its inaccurate way of calculating gradient, especially when parameters are highly correlated. However, it is a highly efficient O(1) algorithm, which makes offline calibration much more efficient and scalable. Besides, Spall (1992) pointed out that SPSA achieves comparable accuracy as finite difference with stochastic optimization problems, after sufficient number of iterations.

The SPSA algorithm is tested for calibrating MITSIM. However, the objective function did not improve significantly after 100 iterations. Although additional iterations might improve the result, that might not be possible because of the heavy computational burden of the simulator. An algorithm with higher convergence rate is needed.

3.3.3 The Weighted SPSA Algorithm

Opportunities for improving calibration efficiency with a transportation network simulator emerge from the fact that correlation between parameter and observation is sparse, since a particular time dependent sensor reading is affected by a small number of time-dependent OD, and this correlation matrix is observable through a traffic assignment matrix generated in the simulation process. The Weighted SPSA (WSPSA) algorithm was proposed by Lu (2013) to take advantage of this opportunity.

WSPSA adds a weight matrix in the gradient estimation process of SPSA, and effectively reduces the noise generated by irrelevant measurements.

$$z_i = \sqrt{\sum_{j=1}^{M} w_{ij}(y_j - h_j(x))^2 / M \Big/ \bar{y}}$$

$$g_i(x) = \frac{z_i^+ - z_i^-}{2d_i}$$

where

w_{ij} is an element in the (usually sparse) weight matrix, which represents the degree of correlation between parameter *i* and measurement *j*

z_i is the objective function associated to the i-the parameter

The WSPSA algorithm was proved effective in calibration to match simulated sensor count with real-world data (Lu, 2013).

However, new tests of the WSPSA algorithm identified that the accuracy of calibration result is not satisfactory, when sensor speed and count are both incorporated in the objective function. It indicated there are still rooms for improvement.

3.4 Improvements to the Weighted SPSA algorithm

The Weighted SPSA (WSPSA) algorithm is an iterative optimization process that tries to improve the objective function value in each iteration. Each iteration starts from initial parameters and generates updated parameters, and consists of the following steps: perturbing the parameters in two opposite directions, running the simulator to obtain sensor readings with perturbed parameters, computing the weight matrix, evaluating the objective function, computing the gradient matrix, and updating the parameters. The following sub-sections discuss the refinements of each step.

3.4.1 Perturbation of Parameters

According to Spall (1992), it is necessary to normalize all parameters to the same scale before perturbing them when using SPSA. Scale can be defined as either standard deviation (S.D.) or range. In the context of transportation network simulation, driving behavior and choice model parameters can be normalized to a number between 0 and 1 according to a priori information on each parameters' lower and upper bounds (ranges). However, normalizing the OD parameters is not straightforward. While lower bound of ODs can be assumed to be 0, upper bound of an OD should consider both the value of this OD and the overall magnitude of all ODs. While larger OD should have larger S.D. than smaller ones, it is not justifiable to assume a constant coefficient of variation.

It is shown that, if each parameter is perturbed by different sizes, it is equivalent to normalizing the parameters, as long as the different perturbation size is accounted for when computing gradient matrix. Therefore, normalization of OD is equivalent to setting different perturbation sizes for each OD.

Determination of perturbation size needs to consider the value of each OD, as well as the upper bound. In the practice of Lu (2013), ODs are perturbed with the same size (results presented as [A] in Figure 3.5), which is equivalent to scaling all the ODs by the same factor, assuming each OD has identical upper bound/S.D. In contrast, a ratio-perturbation method is also tried, which perturbs each OD by a percentage [C]. This is equivalent to scaling all the ODs by a factor that is proportional to their values, which assumes each OD's upper bound/S.D. is proportional to its value. A new method is proposed and tested, which outperforms previous methods in terms of convergence rate and accuracy. The new method perturbs each OD in a size proportional to the square-root of its value [B], which is equivalent to assuming an upper bound/S.D. proportional to the geometric average of this OD's value and the overall magnitude of all ODs.

The three perturbation method are tested in calibrating MITSIM with real-world sensor data, and their performance is shown (Figure 3.5). The three test runs are conducted with the same settings except for the perturbation size. MITSIM is calibrated with WSPSA algorithm, and the proposed method [B] achieved much better convergence rate than the others.

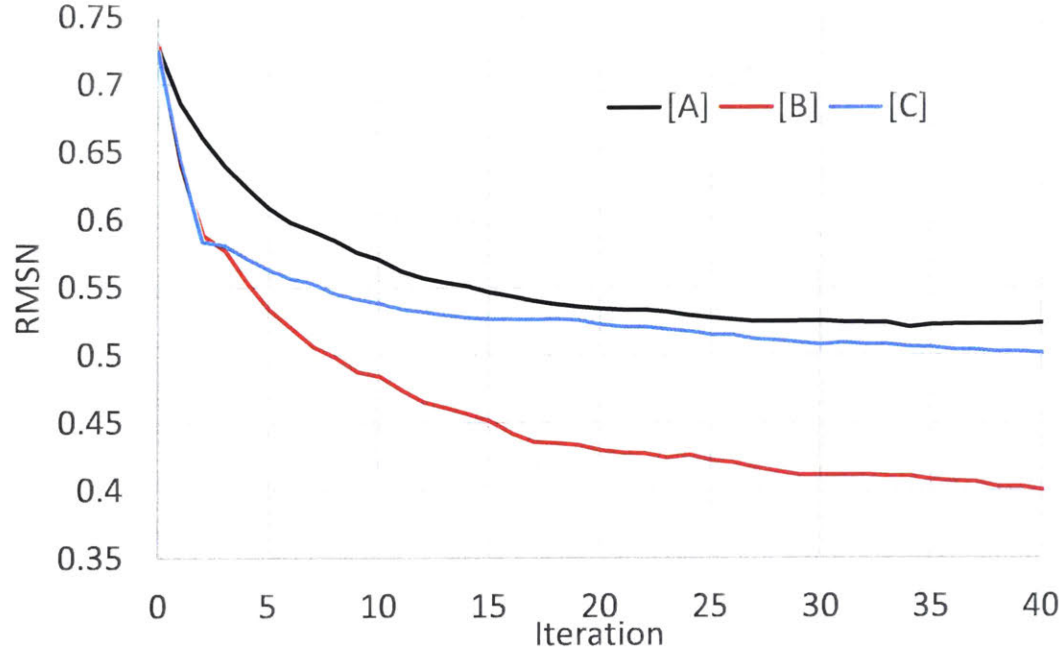

Figure 3.5 Performance of different OD normalization methods

3.4.2 Running the Simulator

To account for the stochasticity of the simulator, for each evaluation of h(x), the simulator runs multiple replications in parallel, and the calibration algorithm uses the average simulated sensor readings and average assignment matrices among multiple replications.

It is observed that sensor measurements in MITSIM deviates by several percent among different runs. Therefore, depending on the needs, either 16 or 4 replications of the simulator are used, which should reduce the stochasticity of simulated measurements by half or by three quarters.

3.4.3 Weight Matrix

In the WSPSA algorithm, weight matrix measures the correlation between parameters and measurements. It is the key component that makes WSPSA superior to SPSA, and accurate estimation of the weight matrix largely determines the efficiency of the algorithm.

Lu (2013) uses assignment matrix as an estimate of the weight matrix between OD and sensor counts, due to its easiness to obtain. However, when speed is included in measurements, assignment matrix would not be successful in measuring the correlations between OD and sensor speed readings, and a new method is proposed to generate a complete and reliable weight matrix.

In recognition that the assignment matrix can be scaled to estimate the gradient matrix between OD and sensor count, and the gradient between count and speed for each sensor can also be estimated by fitting a speed-flow relationship curve with historical sensor data, it is possible to estimate the gradient matrix between OD and speed.

For driving behavior and route choice parameters, it is also desirable to obtain the correlation matrix between those parameters and sensor readings. Since some parameters (e.g., value of time) only have impact on part of the network (e.g., toll lanes and free lanes parallel to toll lanes, but not ramps), this further reduces the noise compared to using a constant in this part of the weight matrix as in previous practice. The gradient between parameters and sensor readings can be obtained by finite difference before performing the calibration. However, since some parameters only have an impact under certain scenarios (acceleration rate of driver have large impact when flow is nearly saturated, but does not have an impact in free-flow state), it is necessary to re-estimate part of the gradient matrix during the WSPSA iterations when calibration of a particular parameter becomes ineffective.

Besides, if the objective function contains penalty terms for deviation from a priori parameters, the gradient matrix should also consider that.

<table>
<tr><th></th><th>Count</th><th>Speed</th><th>OD</th><th>Other parameters</th></tr>
<tr><td>OD</td><td>Assignment matrix (Scaled)</td><td>Assignment matrix * speed-flow derivative</td><td>Identity</td><td>0</td></tr>
<tr><td>Other parameters</td><td colspan="2">Gradient estimated by finite difference before calibration, and updated during calibration</td><td>0</td><td>Identity</td></tr>
</table>

Table 3.1 Components of the complete gradient matrix between measurements and parameters

After a full gradient matrix is obtained, which will be normalized before being used as weight/correlation matrix. A contribution matrix is computed by multiplying the gradient matrix with the perturbation size of each parameter, which represents the contribution of each parameter to the change of each measurement. It is normalized by measurement. The normalized contribution matrix is then divided by the perturbation size, which serves as a better weight matrix to measure the correlation between parameters and measurements.

The new method of constructing weight matrix is tested in calibrating MITSIM with real-world sensor date, and shows much higher convergence rate than the original method (Figure 3.6).

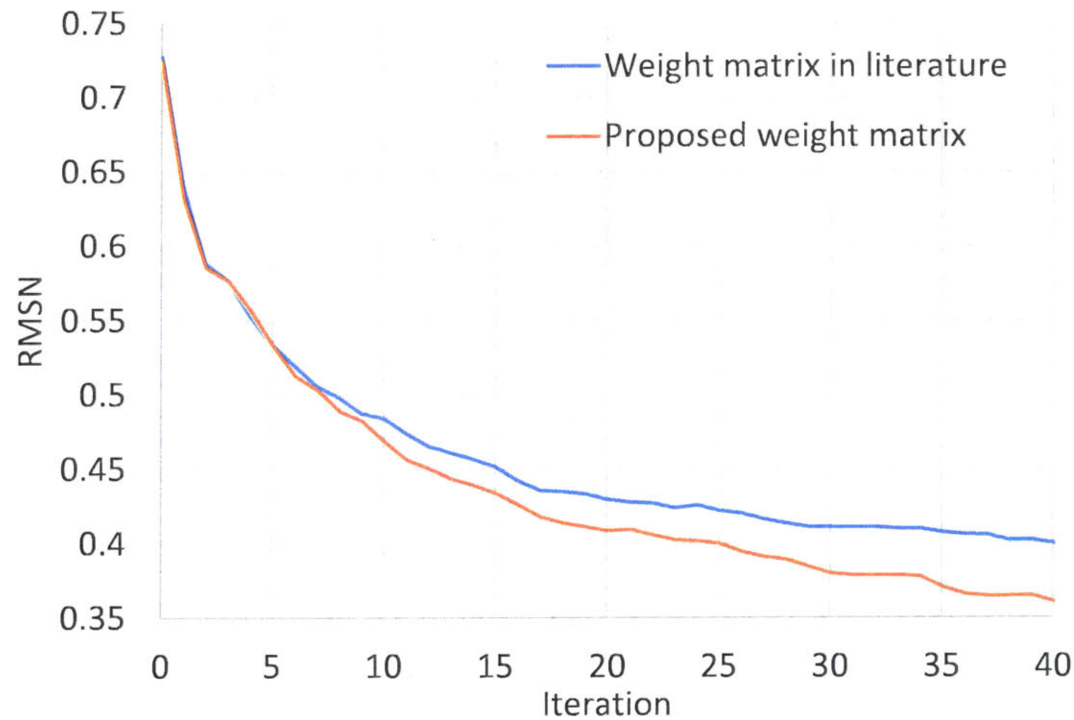

Figure 3.6 Performance of different weight matrices

3.4.4 Objective Function

To account for the randomness in real-world traffic conditions, sensor data from multiple days are averaged to represent an "average day", and the "average day" is calibrated to obtain the parameters that represent the "average day". Then each particular day can be calibrated using the parameters calibrated with the "average day" as initial values, and this calibration should include a penalty on deviation from the initial parameters.

To accommodate for the above calibration framework, a weighted least square (WLS) objective function, instead of RMSN, is proposed:

$$z = \sum_{j=1}^{M} \frac{\left(y_j - h_j(x)\right)^2}{\sigma_j{}^2} + \sum_{i=1}^{N} \frac{(x_i - x_i^a)^2}{q_i{}^2}$$

where

M is the number of measurements;

y_j is the j-th element of real-world measurement;

$h_j()$ is the j-th element of simulated measurement;

σ_j^2 is the S.D. of the j-th element of real-world measurement;

N is the number of parameters;

x_i is the i-th element of estimated parameter;

x_i^a is the a priori value of the i-th parameter;

q_i^2 is the inverse weight for the penalty on deviation of the i-th parameter.

The penalty term is not included when calibrating the average day.

Larger weights are given to sensor measurements that are consistent among multiple days, while smaller weights are given to measurements that vary among days. This helps average day calibration to match measurements that have less variation. Sensors with varying measurements among days are matched when calibrating particular days.

3.4.5 Gradient Matrix

For SPSA, gradient matrix measures how parameters affect the objective function, while for WSPSA, gradient vector of each parameter measures how this parameter affects the portion of objective function that is correlated to this parameter. As a result of modifications introduced in previous sections, calculation of gradient matrix can be re-written as the following:

$$g_i(x) = \frac{z_i^+ - z_i^-}{2d_i}$$

where

g_i is the estimated gradient vector of parameter i;

z_i is the portion of objective function that is correlated to parameter i;

+ and – denotes the two perturbations with opposite directions, and the resulting two simulation runs;

d_i is the perturbation size for parameter i.

$$z_i^+ = \sum_{j=1}^{M} w_{ij} \frac{\left(y_j^+ - h_j(x^+)\right)^2}{\sigma_j^2} + \sum_{n=1}^{N} w_{i(M+n)} \frac{(x_n^+ - x_n^a)^2}{q_n^2}$$
$$= \frac{(x_i^+ - x_i^a)^2}{q_i^2} + \sum_{j=1}^{M} w_{ij} \frac{\left(y_j^+ - h_j(x^+)\right)^2}{\sigma_j^2}$$

where

w_{ij} is the (i,j)-th element of the weight matrix. Note the weight matrix has dimension of N by (M+N).

3.4.6 Update of Parameters

The SPSA algorithm updates parameters by a size proportional to the gradient, in a gradient descent manner:

$$x_{k+1} = x_k - a_k \hat{g}_k(x_k)$$

where

a_k is the update size factor in the k-th iteration, a scaler that decreases with iterations.

Considering that the ideal outcome of calibration is to obtain an objective function value of zero, a Newton-update method is proposed, which assembles Newton's root-finding algorithm:

$$x_{i(k+1)} = x_{ik} - b_k \frac{z_i}{\hat{g}_{ik}(x_k)}$$

where

b_k is the update size factor in the k-th iteration. It takes a scaler value within (0,1).

Inclusion of the update size factor b_k avoids large update size when some parameter already matches the true value very well, while the standard update method for SPSA may have this problem. In the Newton-update method, when gradient is accurate, a larger b_k is desirable to improve convergence rate, while a smaller b_k helps avoid overly updating the parameters when gradient estimate is not good.

Another proposed update method is to update each parameter at a size equal to the perturbation size, to the direction that improves the objective correlated to that parameter:

$$x_{i(k+1)} = x_{ik}^{+} \text{ if } z_i^{+} = \min(z_i, z_i^{+}, z_i^{-});$$

$$x_{i(k+1)} = x_{ik}^{-} \text{ if } z_i^{-} = \min(z_i, z_i^{+}, z_i^{-});$$

$$x_{i(k+1)} = x_{ik} \text{ if } z_i \ = \min(z_i, z_i^{+}, z_i^{-});$$

This method benefits from not updating the parameter that does not tend to improve the objective.

Other updating methods are also explored. In recognition that the objective function is approximately a quadratic function of OD parameters, a new method fits z_i and x_i with quadratic functions and updates x_i to the point with smallest z_i. But this does not show clear improvement.

All three methods are implemented within a single framework, and objective function values corresponding to the three update methods are computed, as well as the objective from the two perturbation runs. The five objective values are then compared, and the parameters that lead to the best objective is chosen as input to next iteration. Since each method has an advantage in different scenarios, it is observed that each method could be the best in different iterations.

Update method	Updated parameters	Objection function value
Gradient descent	x_{k+1}^{G}	Obtained from additional runs of simulator
Shift update	x_{k+1}^{S}	
Newton update	x_{k+1}^{N}	
Plus perturbation	x^{+}	Available from existing simulator runs
Minus perturbation	x^{-}	

Table 3.2 Different update methods

The three update methods, as well as the combined method are tested for calibrating MITSIM with real-world data, and their performances are shown in terms of convergence of objective (Figure 3.7). It is observed that the combined method outperforms others, and the gradient descent method is the worst in this scenario.

It should be noticed that performance of the update methods highly depends on the problem settings, and there is still no evidence to conclude which is the better or not. However, in our setting we find the performance of the combined method more promising.

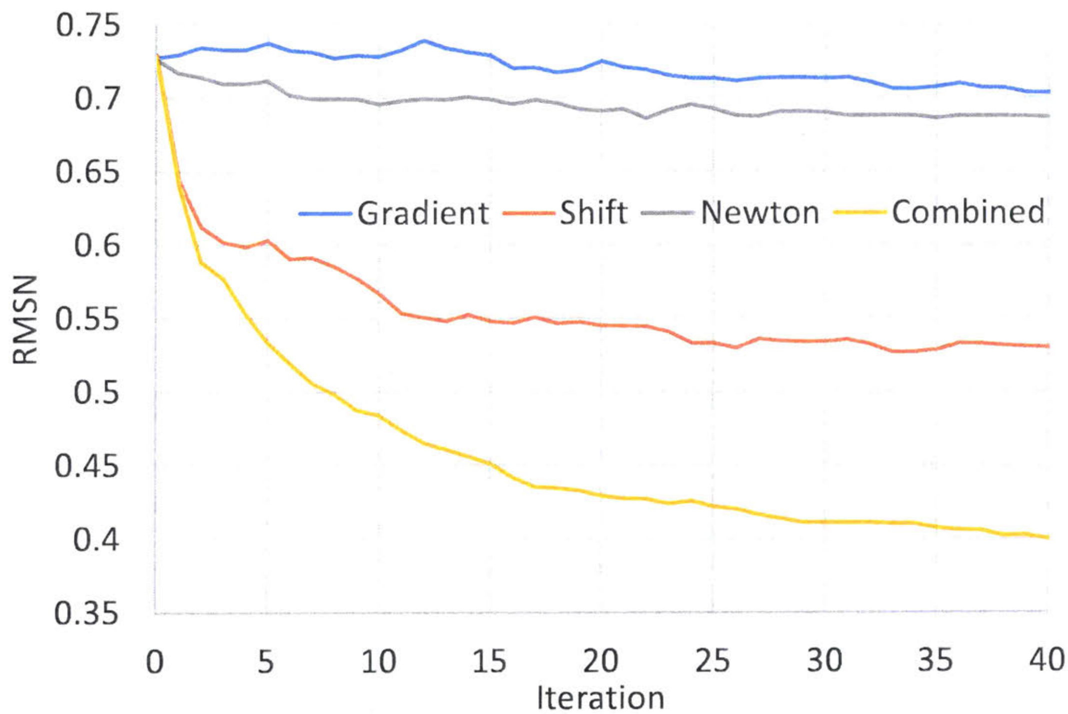

Figure 3.7 Performance of different update methods

Chapter 4

Case Study

This chapter presents a case study where the proposed methodologies are applied in a real-world road network. An overview of the network is provided, followed by detailed description of the implementation, as well as empirical considerations specific to this network. Results are presented, which shows good performance of the algorithm.

4.1 Overview of the Network

The refined WSPSA algorithm is tested with real-world data from a road network that consist of an 11-mile corridor with parallel managed lanes and general purpose lanes (Figure 4.1).

For analysis and presentation purposes, the westbound of the network is divided into 9 Parts, according to the location of entry and exit ramps on the managed lanes. There are no ramps within each Part on the managed lanes, but there could be intermediate ramps on general purpose lanes.

The network has 246 OD pairs, 135 sensors and 11 gantries. Each sensor records traffic flow and speed, while gantries also have provides traffic flow data through transaction record. Therefore, there are 135*2+11=281 measurements in each time interval.

Dynamic tolling is currently in place on the managed lanes. The toll rates are adjusted in real-time in response to observed traffic conditions, in order to keep traffic speed on managed lanes higher than 50mph. The average toll rates are shown for illustration (Table 4.1)

Mon	Tue	Wed	Thu	Fri	Sat	Sun	
6:00 AM	1.79	1.89	1.92	1.86	1.86	0.95	0.95
6:30 AM	2.34	2.47	2.50	2.49	2.44	0.95	0.95
7:00 AM	2.77	2.96	3.00	3.04	2.91	0.95	0.95
7:30 AM	2.83	3.04	3.06	3.03	2.97	0.96	0.95
8:00 AM	2.84	3.12	3.09	3.10	3.00	0.98	0.95
8:30 AM	2.81	3.09	3.08	3.06	2.99	0.96	0.95
9:00 AM	2.42	2.62	2.59	2.53	2.54	0.95	0.95
10:00 AM	2.15	2.32	2.30	2.25	2.25	0.95	0.95
11:00 AM	2.20	2.30	2.29	2.25	2.26	1.47	1.23

Table 4.1 Dynamic toll rates on westbound of the test network

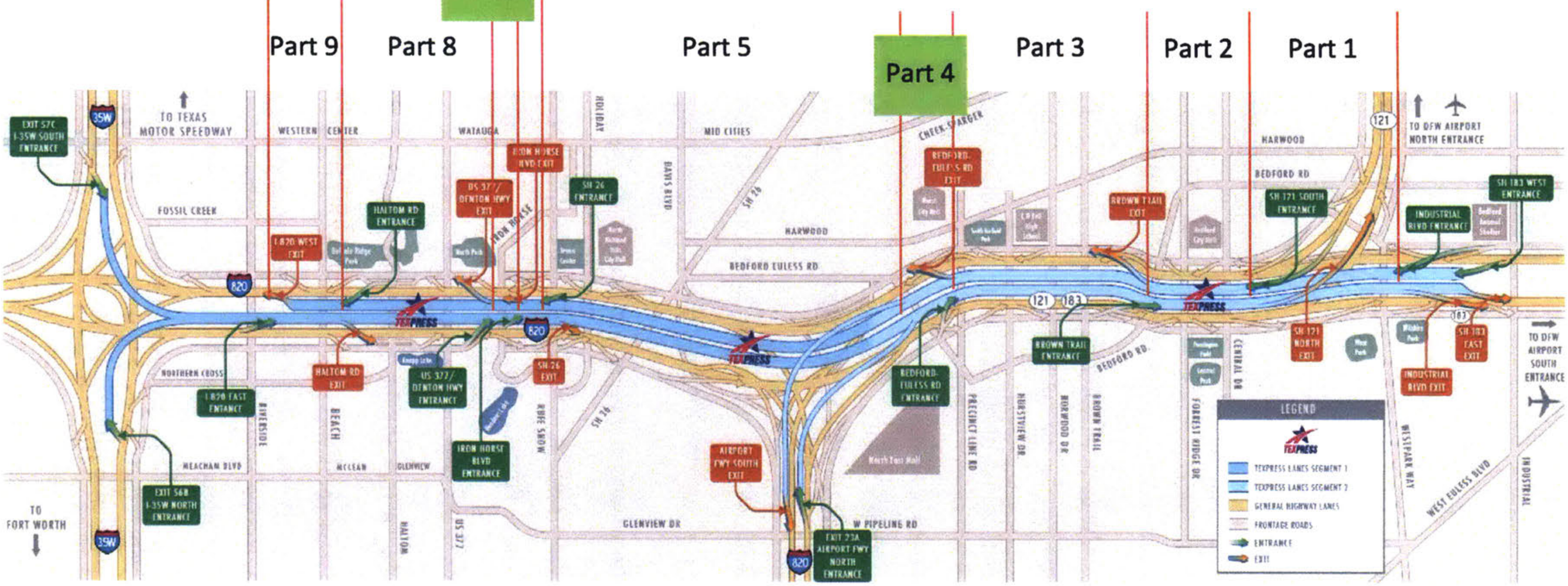

Figure 4.1 The test network, and its 9 Parts on westbound

4.2 Preparation for Calibration

The simulation horizon in this case study is 3 hours, from 6:30am to 9:30am. The first half hour is a warm-up period for vehicles to be loaded onto the network. The time-dependent OD is defined in 15-minutes intervals, so there are 12 time intervals in terms of OD, while the sensor readings are aggregated to 5-minute intervals, thus there are 30 time intervals in terms of measurements, since measurements during the warm-up period are not taken into consideration.

4.2.1 Input Data

Since a simulation-based dynamic toll optimization scheme is being developed with this network, calibration of OD and other parameters is necessary before testing any optimization strategy in the closed-loop framework. The operator of this managed lane system has provided the network topography information, in addition to sensor readings, toll rates and transaction counts for 5 consecutive weekdays obtained in June 2015. Automatic vehicle identification (AVI) data are also available for some other days.

In order to calibrate an "average day", the average of the data among 5 days are obtained, and the variances of each measurement is calculated. The flow data are manually checked against flow-conservation laws, then a number of sensors are identified to have provided abnormal flow readings, and thus all flow readings obtained by those sensors are marked as invalid.

4.2.2 Selection of Parameters to be Calibrated

The microscopic simulator MITSIM has hundreds of behavior parameters that control a series of models to represent driver's route choice and driving behaviors, and to simulate vehicle movements. Most parameters have credible values either from previous research or from highway design manuals. Besides, many parameters may not have a notable impact on the measurements, and are irrelevant to this study. Therefore, 27 driving behavior and route choice parameters are chosen to be calibrated. Those are key parameters that largely affect traffic conditions in this network, which include mean and standard deviation of the distribution of drivers' value of time, their average reaction time, scaling factors to measure their acceleration and deceleration aggressiveness, etc. In this study, these parameters are considered not time-dependent.

So there are 30*281=89430 measurements and 12*246+27=2979 parameters in this calibration problem (Table 4.2).

Number of OD pairs	246	Number of sensors	135
Total number of OD parameters	2952	Number of gantries	11
Number of other parameters	27	Number of measurements per interval	281
Total number of parameters (N)	2979	Total number of measurements (M)	89430

Table 4.2 Characteristics of the calibration problem

4.3 Preliminary Results

For easiness of understanding, and to have a consistent standard for comparison among different objective functions, this chapter uses RMSN, instead of the actual objective function value to present calibration results.

Figure 4.2 shows the convergence of RMSN when calibration starts from an all-zero seed OD.

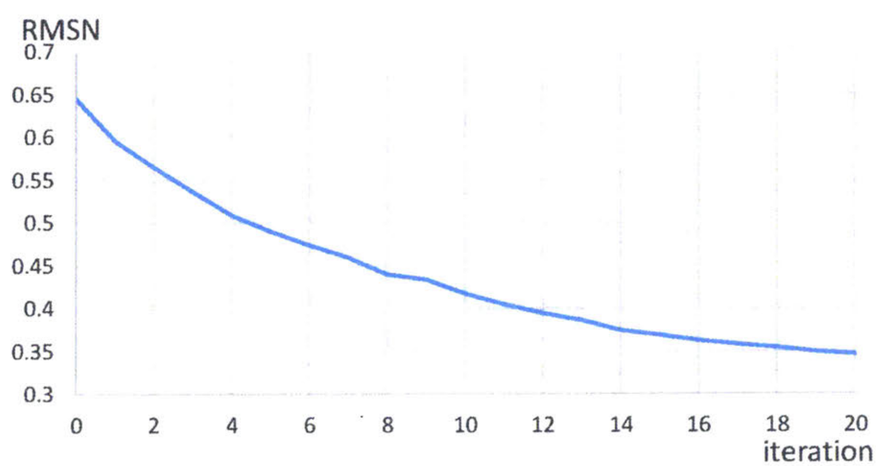

Figure 4.2 Convergence of RMSN from all-zero seed OD

The all-zero seed OD, together with the initial value of behavior parameters, corresponds to a 64% difference between simulated and true sensor measurements. After 20 iterations, the RMSN reduced to 35%, which still has room for improvement.

The preliminary results show that the algorithm is efficient when parameters are still far from their true values. However, after some iterations, convergence slows down, and a satisfactory accuracy cannot be reached.

Then the calibrated parameters and measurements are manually inspected, and the next section discusses about some insights on improving calibration accuracy through some modifications to the modeling approaches.

4.4 Empirical Considerations

When calibrating a model with real-world traffic data, it is necessary to identify unobserved factors that affect traffic conditions, when we notice some traffic condition measurements cannot be captured by simulator after sufficient calibration effort. For example, there might be additional disutility for drivers to switch from general purpose lane to toll managed lane through a ramp (Figure 4.3 red circle), compared to those who enter the managed lane directly (yellow circle). Besides, after inspection satellite photos, it is identified that entering managed lanes from some particular ramp (purple circle) involves additional inconvenience, due to additional detour and turns which cannot be included in our road network. These effects are then modeled with an additional cost on that ramp, which is calibrated altogether with other parameters. This treatment significantly improves the accuracy of calibrated parameters.

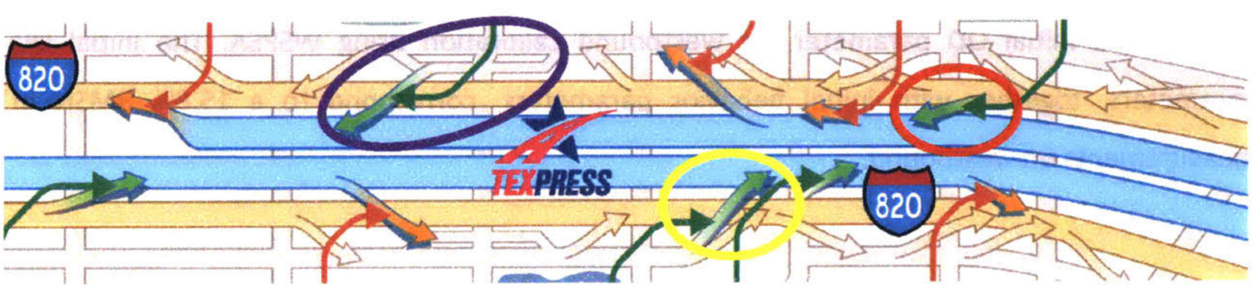

Figure 4.3 Examples of inconvenient merging to managed lanes

Besides, the simulation also reveals that in real world, drivers who are already on the managed lanes tend to continue on the managed lanes, even if our choice model indicated switching to the general purpose lane is more preferable. This inertia is model as a Bonus Utility at the next gantry (Figure 4.4), so that discrete choice analysis can correctly model their route choice in the simulator. The bonus utility only affects drivers who are already on upstream managed lanes, and are making a decision whether to exit to general purpose lanes or to continue on the managed lanes. This modeling approach is implemented for both eastbound (Figure 4.4) and westbound (not shown in figure) on the test network, and improved calibration result.

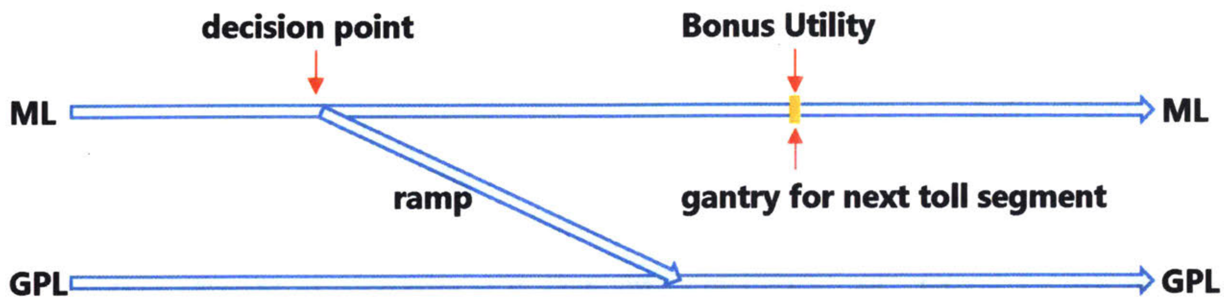

Figure 4.4 Bonus Utility to model inertia on switching

The above empirical modeling approaches are incorporated in the simulation model, and further improvements are developed based on the new model.

In the context of a transportation network, there is other information to help with calibration. Automatic vehicle identification (AVI) data give direct information on OD demand, and can be scaled by a penetration ratio to generate a seed OD for better calibration (Wang, 2016).

A westbound seed OD generated with AVI data is provided by the managed lane operator, and is used as the initial OD parameter for westbound calibration using WSPSA. The initial OD parameters, together with initial behavior parameters, correspond to a 15.8% RMSN in measurements on westbound.

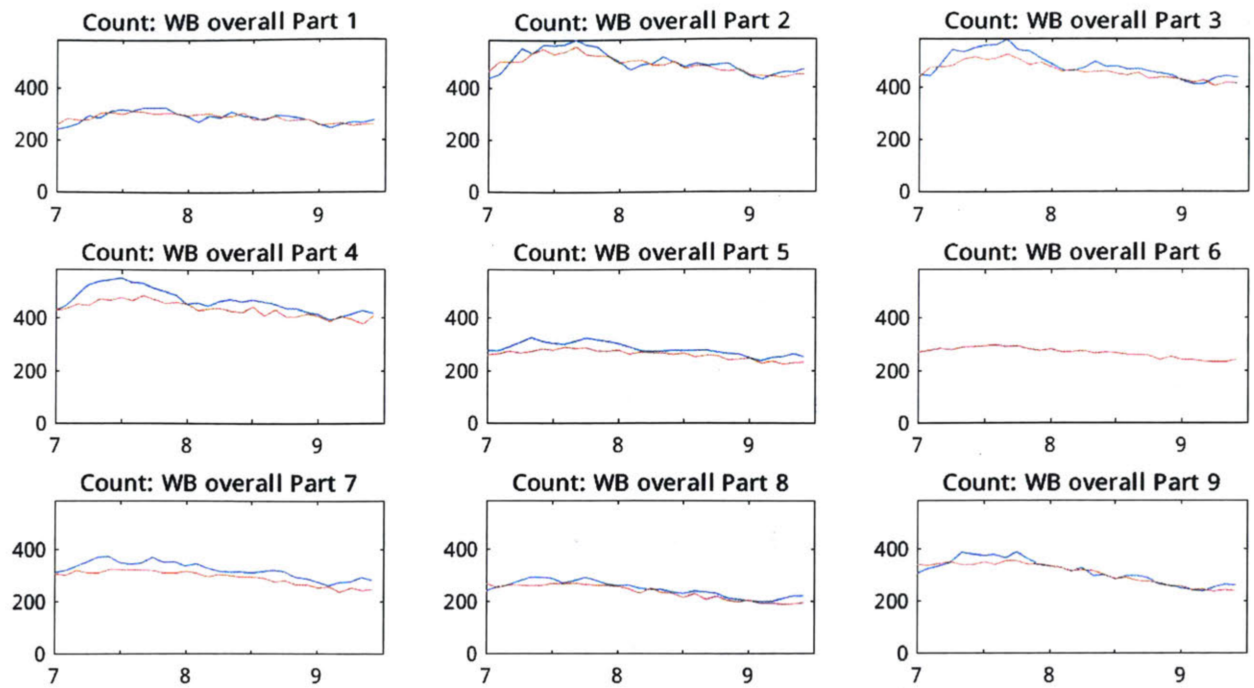

Figure 4.5a WSPSA calibration results w.r.t. overall sensor count

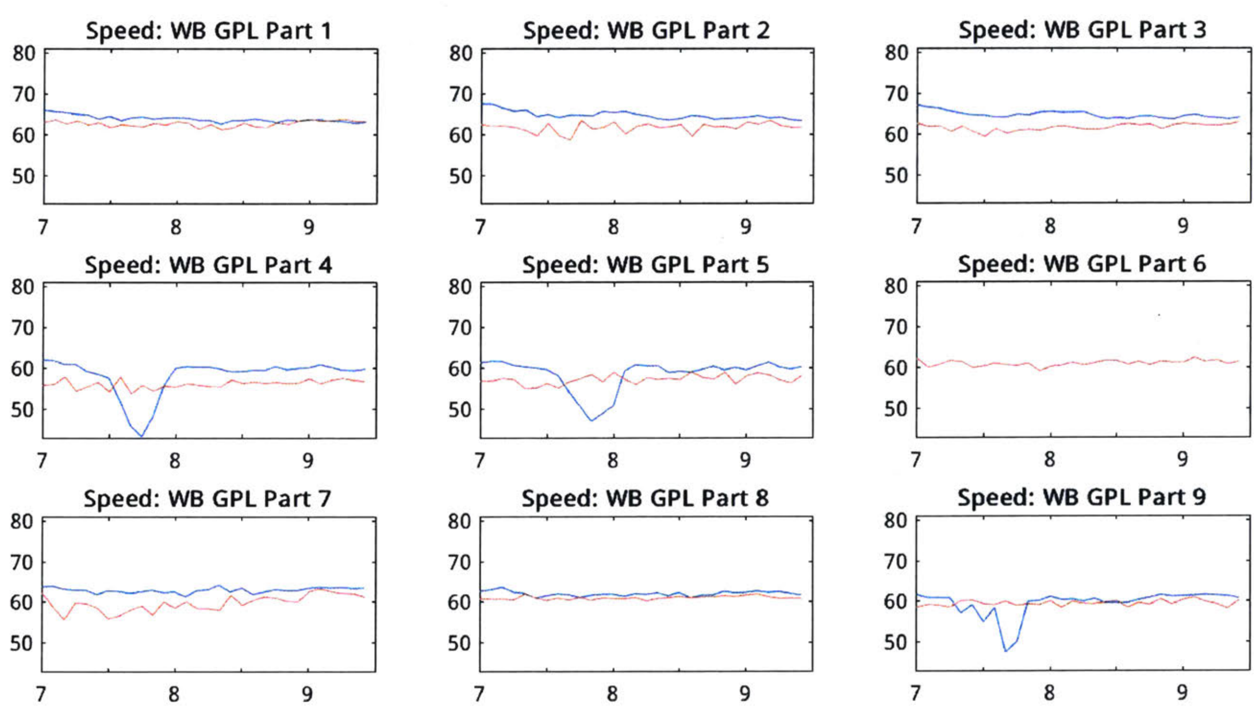

Figure 4.5b WSPSA calibration results w.r.t. sensor speed on GPL

Based on this seed OD, after 20 iterations of WSPSA calibration on westbound, the RMSN reduced to 8.9%. There is no further improvement with more iterations. The plots (Figure 4.5) show the calibration results with respect to overall sensor count and GPL speed on the test network.

The 9 plots correspond to 9 Parts of the network, from upstream to downstream. Y-axis values are sensor counts (vehicles per time interval) or average speed (mph) in 5-minutes intervals, and X-axis is the starting time of each time interval, excluding the warm-up period (7:00-9:25). Red curves are the simulated measurements, while blue curves are field observations.

It is observed that there is still significant deviation from the true values at certain parts of the network, especially the speed reduction between 7:30 and 8:00 in Part 4, 5 and 9. The calibration rather generates a bit lower speed, instead of capturing the exact trend of speed profile, which highlights the need for further inspection.

Components of the objective function is reviewed and analyzed manually in a case by case manner, and changes to the parameters are made according to real-life knowledge to the network behavior. The main change that improved the result was the inclusion of acceleration/deceleration scaling factors into the set of parameters to be calibrated. Higher value of these parameters enables the simulator to represent larger disruption of merging and diverging vehicles to the overall traffic conditions.

These changes further improve the calibration result to 7.7% in RMSN.

4.5 Final Results

Westbound final calibration results are presented in the plots (Figure 4.6).

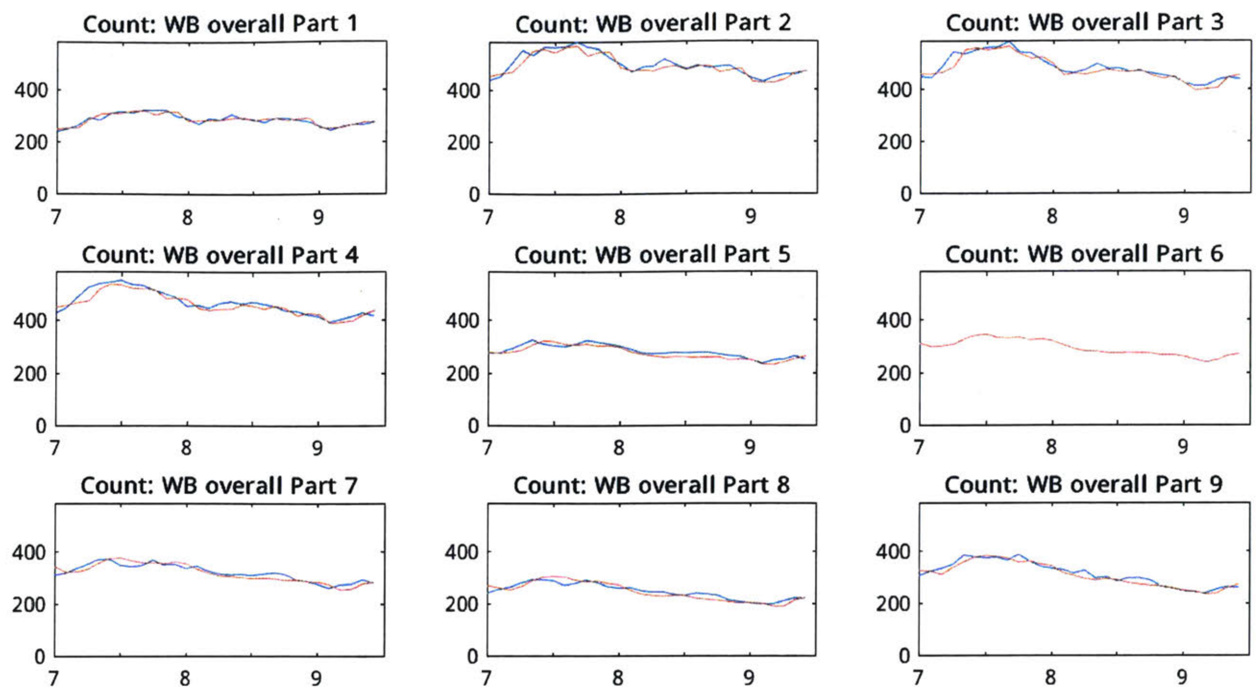

Figure 4.6a Final results w.r.t. overall sensor count

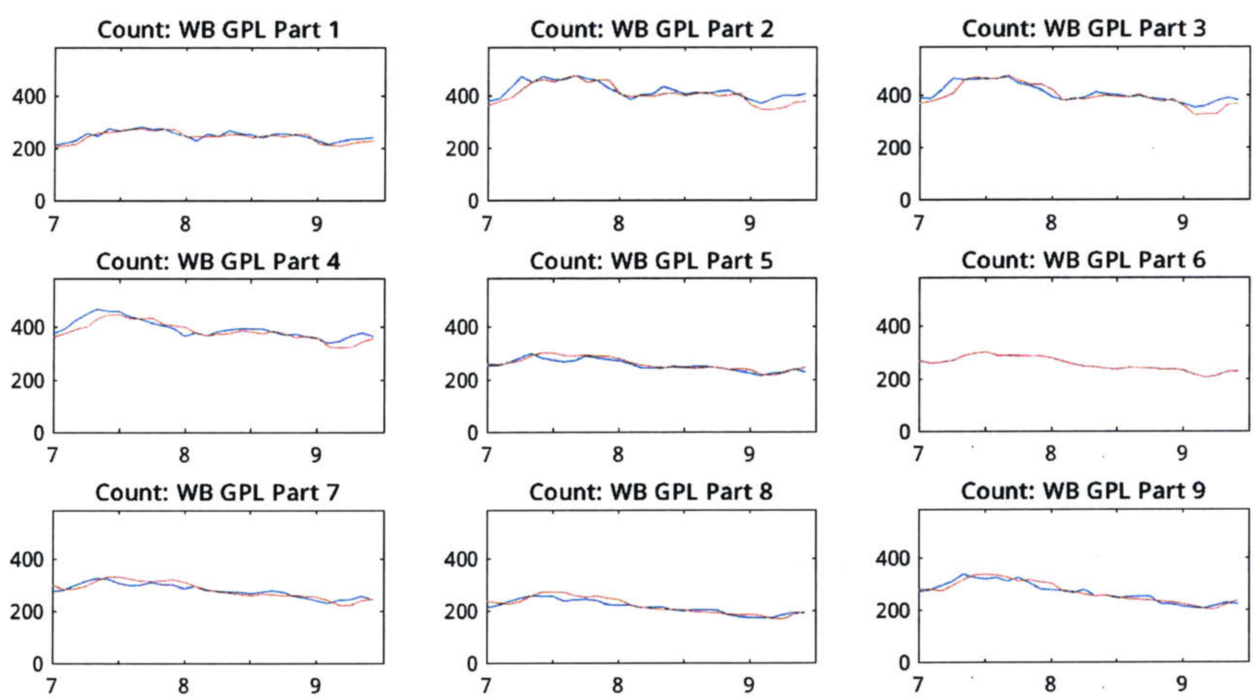

Figure 4.6b Final results w.r.t. sensor count on GPL

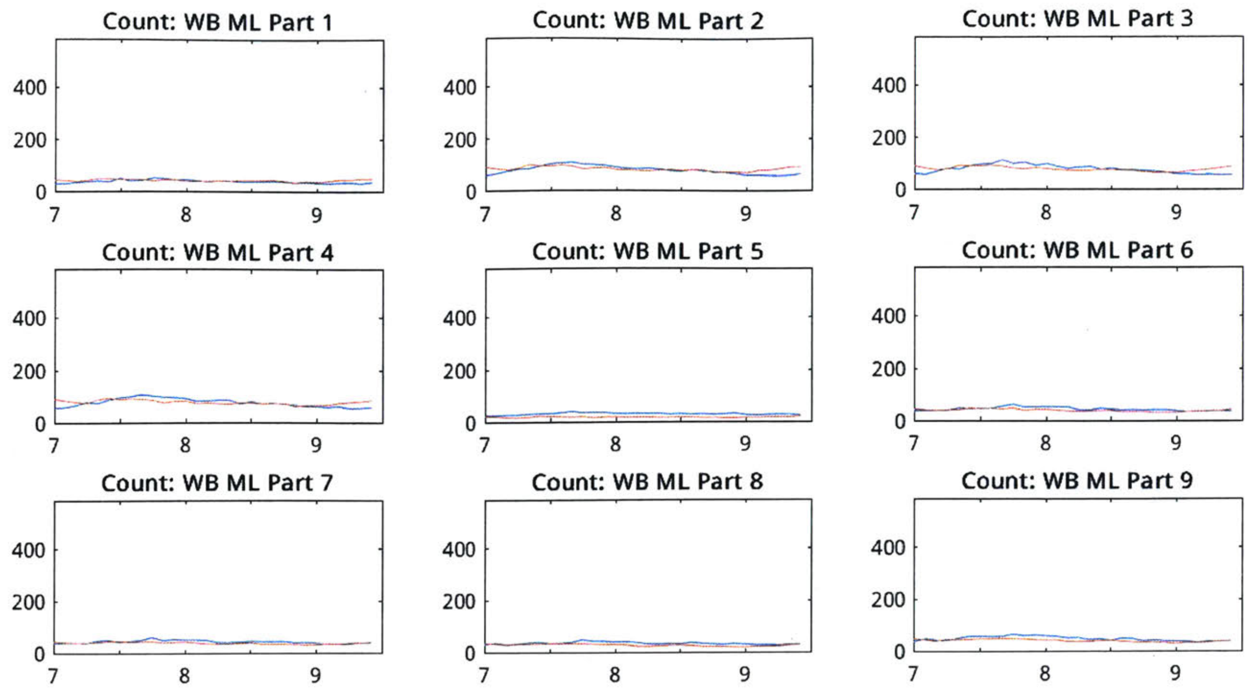

Figure 4.6c Final results w.r.t. sensor count on ML

The calibrated parameters are to be used in the closed-loop testing framework of the toll optimization research, and the accuracy of these parameters is essential to ensure the validity of the closed-loop testing results.

The overall accuracy of the simulator can be measured by the RMSN, which is 7.7%. This result indicates that the simulator closely replicates the real world.

The plot of overall traffic flow indicated that accurate time-dependent OD demand has been successfully obtained, thus the simulated flow on each Part of the corridor is accurate, except for some small discrepancies on Part 2 to 4 at 7:10 and 9:10.

The two plots of traffic flow on general purpose lanes and managed lanes both indicates that the simulator correctly models travelers' choice between the two options, thus the split of flow between GPL and ML is accurate.

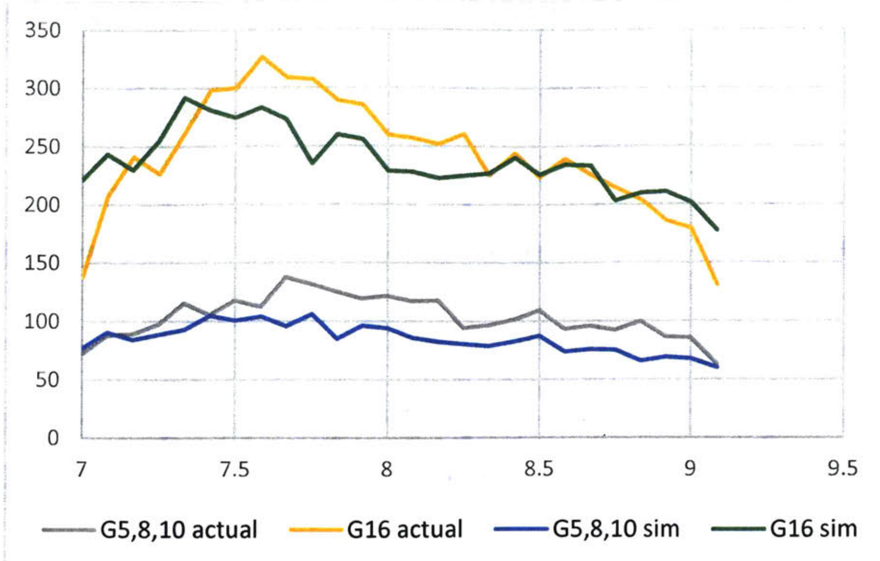

Figure 4.6d Final results w.r.t. toll revenue

The comparison of true and simulated toll revenues collected at 4 gantries on westbound is presented in Figure 4.6d, where y-axis is the revenue in dollars during each 5-minute time interval. Revenue for Gantry 5, 8 and 10 are summed up because they belongs to the same tolling segments.

The deviation between simulated and actual revenue is the deviation of flows on the corresponding part of the managed lanes, scaled by the toll rates at each time interval. For example, the revenue at Gantry 16 follows the flow on managed lane Part 2.

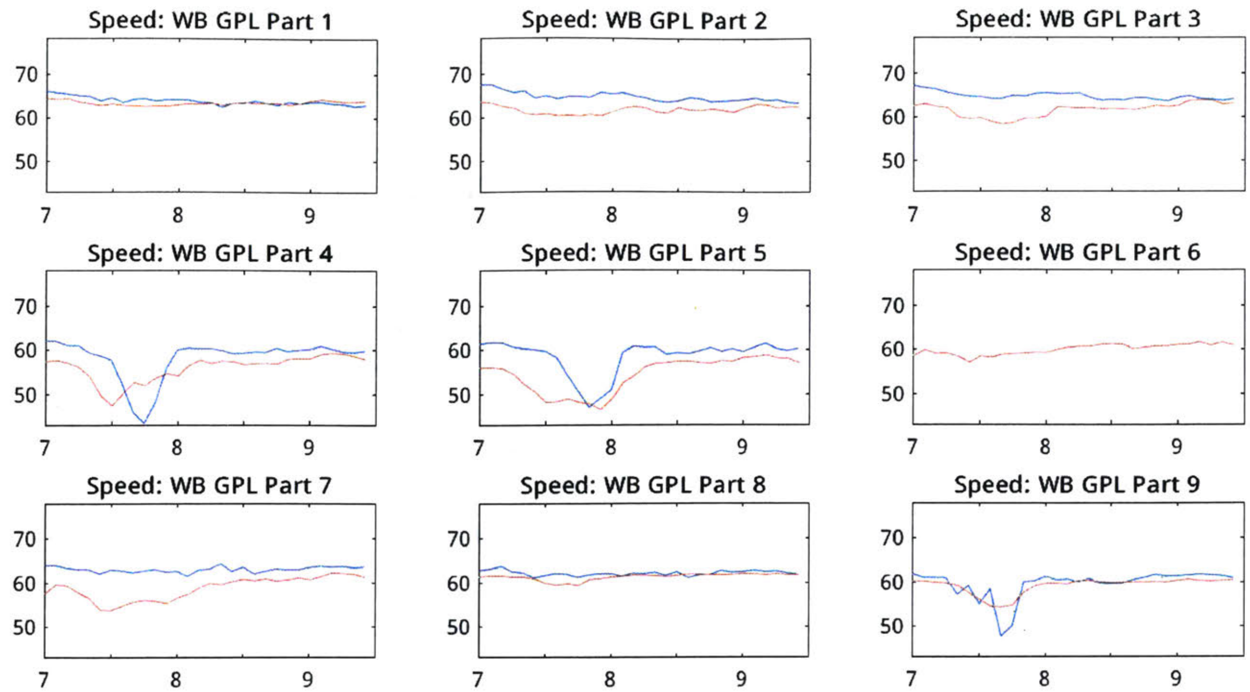

Figure 4.6e Final results w.r.t. sensor speed on GPL

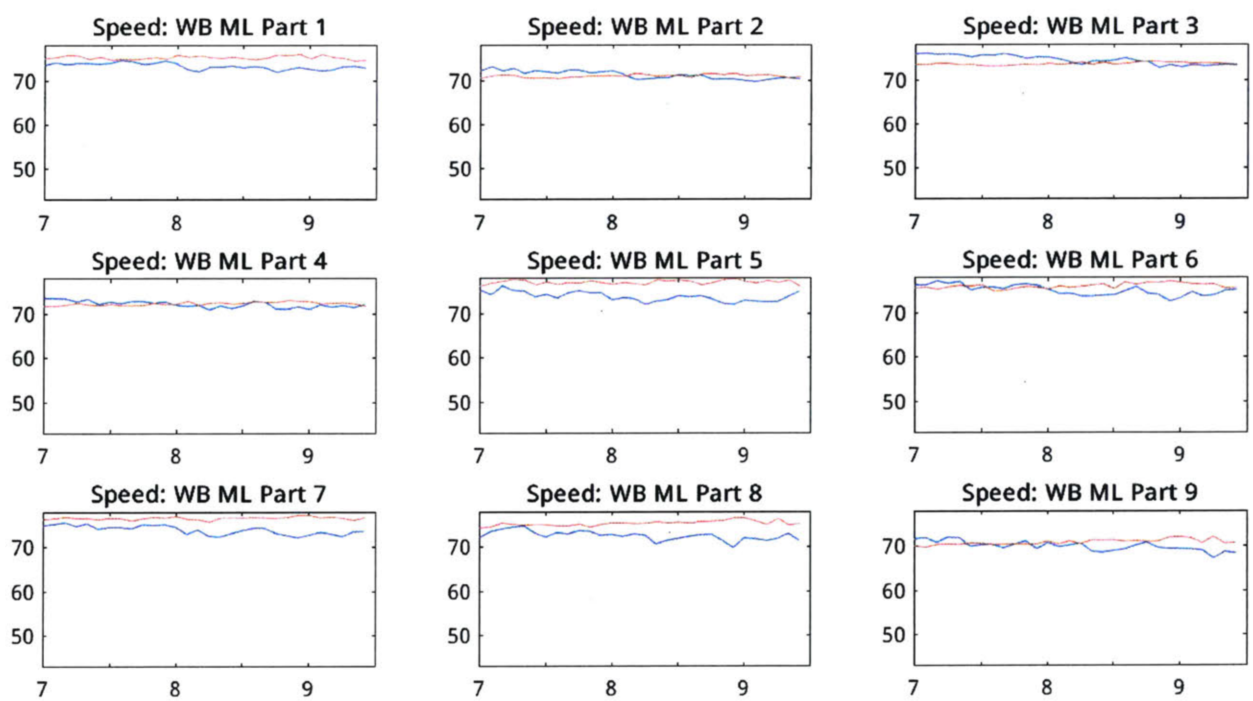

Figure 4.6f Final results w.r.t. sensor speed on ML

As for speed, simulated and actual speed on the network are different by less than 10mph in most cases. However, there exist some incidence of short-term speed drop in the real network, while the simulator is not capable of simulating such sudden changes of speed.

Some of the calibrated parameters are shown in Table 4.3.

Mean value of time ($/h)	35.627
S.D. of value of time ($/h)	14.618
Disutility at Gantry 8 ($)	4.17
Disutility at Gantry 5 ($)	3.50

Table 4.3 Calibrated value of selected parameters

Note that the value of time of drivers are assumed to follow a lognormal distribution. Each driver is assigned a random value of time based on the mean and S.D. of the distribution. The calibrated mean value of time is about 36 dollars per hour. This means, if managed lanes save 3 minutes, an average drivers would be more likely to choose managed lanes when toll is no more than 2 dollars.

Disutility at Gantry 8 and 5 (Figure 4.3 red circle and purple circle, respectively) is 4.17 dollars and 3.50 dollars, indicating drivers perceive the toll rates at these gantries by 4.17 and 3.50 dollars higher than actual toll charged.

Chapter 5

Conclusions

This chapter summarizes the methodologies and the results in the case study, as well as the significance, followed by discussions on several future research topics.

5.1 Summary

This thesis proposes improvements to the WSPSA algorithm for calibration of a traffic simulator. Calibration is formulated as an optimization problem, and different solution algorithms are reviewed and tested. The proposed WSPSA algorithm is proved to be more efficient and accurate. The refinements are made to the weight matrix, which is the key feature of the WSPSA algorithm that makes the algorithm superior to SPSA. Refinements are also made to other parts of the algorithm, including perturbing and updating the parameters, etc.

Case study indicates the proposed algorithm outperforms the original WSPSA, and satisfactory calibration results are obtained. This case study is conducted on a real-world road network using real data, and used non-homogeneous measurement data including traffic flow and speed. Calibration of demand and behavior parameters are conducted in a simultaneous way, which also makes the problem more challenging. With non-homogeneous data, successfully conducting simultaneous demand-supply calibration of this simulator in the real-world scenario proved satisfactory performance of the proposed WSPSA algorithm.

5.2 Future Research

Future research includes follow-up research of toll optimization on this road network, transferring the learning from the calibration algorithm to the toll optimization problem, as well as exploration of other calibration algorithms. The thesis also discusses application of PCA dimension reduction in a transportation network context to facilitate calibration and optimization.

5.2.1 Closed-loop Testing Framework of Toll Optimization

The simulator is calibrated for this test network as a preparation for the closed-loop testing framework of the dynamic toll optimization problem.

While we currently treat all vehicles as single-occupancy cars, the real network actually makes a distinction among different types of vehicles, and gives a discount to high occupancy vehicles. Figure 5.1 illustrates different toll rate factors for different vehicles. This distinction should be included in the toll optimization framework, and thus the simulator should be calibrated with expanded parameters to account for other types of vehicles.

Besides, other data that can be obtained in the real network could be exploited to improve the calibration and optimization. For example, calibration might consider origin to destination travel time as a measurement, and include this in the objective function. AVI data, or even customer demographic data can be made better use of in toll optimization.

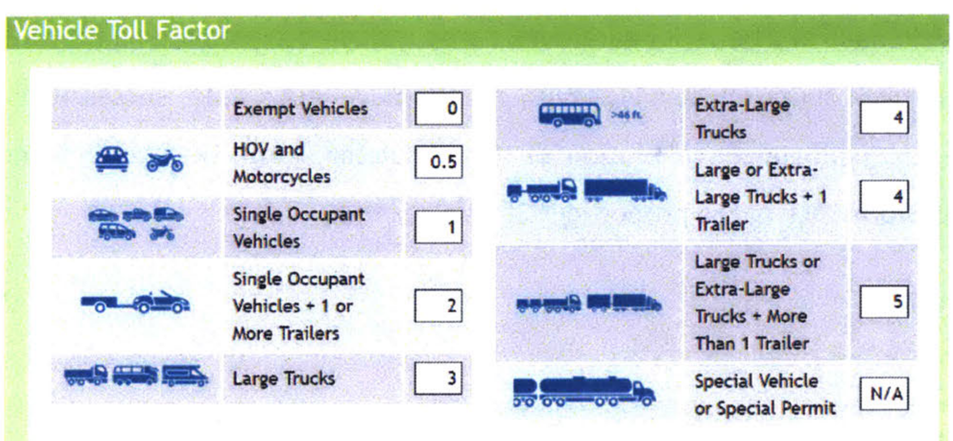

Figure 5.1 Toll rate factors for different types of vehicles

5.2.2 Toll Optimization Algorithms

The simulation-based dynamic toll optimization scheme proactively makes toll rate decisions to optimize some objective function, instead of reactively adjusting toll rates based on traffic conditions as in current practice.

Since toll rates usually take discrete values with step size of 5 cents, previous practice on toll optimization used exhaustive searches (Wang, 2016). However, the toll optimization scheme requires real-time performance, where a more effective algorithm is desirable, especially when the toll rate parameter has a large dimension.

The calibration work serves the toll optimization research in terms that not only the calibrated parameter values, but also the calibration algorithms can be transferred to the toll optimization problem. Toll optimization problem is different from simulator calibration in terms that parameters and objective functions are different, but the two problems use the same sensor measurements to compute objective, and share the same function structure among parameters, measurements, and objective.

The idea of decomposing the overall objective function into objective function for each parameter (toll rate at each gantry) can be used in the optimization problem, so that after a single run of the prediction module in DynaMIT, toll rate at each gantry can be updated independently. The gradient/weight matrix required for this decomposition process can be estimated offline and updated online.

Besides, a historical database of toll rates and any gradient/weight matrices under different demand levels can be established offline, in order to provide a better starting point and search space for optimization at each tolling interval.

5.2.3 Alternative Algorithms for Calibration

As discussed in the Case Study, calibration accuracy still has room for improvement after sufficient WSPSA iterations. Alternative calibration algorithms can be explored.

The calibration problem can be formulated as a GLS problem which solves a system of linear equations $y=h(x)$, where standard solution methods exist. However, the simulator $h(x)$ is not linear, so the main difficulty is obtaining a reliable gradient matrix in an efficient way, in order to linearize $h(x)$. This thesis has proposed a method for estimating the gradient matrix between parameters and measurements, thus the GLS algorithm can be used.

With a reliable gradient matrix estimated by an efficient approach, other gradient descent algorithms can also be used, which minimize the objective iteratively. During the iterative process, update size can be reduced to avoid fluctuation among iterations. Besides, the gradient matrix should be re-estimated when necessary.

Other algorithms can be explored, such as formulating the parameter evolvement process as a state-space model, and solving the calibration problem by constrained extended Kalman filter (Constrained-EKF) algorithm (Zhang, 2016), with augmented state vector to account for correlations among adjacent time intervals.

5.2.4 PCA Dimension Reduction on Sensor Data and OD Parameters

Principal component analysis (PCA) is a powerful data analysis and processing tool, which identifies hidden structures of the data. It can be used to reduce dimension of data, by decomposing each data vector into a new vector with lower dimension multiplied by a matrix that consists of principal component vectors which are estimated offline from a large set of data. This process eliminates the correlations among variables in the data, and also reduces noise.

Performing PCA dimension reduction on sensor data can reduce the effect of stochasticity in sensor readings, so that randomness in data has less impact on calibration. Besides, this also helps detecting traffic anomaly (Ringberg, 2007), which indicates the necessity for special treatment or manual adjustment.

PCA dimension reduction can also be applied on parameters (OD and others), either space-wise or time-wise, which reduces the number of parameters to be calibrated (Djukic, 2014), and thus improves the efficiency of gradient estimation by finite difference.

In practice, the GLS objective function is often simplified to WLS form, due to the difficulty of estimating the covariance matrix, and heavy computation burden when performing matrix multiplications. PCA dimension reduction eliminates the correlations among sensor data or among parameters, which makes it justifiable to assume zero covariance and simplify the objective function.

Reference

Antoniou, C. (2004). On-line calibration for dynamic traffic assignment (Doctoral dissertation, Massachusetts Institute of Technology).

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.

Antoniou, C., Azevedo, C. L., Lu, L., Pereira, F., & Ben-Akiva, M. (2015). W–SPSA in practice: Approximation of weight matrices and calibration of traffic simulation models. Transportation Research Procedia, 7, 233-253.

Balakrishna, R., Koutsopoulos, H. N., & Ben-Akiva, M. E. (2005). Calibration and validation of dynamic traffic assignment systems. In Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic Theory.

Balakrishna, R. (2006). Off-line calibration of dynamic traffic assignment models (Doctoral dissertation, Massachusetts Institute of Technology).

Balakrishna, R., Antoniou, C., Ben-Akiva, M., Koutsopoulos, H., & Wen, Y. (2007). Calibration of microscopic traffic simulation models: Methods and application. Transportation Research Record: Journal of the Transportation Research Board, (1999), 198-207.

Ben-Akiva, M. E., & Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand (Vol. 9). MIT Press.

Ben-Akiva, M., & Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. In Handbook of transportation science (pp. 5-33). Springer, New York.

Ben-Akiva, M., Koutsopoulos, H. N., Antoniou, C., & Balakrishna, R. (2010). Traffic simulation with dynamit. In Fundamentals of traffic simulation (pp. 363-398). Springer, New York.

Darda, D. (2002). Joint calibration of a microscopic traffic simulator and estimation of origin-destination flows (Master thesis, Massachusetts Institute of Technology).

Djukic, T. (2014). Dynamic OD demand estimation and prediction for dynamic traffic management (Doctoral dissertation, Delft University of Technology).

Greenshields, B. D., Channing, W., & Miller, H. (1935). A study of traffic capacity. In Highway research board proceedings (Vol. 1935). National Research Council (USA), Highway Research Board.

Gupta, A. (2005). Observability of origin-destination matrices for dynamic traffic assignment (Master thesis, Massachusetts Institute of Technology).

Kim, K. O., & Rilett, L. R. (2004, January). A genetic algorithm based approach to traffic micro-simulation calibration using ITS data. In 83rd Annual Meeting of the Transportation Research Board, Washington, DC.

Kundie, K. K. (2002). Calibration of mesoscopic traffic simulation models for dynamic traffic assignment (Doctoral dissertation, Massachusetts Institute of Technology).

Lu, L. (2013). W-SPSA: an Efficient Stochastic Approximation Algorithm for the off-line calibration of Dynamic Traffic Assignment models (Master thesis, Massachusetts Institute of Technology).

Park, B. B., Pampati, D. M., & Balakrishna, R. (2006). Architecture for on-line deployment of DynaMIT in hampton roads, VA. In Applications of Advanced Technology in Transportation (pp. 605-610).

Ringberg, H., Soule, A., Rexford, J., & Diot, C. (2007). Sensitivity of PCA for traffic anomaly detection. ACM SIGMETRICS Performance Evaluation Review, 35(1), 109-120.

Spall, J. C. (1992). Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE transactions on automatic control, 37(3), 332-341.

Toledo, T., Ben-Akiva, M., Darda, D., Jha, M., & Koutsopoulos, H. (2004). Calibration of microscopic traffic simulation models with aggregate data. Transportation Research Record: Journal of the Transportation Research Board, (1876), 10-19.

Toledo, T., Kolechkina, T., Wagner, P., Ciuffo, B., Azevedo, C., Marzano, V., & Flötteröd, G. (2014). Network model calibration studies. Traffic Simulation and Data: Validation Methods and Applications, 141.

Van Aerde, M., & Rakha, H. (1995, July). Multivariate calibration of single regime speed-flow-density relationships [road traffic management]. In Vehicle Navigation and Information Systems Conference, 1995. Proceedings. In conjunction with the Pacific Rim TransTech Conference. 6th International VNIS.'A Ride into the Future' (pp. 334-341). IEEE.

Wang, S., Atasoy, B., & Ben-Akiva, M. (2016). Real-Time Toll Optimization Based on Predicted Traffic Conditions. In Transportation Research Board 95th Annual Meeting (No. 16-2073).

Zhang, H. (2016). Constrained extended Kalman filter: an efficient improvement of calibration for dynamic traffic assignment models (Master thesis, Massachusetts Institute of Technology).