IMPROVING TRAFFIC OPERATIONS UNDER INCIDENTS USING LARGE-SCALE TRAFFIC SIMULATIONS AND SIMULATION-BASED OPTIMIZATION METHODS

by

Guanqi Liu
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Civil, Environmental, and Infrastructure Engineering

Committee:

___________________________________ Dr. Shanjiang Zhu, Dissertation Director

___________________________________ Dr. Mohan Venigalla, Committee Member

___________________________________ Dr. Elise Miller-Hooks, Committee Member

___________________________________ Dr. Jie Xu, Committee Member

___________________________________ Dr. Ossama Salem, Department Chair

___________________________________ Dr. Kenneth S. Ball, Dean, Volgenau School of Engineering

Date: ______________________________ Spring 2019
George Mason University
Fairfax, VA

Improving Traffic Operations under Incidents Using Large-Scale Traffic Simulations and Simulation-Based Optimization Methods

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Guanqi Liu
Master of Transportation Planning and Management Science
Texas Southern University, 2013
Bachelor of Optical Electronics Engineering
Science and Technology University of Changchun, 2010

Director: Shanjiang Zhu, Associate Professor
Department of Civil, Environmental, and Infrastructure Engineering

Spring 2019
George Mason University
Fairfax, VA

DEDICATION

To

My beloved parents Haijun Liu and Derong Sun, for their gracious love and endless support.

&

All my family members and dear friends, who have always been there for me

ACKNOWLEDGEMENTS

First of all, I am thankful for having a great and supportive advisor, Dr. Shanjiang Zhu. His inspiration guided me throughout my adventure of academic research. There would not have been this dissertation without Dr. Shanjiang Zhu's valuable and constant mentorship and advice. His knowledge and advice not only helped with my academic life, but also will surely guide me through my future professional career.

I would also like to thank all my committee members. Dr. Jie Xu Department of Systems Engineering & Operations Research at George Mason University for giving me so much insights on the optimization algorithms that I used in my research. His expertise in operations research is the back stone of this work. I also thank Dr. Elise Miller-Hooks and Dr. Mohan Venigalla for their valuable advice and suggestions on my research. I am so honored to have all these wonderful professors to be my dissertation committees.

I would like to give my appreciations to all my friends during the time I live in Virginia. I would like to thank beloved Yina Guo, my beautiful roommates Zhuo Yang, Shiying Zhang, Wenjie Li, as well as Yujue Yang, Ya Peng, Teng Sun and her cute adorable cat Jiubai. Thank you all for your love and friendship, and all the wonderful memories that we had together during these years in Fairfax. I also want to thank Ding Zhang, Peter Wang, Dr. Zhuo Yang, Dr. Chen Tian, Dr. Yangfeng Wu, Wensong Zhang, Jerry Ji, Zida Wang, Weichen Li, and all my amazing teammates from our soccer team. I am glad to have same passion and interest as you guys do. We played as a team and fought as a team, and I am grateful to have so many amazing moments with you on the pitch.

Lastly, I would like to thank the sponsors of my research, the USDOT, University Transportation Center program, without which this dissertation would not be possible.

TABLE OF CONTENTS

Page

List of Tables ... vii

List of Figures ... viii

List of Equations ... ix

List of Abbreviations ... x

Abstract ... xi

Chapter One INTRODUCTION ... 1

Background and Research Motivation ... 1

Simulation Models ... 3

Analytical Models ... 5

Objectives and Methodology ... 6

Dissertation Outline ... 8

Chapter Two LITERATURE REVIEW ... 10

Simulation Model ... 11

Analytical Models ... 17

Simulation-Based Optimization Methods ... 19

Summary ... 24

Chapter Three IMPROVING INCIDENT RESPONSE OPERATION STRATEGIES USING SIMULATION-BASED OPTIMIZATION METHODS: FEASIBILITY EVALUATION AND PERFORMANCE COMPARISON ... 26

Introduction ... 26

Relevant Studies ... 28

Methodology ... 29

Quadratic Polynomial Regression ... 31

Radial Basis Function ... 32

Bayesian Optimization ... 33

Numerical Test ... 35

Analytical Testing Bed 35
Numerical Results Analysis 39
Conclusions 46
Chapter Four IDENTIFYING OPTIMAL TRAFFIC OPERATION STRATEGY ON A LARGE-SCALE SIMULATION NETWORK 48
Introduction 48
Relevant Studies 49
Methodology 51
Integrated Framework 51
Experimental Performance of Integrated Model 57
Simulation Model 57
Model Calibration and Validation 59
Integrated Optimization Model 62
System Performance 65
Conclusions 69
Chapter Five OPTIMIZATION OF MULTIVARIATE TRAFFIC OPERATION STRATEGY 72
Introduction 72
Methodology 73
Integrated Framework 73
Design of Experiment 76
Integrated Optimization Model 79
System Performance 83
Conclusions 87
Chapter Six CONCLUSIONS 90
References 94

LIST OF TABLES

Table Page
Table 1. Optimal Search Performance 43
Table 2. Comparison of System-wide Total Travel Time under Different Scenarios 66
Table 3. Performance of Ramp Meter System under No-Incident Scenario 67
Table 4. Properties of Different Scenarios for Experimental Design 79
Table 5. Comparison of System-wide Total Travel Time under Different Scenarios 84

LIST OF FIGURES

Figure Page
Figure 1. Categorization of Simulation-based Optimization Models 21
Figure 2. Framework of Surrogate Models. 31
Figure 3. Sioux Falls Traffic Assignment Network. 37
Figure 4. RMSE Comparisons..41
Figure 5. Optima Searching Performance 43
Figure 6. Computation Time Comparison 45
Figure 7. Comprehensive Performance Comparison 45
Figure 8. Framework of the Incident Response Optimization Model (Adapted from Chapter Three, New Components Highlighted in Red)...53
Figure 9. Study Area in Northern Virginia 58
Figure 10. Simulation Network in DTALite 59
Figure 11.Comparison between Observed and Estimated Link Counts and Speed 61
Figure 12. Observed vs. Estimated Corridor Travel Time on I-66WB 62
Figure 13. Convergence of System Travel Time 65
Figure 14. Computing Time Comparison of the Three SBO Algorithms 65
Figure 15. Corridor Travel Speed Heat-map under Different Incident and Control Scenarios..69
Figure 16. Framework of the Incident Response Optimization Model (Adapted from Chapter Four, New Components Highlighted in Red)..74
Figure 17. Simulation Network in DTALite with Multiple Traffic Operation Controls...76
Figure 18. Convergence of System Travel Time 82
Figure 19. Computing Time Comparison 83
Figure 20. Corridor Travel Speed Heat-map under Different Incident and Control Scenarios..86

LIST OF EQUATIONS

Equation Page

$fx,\beta = \beta 1 + m = 1d\beta m + 1xm + m = 1d\beta m + d + 1xm2, \; xm \in$ $Data\ set\ x1, x2, x3, ..xd$ (1)...31

$\boldsymbol{f(x) = \sum_{i=1}^{n} w_n * exp(-\gamma \|x - X_i\|^2)}$ $\boldsymbol{f(x) = \sum_{i=1}^{n} w_n * exp(-\gamma \|x - X_i\|^2)}$ (2).....32

$\boldsymbol{\varphi = exp(-\gamma \|x - X_i\|^2)}$ $\boldsymbol{\varphi = exp(-\gamma \|x - X_i\|^2)}$ (3).....32

$\boldsymbol{p(f(x)|x, X_i, F(X_i)) = \int p(f(x)|x, F) * p(F|X_i, F(X_i))\, dF = GP(f(x)|\mu_*(x), \sigma_*{}^2(x))}$ (4)..33

$\begin{bmatrix} F(X_i) \\ f(x) \end{bmatrix} \sim GP(\begin{bmatrix} \mu(X_i) \\ \mu_*(x) \end{bmatrix}, \begin{bmatrix} K & k \\ k^{Transpose} & k^* \end{bmatrix})$ $\begin{bmatrix} F(X_i) \\ f(x) \end{bmatrix} \sim GP(\begin{bmatrix} \mu(X_i) \\ \mu_*(x) \end{bmatrix}, \begin{bmatrix} K & k \\ k^{Transpose} & k^* \end{bmatrix})$ (5)..34

$k(X_i, \boldsymbol{x}) = \sigma_f^2 * exp(-\sum_{m=1}^{d} \frac{1}{2l_m^2} \| X_i^m - \boldsymbol{x}^m \|^2)$ $k(X_i, \boldsymbol{x}) = \sigma_f^2 * exp(-\sum_{m=1}^{d} \frac{1}{2l_m^2} \| X_i^m - \boldsymbol{x}^m \|^2)$ (6)..34

$Logp(F(X_i)|\, X_i) = -\frac{1}{2} F(X_i)^{Transpose} K^{-1} F(X_i) - \frac{1}{2} Log|K| - \frac{N}{2} Log 2\pi$ (7)..34

$\mu_*(x) = k^{Transpose} * [K]^{-1} * f(X)$ (8)..35

$\sigma_*{}^2(\boldsymbol{x}) = k^* - k^{Transpose} * [K]^{-1} * k$ $\sigma_*{}^2(\boldsymbol{x}) = k^* - k^{Transpose} * [K]^{-1} * k$ (9)..35

$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (F(\boldsymbol{x}_i) - f(\boldsymbol{x}_i))^2}$ $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (F(\boldsymbol{x}_i) - f(\boldsymbol{x}_i))^2}$ (10).40

$\text{argmin}\, f(X) = \sum_{t=1}^{T} F(x_1^t, x_2^t, \ldots x_m^t \ldots, x_d^t, O)$ $\text{argmin}\, f(X) = \sum_{t=1}^{T} F(x_1^t, x_2^t, \ldots {}_m^t \ldots, x_d^t, O)$ (11)....................53

$\{(X_1^t\ F(X_1^t)), (X_2^t, F(X_2^t)), (X_3^t, F(X_3^t)), \ldots\ (X_i^t, F(X_i^t))\} \epsilon\ \Phi,$ (12)...............55

$\{(X_1^t\ F(X_1^t)), (X_2^t, F(X_2^t)), (X_3^t, F(X_3^t)), \ldots\ (X_i^t, F(X_i^t))\} \epsilon\ \Phi,$ (13)...............75

LIST OF ABBREVIATIONS

Simulation-Based Optimization	SBO
Bayesian Optimization	BO
Radial Basis Function	RBF
Quadratic Polynomial Regression	QPR
Gaussian Process	GP
Ramp Metering	RM
Shoulder Lane	SL
Variable Message Sign	VMS
Simulated Annealing	SA
Genetic Algorithm	GA
Total Travel Time	TTT

ABSTRACT

IMPROVING TRAFFIC OPERATIONS UNDER INCIDENTS USING LARGE-SCALE TRAFFIC SIMULATIONS AND SIMULATION-BASED OPTIMIZATION METHODS

Guanqi Liu, Ph.D.

George Mason University, 2019

Dissertation Director: Dr. Shanjiang Zhu

Traffic incidents are a major contributing factor towards traffic congestion. Effective incident response strategies can significantly reduce non-recurrent congestion and thus represent important research topics. In many cases, traffic operators only consider incident response strategies for local conditions on a relative basis. There is a need for efficient algorithms that could help identify optimal traffic operation control strategies on a regional network with complex traffic dynamics fully considered. However, identifying such optimal incident response strategies is a complex problem because of the large number of possible scenarios due to the combination of different type of control strategies, multiple time periods, different control parameters, multiple locations, and traffic dynamics. This optimization problem is high-dimensional and nonlinear. The underlying traffic dynamics do not have a close-form objective function and are also very time-consuming to evaluate. To address these challenges, an integrated optimization

framework that combines traffic simulation model with Simulation-Based Optimization (SBO) methods will be developed to identify the optimal traffic operation strategies under incident conditions. SBO methods can approximate the input-output relationship of a complex system using information from a set of sampling points and improve the response surface (objective function) effectively using new inputs guided from previous knowledge. Surrogate Models, a sub-category of SBO family, is a powerful method for solving optimization problems where the response surface is complex and unknown. The proposed integrated optimization framework was tested by combining Surrogate Models with a small analytical assignment network where the theoretical optimal can be analytically derived. Three Surrogate Models were evaluated: Quadratic Polynomial Regression (QPR), Radial Basis Function (RBF) and Bayesian Optimization (BO).The performance of the proposed framework is discussed. Results showed that BO has the best optimal solution searching capability yet with great running time. Though QPR requires much less computational time, it was shown to have poor optimal solution searching capability. This research then integrated a well calibrated traffic simulation model with SBO methods to identify the optimal traffic operation strategies under incident conditions on an urban area roadway network in Northern Virginia and the network-wise performance of the proposed framework was studied. This study presented a sequential method to further reduce the research space and improve the efficiency of the proposed model. Three different SBO algorithms are tested and compared, the performance of the proposed framework is discussed. The optimization results show the significant improvements of system total travel time and corridor congestion pattern. The results indicated that the proposed optimization model is capable of enhancing response strategies that are high dimensional and with an unknown objective function, solving both of which by either derivative optimization algorithms or brute-force searching method is infeasible. This study then introduced multiple choices of traffic operation strategies (Ramp Metering, Shoulder Lane, and Variable Message Sign) that function cooperatively and simultaneously to respond incident related traffic congestion. The optimal traffic operation strategy was then identified by using the proposed optimization framework. The results also showed significant improvements in total travel time in the system when multiple traffic operation strategies are implemented under incident scenarios. Concurrent deployment of synergistic multiple operational strategies are found to outperform each of the strategies working in isolation. This research is one of the first studies that aims to identify the optimal incident response strategies with multiple choices of traffic operation controls for responding to non-recurrent congestion on a large scale time-dependent simulation network with detailed traffic dynamics (e.g. users' reactions, queuing propagation, bottlenecks) fully considered.

CHAPTER ONE INTRODUCTION

Background and Research Motivation

As population grows and economy expands, traffic congestion has become a serious problem for major metropolitan areas worldwide. According to the Texas A&M Transportation Institute's 2015 Urban Mobility Report, traffic congestion issue has been continuously deteriorating since the economic recession of 2008 (Schrank et al., 2015). The average travel delay per driver has doubled since 1982. Peak hour commuters experienced almost 42 hours of travel delay, and an estimated 19 gallons of fuel and 960 US dollars are wasted on traffic congestion for each peak-hour driver in year of 2014. Nationally, the cost of congestion amounts to an extra 6.9 billion hours of travel delay, an extra 3.1 billion gallons of fuel and 160 billion dollars in United States. On average, drivers must plan their trip twice as long as the free flow travel time to ensure reliable arrival for important trips. Traffic congestion affects not only rush-hour commuters, but also drivers in other hours. Approximately 41% of total delay occurs during the mid-day and night hours when drivers should expect free flow traveling. Delays during non-peak hours will have a negative impact on the efficiency of logistics and manufacturing operations (Schrank et al., 2015). A recent study by INRIX indicates that 5 out of the top 10 most congested cities are from the United States in 2017 (INRIX, 2017). According to this study, the Washington D.C. Metropolitan area ranks No.6 among the most congested urban areas in the United States and No. 18 globally. Drivers spent almost 11 percent of their travel time (about 63 hours) in congestion in the year 2017, costing $1,694 per driver. Traffic congestion will also lead to other issues such as pollutions and energy efficiency.

Although many factors (e.g. increasing demand, poor geometry design, traffic incidents, bad weather, etc.) may contribute to traffic congestion, traffic incidents (non-recurrent congestion) are one of the major sources in urban area (Schrank et al., 2015). A traffic incident can be defined as "any non-recurring event that causes a reduction of roadway capacity or an abnormal increase in demand" (Owens, 2010). Such events include traffic crashes, disabled vehicles, spilled cargo, highway maintenance and reconstruction projects, and special non-emergency events. One previous study indicated that incident related delay caused 13-30 percent of peak hour congestion (Skabardonis, 2003). According to TTI's 2011 Urban Mobility Report (Schrank et al., 2011), 13 among 15 large metropolitan areas and 30 out of 32 medium urban areas considered incident management systems as a key solution to traffic congestion mitigation. According to the National Traffic Incident Management Coalition report (NTIMC, 2016), traffic incidents accounted for approximately 25% of traffic congestion on U.S. roadways. For every minute that a freeway lane is blocked during a peak travel period because of an incident, an estimated 4 minutes of delay occurs until the incident is cleared. Traffic incidents also lead to safety issues. On average, in the U.S. at least two emergency responders are struck each day and at least one U.S. law enforcement officer is struck and killed each month by a passing vehicle during traffic incidents. Furthermore, about 20% of all firefighter deaths are due to vehicle-related incidents (NTIMC, 2016).

It is important to identify a proper response strategy that could help to quickly detect, and respond to the incidents in order to expedite its clearance (Zhu et al., 2012; Zhu et al., 2014). However, in many cases, the required clearance time is out of the control for traffic operators. To mitigate traffic congestion, and the related safety and environmental issues due to traffic incidents, it is equally important to explore traffic control and travel demand management strategies that would help mitigate the impact of such non-recurrent congestion during the response period. Therefore, identifying the optimal traffic operation strategy will be yield huge benefits, especially when the transportation system has been over-saturated due to normal travel demands. However, identifying optimal traffic management strategies in response to traffic incidents is a big challenge due to the complex traffic dynamics and a large number of possible operation strategies. Each of these operation strategies may also include multiple decision variables, either discrete or continuous, which leads to a large number of combinations. These operation strategies will interact with each other through complicated traffic dynamics, posing additional challenges for identifying the optimal solutions. Conventional optimization methods, either based on analytical models or simulation models, cannot provide quick and satisfactory answers to our question.

Simulation Models

As computing power becomes cheaper and more readily available, simulation models have been widely used by researchers and transportation engineers to analyze transportation systems, and particularly traffic operation strategies. Simulation models can be grouped into three categories: macroscopic, microscopic, and mesoscopic. There three types of models differ in the level of details in traffic dynamics they are trying to capture, and at the same time the amount of computing time they require. A macroscopic simulation model describes the interaction of traffic flow, density, and speed at the aggregated level and keeps an analytical approach to maintain tractability to the maximum extent. Microscopic and Mesoscopic models usually capture more detailed traffic dynamics and if well calibrated, could provide a better description of real world traffic phenomena.

Traditionally, microscopic and mesoscopic traffic simulation models are commonly used to evaluate the performance of different transportation designs and traffic operation strategies. Transportation simulation models normally integrate a series of sub-models (car-following, lane-changing, etc.) to better simulate the microscopic traffic dynamics, thereby capturing the detailed interactions between vehicles and those between vehicles and traffic controls. Because of these complicated interactions and inter-connected models, traffic simulation does not offer explicit close-form objective functions that researchers, engineers, and policy makers may be interested in. Therefore, it is very challenging to identify the optimal traffic control strategy when simulation models are used. Instead of developing a systematic approach for optimization, most studies utilized simulation models to conduct scenario-based analysis. Researchers often evaluated several pre-determined design scenarios based on previous experience and/or engineering judgement, while do not directly address the question of identifying global optimal.

This scenario-based approach is not suitable for identifying optimal operation strategies since many corridors equipped with active traffic management hardware often include multiple traffic control strategies working cooperatively. The number of possible scenarios will quick explode when the number of time periods, possible control parameters, and combinations of strategies are considered. As running traffic simulation models on a large network is usually time-consuming, it is extremely hard, if possible at all, to test every possible scenario exhaustively in a reasonable timeframe to identify the optimal one. Moreover, operation parameters could be either continuous or discrete, making the enumeration of possible scenarios impossible in many cases.

Analytical Models

In contrast to simulation models, analytical models try to maintain the tractability and computing efficiency by adopting simpler traffic models. Most models adopt either point-queue models or link-delay functions to represent traffic dynamics. These models are usually formulated as a bi-level optimization problem, where the lower-level problem usually follows a user equilibrium paradigm. Because of the simplification in traffic models under an analytical framework, analytical models cannot capture the complexity of real-world traffic dynamics (such as capacity constraints at intersections and weaving segments, blockage of lanes, queue spillback, etc.). Arguably, these details are crucial for faithfully representing many traffic management strategies, and effectively evaluating their impact on regional traffic patterns and network performance. Therefore, we need to explore new methods that would strike a balance between computing efficiency and faithful representation of key traffic dynamic features.

Objectives and Methodology

Many traffic operation strategies may be implemented in response to non-recurrent traffic congestion (i.e. traffic incidents). For example, if an incident occurred on a freeway segment, traffic operators may limit access to a major freeway by using ramp meters on upstream on-ramps. Traffic operators may also increase the capacity of some congested freeway segments by opening shoulder lanes. They may also rebalance regional travel demands by releasing travel information through variable message signs (VMS), or other communication channels (e.g. 511, radio, smart phone apps). These strategies do not work in isolation. Adopting multiple strategies may be synergetic, or counter-productive. Therefore, it is important to develop a method to evaluate different alternatives with reasonable computing time and identify the optimal traffic operation strategy or strategies. However, given the large number of possible scenarios, a mix of discrete and continuous operating parameters, and complex interactions between strategies, finding the best solution in a reasonable time is a big challenge, especially when the transportation system is time-dependent, which leads to a high dimensional and nonlinear optimization problem to solve.

To address this challenge with profound implication for informed traffic management, this dissertation is seeking to develop a systematic method which is capable of combining the strengths of both simulation approach and analytical approach. This dissertation is aiming to develop a method which can effectively identify the optimal traffic operation strategy or a combination of strategies that optimize the roadway network performance under incident conditions. This study proposes an optimization framework that integrates traffic simulation models with simulation-based optimization algorithms to identify the optimal traffic controls on a time-dependent large-scale simulation network under incident condition. The proposed framework will address the challenge of high search dimensions and nonlinearity of system performance as a function of control strategies. To achieve this objective, this study proposes to use Surrogate Models, a subcategory in the family of Simulation Based Optimization (SBO) methods to estimate the system performance under different scenarios in an efficient way. The proposed method could search the entire input domain of multiple controls (e.g. ramp-metering, shoulder lane, VMS) and identify the optimal/near optimal solution within limited time. Three Surrogate Models, including Quadratic Polynomial Regression (QPR), Radial Basis Function (RBF) and Bayesian Optimization (BO) will be integrated with transportation simulation models, each at a time using the proposed optimization framework. Their performance will be evaluated and compared first on a small analytical network where the theoretical optimal could be identified through brutal force method. This demonstration on a small network serves as prove of concept for the proposed optimization model and shed lights on the strengths and weaknesses of the three simulation-based optimization algorithms. This dissertation will then apply the proposed integrated optimization framework on a large scale time-dependent simulation model with high dimensional decision variables to demonstrate its feasibility and scalability on a regional network. The performance of the proposed optimization framework with the three Surrogate Models will be compared and discussed. Finally, this study will introduce more traffic operation strategies into the network, and particularly those that would involve more complicated behavioral responses from road users (e.g. VMS). The results of this study would provide researchers and traffic operators better insights into deploying multiple traffic operation strategies under incident conditions. As managed lanes and the concept of Active Traffic Management (ATM) become more popular in the U.S., methods developed in this study could be extend to analyze scenarios beyond incident responses and help mitigate both recurrent and non-recurrent congestion through better traffic operations.

Dissertation Outline

The rest of this dissertation is organized as follows: the second chapter reviews previous research on incident responses based on analytical methods, simulation methods and/or simulation-based optimization methods. The third chapter will introduce the proposed integrated optimization framework. It then applies the proposed optimization framework on a small analytical network where the theoretical optimal is identified through a brutal force approach. This case study provides as a benchmark to investigate the performance of the proposed framework and compares the performance of three commonly used Surrogate Models. Chapter Four applies the proposed optimization framework on a large scale mesoscopic simulation model and addresses technical challenges associated with the scalability, computing efficiency, and some limitation of the Bayesian framework. The proposed optimization framework is capable of reducing the dimension of decision variables through a sequential approach and maintain the scalability of the proposed framework on a large transportation simulation network. In the fifth chapter, traffic operation strategies with more complicated behavioral reactions from road users (e.g. VMS) will be introduced to the integrated optimization framework and technical challenges related to the randomness and sampling issues will be discussed. This chapter will demonstrate that the integrated optimization model is capable of optimizing the multivariate traffic operation strategies simultaneously on a large scale time dependent traffic simulation network. The final chapter will conclude this dissertation with a brief discussion regarding the findings and insights of this dissertation.

CHAPTER TWO LITERATURE REVIEW

This study proposes an integrated optimization framework that combines traffic simulation models and Simulation-Based Optimization (SBO) algorithms to identify the optimal traffic operation strategies on a time-dependent large-scale simulation network under incident conditions. The proposed optimization model will be combining the strength of computational capability of simulation model and tractability of analytical model to optimize the traffic operation strategy which bridges the gap between previous studies. Both simulation and analytical methods have previously been applied in transportation studies. However, integrating these two methods to address challenges in traffic operations, and particularly in the context of incident responses, has not been well investigated. Many questions, including the performance of different simulation-based optimization algorithms, the scalability of such method on large-scale simulation models, the interaction between different control strategies and its implication of optimization results, have not been addressed in the literature. This chapter will review previous efforts in three different research areas: 1). Research on traffic operation and management using simulation models; 2). Research on the same topic using analytical models; 3). applications of simulation-based optimization methods in transportation research. Findings from this chapter will identify the gap in the existing literature, highlight the value and contributions of the proposed modeling framework, and provide guidance for the research to be conduct in this dissertation.

Simulation Model

Traffic simulation models have been widely used to evaluate the performance of pre-determined traffic operation strategies under various scenarios, many of which are in the context of incident responses. Simulation models can be grouped into three categories: macroscopic, microscopic, and mesoscopic simulation. The essential difference of the three simulation models is the resolution of the models will be focusing.

Transportation simulation models had been developed and introduced to solve transportation design and operation problems since 1950's. Various simulation software models such as TransModeler, VISSIM, Paramics had been widely utilized for diverse research purposes in transportation area (Chu and Yang, 2003; Fellendorf and Vortisch, 2010; Zhang, et al., 2012). Simulation models have also been used for modeling incident response actions and operations. In some of the early research on real time incident response, Southworth et al. (1992) proposed a Dynamic Traffic Assignment (DTA) modeling framework for real time traffic routing during mass evacuations. Hasan (1999) investigated two ramp metering algorithms (ALINEA and FLOW) and compared their performance using microscopic traffic simulation models. Pal and Sinha (2002) used a mesoscopic simulation model to systematically develop freeway service patrol strategies to improve the total vehicle hours and response patrol deployment schedule. Pulugurtha et al. (2002) used CORSIM and VISSIM models to model and analyze freeway incident by using Las Vegas metropolitan area traffic data on a 2.5-mile freeway segment. Their study compared CORSIM and VISSIM under several different incident conditions. Results showed that CORSIM performed better than VISSIM on simulating incident. Ben-Akiva et al. (2003) used a microscopic traffic simulation model to evaluate the performance of a wide range of freeway traffic operation strategies (e.g. ramp control, lane control, integrated control with route guidance) separately under various scenarios in Boston. Chauhan (2003) used Mesoscopic Dynamic Traffic Assignment (DTA) model to investigate the impacts of Variable Message Sign (VMS) as a traffic operation strategy under various traffic and incident scenarios. The study attempted to improve the computational efficiency in scenario evaluation and investigated the impacts of VMS diversion strategies under incident situation. The author concluded that the predictive VMS strategy behaved the best on travel time reduction. The impacts of VMS strategy also hugely depend on the incident location and number of available alternative routes. Similarly, Wirtz et al. (2005) also used a mesoscopic DTA model to investigate incident response strategies. The authors applied a self-developed mesoscopic software called VISTA to model the performance of freeway incident response actions under various incident durations and scales on northern Chicago highway network. This research concluded that implementing the wrong incident response strategy could worsen the congestion on the directly impacted freeway segment and its surrounding highway region. Incident scales and durations have the most statistically significant impact on congestion severity. Closing on ramps upstream of the incident location would be an effective way to reduce congestion but only for incidents with long duration and huge lane closing. This research also has limitation, such that, the model that is used in the study assumes drivers have perfect information and always switch to the routes with shorter travel time. It is impossible for people to know the perfect information of all routes and User Equilibrium is most unlikely to be reached in the real world, especially when a non-recurrent congestion occurs (e.g. incident related congestion). Chu et al. (2004) used Paramics, a microsimulation tool to evaluate how ITS strategies can reduce non-recurrent traffic congestion over a network through scenario analysis. The evaluated ITS strategies include incident management, adaptive ramp metering, traveler information, arterial management, and a combination of those strategies. Their results indicated that all ITS strategies have positive effects on network performance. However, adaptive ramp metering cannot improve system performance effectively under the incident scenario. This research study modeled incident scenarios hypothetically in the simulation model and called up API quantitatively evaluate embedded ITS strategies. This research also concluded that the proper combination of ITS strategies leads to greater benefits and variable message signs play the most important role in mitigating traffic congestion under incident situation. Sisiopiku (2007) developed and evaluated emergency response scenarios on regional transportation network by using large-scale CORSIM micro-simulation model. Three emergency scenarios were developed and tested by simulation. The outputs of different scenarios were compared and recommendations were made for the future needs. This study proves the capability of simulation approaches in modeling emergency response management. Huaguo (2008) used CORRIM simulation model to investigate the impacts of incident management and designated signal timing plan modifications on traffic diversion operations. The study claimed that the purposed signalized diversion operation would significantly reduce overall network delay by 21%. The researcher used SYNCHRO to design the signal plans for six different incident scenarios, and indicated that traffic diversion has a critical impact on total delay to the network. This study only tested one strategy within small network. Zhou (2008) used simulation package CORSIM and Synchro to evaluate the performance of route diversion plans for incident response management. The results demonstrate that the percentage of diverted traffic is a critical factor on travel delay reduction on network-wide performance. The authors found that a 10% diversion rate offers the optimal total network delay based on the case study. Ozbay et al. (2009) applied a self-developed a mesoscopic cell transmission based simulation model to evaluate incident management strategies including the deployment of Variable Message Signs (VMS) to divert traffic during incidents and the use of Freeway Service Patrols (FSPs) for detecting and verifying incidents efficiently. This research mainly focus on the benefit/cost analysis of deployment of VMS, and the incident duration reduction and also benefit/cost analysis due to increasing on number of FSP units. The results showed that the implementation of VMS has the positive impact on congestion relief. Moreover, increasing the number of FSP unit generate more benefit by reducing incident detection time. Chou & Miller-Hooks (2011) investigated the potential benefits of diverting traffic to managed lanes on freeway during incident by using microscopic simulation tool VISSIM and quantifying the impacts and benefits of traffic mitigation strategies. The results showed that the impacts mainly depend on the location of the incident scene relative to the starting point of the diversion opening, the total length of access to the managed lane under the diversion strategy, incident duration, and incident severity (number of lane being closed). The authors then proposed several approaches (response actions) to be evaluated in the simulation model. The results showed that diverting point closer to the incident location the better, GP vehicles can jump into the managed lane immediately after detecting queue formation in the GP lanes. The longer the incident duration, the greater the benefit. Permitting continuous access to managed lane produced significant benefit on travel time saving. Tasic (2012) used simulation model to develop a set of incident response strategies that optimize decision making in terms of time and response level minimize users' cost due to incident induced delay on the freeway network. This research conducted a comparison study for different combinations of incident duration, level of closure, VMS level and response time and determined the responses that give the optimal results. The limitations of this study were that the VMS diversion rate could only be obtained through survey or previous studies, and the lack of data from arterial. Another drawback of this study is that the comparison of such a great number of different combinations of incident response strategies would be computationally heavy and lack of determination of global optimal. Henchey et al. (2014) built a simulation model to well replicate the real-world traffic condition. Most importantly, this model can simulate the emergency vehicles responding to the scene of a crash and provide the basic information for future needs. Jiang & Chung (2014) proposed a zone rapid congestion recovery strategy of ramp metering to mitigate the recurrent congestion during morning peak hours. In this research, the author only evaluated single traffic operation strategy by conducting a case study in Brisbane, Australia. The case study used microsimulation model with static traffic demand and the proposed ramp metering strategy was integrated into the simulation model through API. The results indicated that the proposed ramp metering has better performance on congestion recovery than the basic ramp metering plan. Nitsche et al. (2016) investigated a scenario-based study of evaluating the overall performance (incident duration and response time reduction) of the novel traffic incident management techniques (e.g. eCall, C2X or xFCD) by using macro-simulation model. Papayannoulis et al. (2011) used an integrated simulation model to simulate the performance of traffic operation scenarios under different traffic operation designs and geographical designs in Manhattan, New York. Vadde et al. (2012) used microscopic simulation model VISSIM to evaluate the effectiveness of using various active traffic management (ATM) strategies such as speed harmonization, hard shoulder, and ramp metering on freeways. The simulation results indicated that the implement of these traffic operation strategies has significant improvements on speed, travel time, and crash rates on the study locations. Dion et al. (2012) developed traffic simulation models to evaluate the impacts of different Corridor System Management strategies for California freeway corridors. Xiong et al. (2015) developed a 24-hour large-scale microscopic traffic simulation model on a tollway system that was implemented for the Washington, DC, metropolitan area. The research team used TransModeler to conduct a case study that evaluates the performance and impacts of the tollway system both in network level and corridor level. The simulation network that was developed in this research is applicable for traffic operation analysis of a wide range of planning policies, traffic control and active traffic operation strategies such as dynamic lane controls, ramp metering, and en-route diversion. Lee et al. (2017) used real time responsive ramp metering embedded in a large scale simulation-based mesoscopic DTA model to investigate the network-wide and local-wide impacts. This research provides a prototype planning framework beyond the corridor level with traffic operation management tool.

Although many researchers have used various simulation models to evaluate the effectiveness of a wide range of traffic operation controls (e.g. shoulder lane, ramp metering, variable message sign, toll road, etc.). Most of these efforts were still limited to scenario-based evaluation of several pre-determined strategies and did not search for the optimal solution, especially in the context of combining multiple traffic operation controls. The problem becomes intractable because of the large number of strategy combinations and the long computing time for simulating a large network.

Analytical Models

In addition to simulation models being used on modeling the performance of traffic operation management tools. Some previous research adopted analytical models to study incident responses to tackle with the intractability and expensive computing efforts issues of simulation models. However, there are few studies focused on solving traffic operation strategy optimization from analytical perspective. Chen et al. (1997) formulated a dynamic ramp metering control problem and solved the optimal metering rate using a non-linear programming method. Zhang et al. (2001) formulated the ramp metering control problem based on a discretized traffic flow model. Levinson (2003) utilized probit choice model and empirical traffic data to quantitatively investigate the effectiveness of variable message sign (VMS) on network travel time and total delay mitigation. Results indicates that the VMS stands only has no obvious effect on reduction of travel time, however, it will be obviously improving on travel time reduction if VMS works with ramp meters simultaneously, especially when incident occurs during rush hours. Kotsialos et al. (2001) and Kotsialos et al. (2002) integrated various traffic control strategies (VMS and Ramp Metering) with a second-order macroscopic traffic flow model. The authors built a discretized nonlinear optimal control problem and used a feasible-direction nonlinear optimization algorithm to obtain the optimal traffic control strategies. Some researchers extended the study of Kotsialos et al. (2001) and studied the optimization problem of implementing the variable speed limit in coordination with ramp meters (Hegyi et al., 2005; Carlson et al., 2010; and Carlson et al., 2010). They concluded that the combination of two strategies would lead to better performance than the cases of a single strategy. Most studies based on analytical models are limited to a single corridor and do not consider complex traffic dynamics as those captured by traffic simulation models. Zhao et al. (2017) proposed an integrated model to tackle the off-ramp congestion problem by maximizing the capacity of downstream intersection. The research adopted a nonlinear mixed-integer program model to optimize pre-signal strategy and signal timing problem on off-ramps and signalized intersection. The optimization is solved by branch-and-bound technique. Brennan et al. (2018; 2019) used large data repositories obtained from probe vehicles to evaluate congestion propagation on a regional network. The studies identified the need for evaluating congestion propagation and simulating incident response strategies in real time. There has not been many research used analytical methods to investigate incident respond problem from traffic operation management perspective, which leave us a big research gap. Most studies based on analytical models are limited to a single corridor and do not consider complex traffic dynamics as those captured by traffic simulation models.

Simulation-Based Optimization Methods

Analytical models were used to keep the tractability and clear mathematical formulations. Such models would trade-off the realistic and stochastic nature of traffic dynamics and incident patterns. In contrast, simulation models have been widely used to replicate the real-world traffic patterns and dynamics to help traffic decision makers to make the best decision. To find the optimal operation strategies in a reasonable time duration, a clear and well-behaved objective function (such as those adopted in analytic models) is needed, yet since simulation models introduce noise and stochasticity into the model, the analytical relationship between traffic operation strategies and its impact on transportation system is hard to estimate, such an objective function is usually unavailable if we want to capture complex traffic dynamics through simulation models. Moreover, it is usually very time-consuming to evaluate even a single scenario on a large network, which makes it computationally intractable to use brutal force methods. Therefore, we need to develop an efficient method to estimate the objective function and the optimal solutions simultaneously. Increasing number of studies have used simulation-based optimization (SBO) methods to reduce the number of scenarios that modelers need to simulate and accelerate the decision making process. Such models combined simulation power and tractability of mathematical optimization.

There are various types of simulation-based optimization methods. Simulation-based optimization will build or estimate an input-output relationship of the system when analytical expressions are unavailable. Fu (1994) grouped simulation optimization problems into two categories: discrete and continuous. For the problems with discrete input and finite set, multiple-comparison and ranking-and-selection methods would be the best approaches to use. For the problems with continuous input parameters, gradient-based methods including perturbation analysis, the likelihood ratio method, Finite Difference, and frequency domain analysis can be applied. Carson and Maria (1997) introduced an overview of simulation-based methods and grouped them into six categories. Ammeri et al. (2010) comprehensively introduced simulation-based optimization methods. The authors categorized them by the continuity of input variables, and the shape of the response surface. Rios & Sahinidis (2013) and Amaran et al. (2014) also conducted a systematic review on SBO methods. This paper introduced the basic concept of simulation optimization. This paper also compared SBO and derivative-free optimization. The author classified the algorithms by different research needs and characteristic of the problems (e.g. infinite/finite parameter set, discrete/continuous variables, etc.). Xu et al. (2015) also gave a brief review of the state-of-the-art simulation-based optimization algorithms and illustrated how simulation optimization methods would work around the uncertainty and stochasticy of simulations. To be convenient and reader friendly, a categorization chart of existing simulation-based optimization models has been showed into Figure 1.

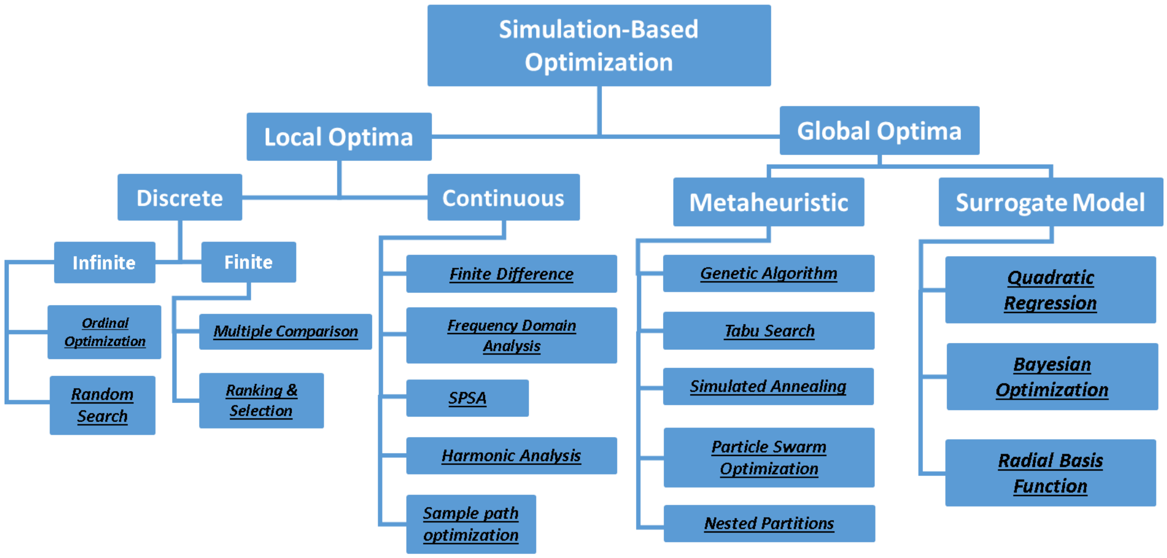

Figure 1. Categorization of Simulation-based Optimization Models

SBO algorithms have been also applied in transportation areas. For instance, Ting & Schonfeld (1998) used Simultaneous Perturbation Stochastic Approximation (SPSA), a gradient based optimization method to optimize waterway transportation investment problem. Adeli & Karim (2000) developed a radial basis function neural network (RBFNN) to classify the de-noised and clustered observed data. The new model produced excellent incident detection rates with no false alarms when tested using both real and simulated data. Fu & Howell (2003) applied perturbation analysis (PA) method to determine the traffic signal timing for a single intersection. Wang & Schonfeld (2005) and Tao & Schonfeld (2005) used Genetic Algorithm (GA) to optimize the transportation budgeting problems. Some researchers also used GA algorithm to locate the global optima of the traffic signal control system (Teklu et al. 2007; Stevanovic et al. 2008). PIZANO (2010) used surrogate model to optimize regional signal timing strategies.

Osorio & Bierlaire (2013) proposed a simulation-based optimization method that combines with quadratic polynomial regression (QPR) and queuing model to optimize traffic signal timing control problem. The results indicated the model reduced average travel time. Osorio and her research team also did some extensions based on their original research from 2013. Osorio & Chong (2015) used SBO method to efficiently solve signal optimization problem on a large-scale transportation network. Based on the previous ground research, Chong & Osorio (2018) extended their previous study to SBO method to optimize signal control plans on a dynamic network. Li et al. (2010) applied a heuristic approach that is using Lagrange relaxation technique to optimize highway transportation investment problems under uncertain budget. More recently, Melouk et al. (2011) adopted a metaheuristic algorithm (Tabu Search) to search for the best congestion mitigation strategies. This research simulated traffic using a macroscopic simulation model. The optimization method identifies the best strategy of which congested link is necessary for capacity boost to mitigate the congestion of roadway networks. This study only considered lane capacity increases as a traffic congestion mitigation solution. Ciuffo et al. (2011) used kriging to verify different micro-simulation calibration methods and improve the calibration of traffic simulation models. In some of the early research on real time incident response, Southworth et al. (1992) proposed a dynamic traffic assignment (DTA) modeling framework for real time traffic routing during mass evacuations. Zhu et al. (2018) also used kriging to estimate the best supply parameter (capacity, jam density etc.) that calibrates the large-scale dynamic traffic assignment (DTA) network. This method is able to reduce the number of simulation runs for function evaluation in calibration process and surpass the performance of genetic algorithm in searching for better solutions. Pal & Sinha (2000) is the first research that used simulation-based optimization method to solve incident response related optimization problem. This research focused on optimizing the resource allocation, hours of operation, fleet and crew sizes, dispatching policies, areas of operation, and routing patterns of incident response units through dynamic transportation simulation incorporated with optimization algorithm that enhance the patrol service and improve the efficiency. This research used nested partitions method to optimize the performance of the system. Nested partitions method was selected because of the high efficiency of computing capability. This method divided the whole search space into some sub-regions and those sub-regions with good solutions are clustered. He (2014) is one of the first researchers to implement simulation-based optimization methods on large-scale transportation network problems. This research used SBO methods to evaluate the unknown objective function of a large-scale toll road system and used Genetic Algorism to find the optimal toll rate. The author compared and tested some of the most popular surrogate models, such as Bayesian Optimization, Radial Basis Function, and Quadratic Polynomial Regression. This research was focused on large-scale transportation simulation network, and it's capable of solving multi-objective optimization problem with consideration of simulation noise. Chen et al. (2014) evaluated and compared the family of surrogate models on approximating objective function for transportation simulation input-output mapping. The research applied kriging to optimize highway toll rates in order to minimize the average travel time. However, the computational expense is huge and it's not good enough for optimizing transportation problem with high dimensional variables. Zhang et al. (2014) also applied kriging method to investigate the usage of HOT lanes and Dynamic message signs. Chen et al. (2017) integrated simulation based optimization algorithm and transportation simulation model to identify the optimal toll pricing strategies to improve network-wide reliability.

Summary

Traffic simulation models have been widely used to evaluate the performance of different traffic operation and management strategies. Most of these studies were scenario-based and only a few alternatives were selected based on engineering judgement. Most studies either evaluated a single strategy, or a few strategies but with a searching space reduced to a sparse grid. This approach is not systematic and can only identify the best performer among a few pre-determined scenarios, thus does not qualify as an optimization method. Expanding the searching grid is computationally inefficient, and optima are not guaranteed.

Analytical methods on the other hand, did not fully capture traffic dynamics related to many operation strategies because of the limitation of analytical models. Most of the previous analytical-based research on incident responses mainly focused on minimizing the response time, optimizing response team allocation and deployment, and reducing impacts of incidents by reducing incident duration. Although these issues are important for incident responses, they are different from optimizing traffic management strategies (e.g. ramp meters, VMS, and should lanes) to reduce system congestion during and after incident clearance.

In addition, it is extremely difficult to build up a close-form analytical objective function, which means the gradient and derivative information is usually very difficult to obtain. Therefore, neither simulation approach nor analytical approach would satisfactorily address the incident response strategy problems traffic operators are facing. Simulation-based optimization methods provide a viable alternative. A large number of research has been dedicated to exploring the simulation-based optimization methods for transportation applications. However, few of those focused on addressing incident response optimization problems. A framework based on simulation-based optimization approach has the potential of addressing both the issue of replicating complex traffic dynamics and interaction between operation strategies, and the tractability for optimization problems on a large-scale transportation roadway network. However, such method has not been sufficiently developed. Many technical issues, including the best simulation-based optimization algorithm to be used, the scalability, and best way to capture behavioral reactions to control strategies, have not been explored. To bridge the gap, this research will integrate traffic simulation model with simulation-based optimization methods to efficiently identify the optimal traffic operation strategies for mitigating non-recurrent congestion while still considering the complex traffic dynamics on a large network.

CHAPTER THREE IMPROVING INCIDENT RESPONSE OPERATION STRATEGIES USING SIMULATION-BASED OPTIMIZATION METHODS: FEASIBILITY EVALUATION AND PERFORMANCE COMPARISON

Introduction

As indicated in the introductory section, with active traffic management (ATM) systems being increasingly deployed nationwide, there is an urgent need for methods that help to optimize traffic control and travel demand management strategies, and particularly for mitigating the impacts of traffic incidents during the response period. There are many traffic operation strategies that may be deployed as part of the ATM. For example, traffic operators may limit the access to a major freeway and divert traffic using ramp meters when an incident occurs. Traffic operators may also increase the capacity of some freeway segments by opening shoulder lanes. They may also rebalance regional travel demand by releasing travel information through variable message signs (VMS), or through other communication channels. These strategies do not work in isolation. Given the large number of possible traffic operation scenarios and complex traffic dynamics, finding the best solution quickly is a big challenge.

This study proposes to use Surrogate Models, a subcategory of SBO family, to explore the optimal traffic operation strategies for incident response management. The proposed method will estimate the system performance (i.e. total system travel time in this case) as a function of control strategies and/or parameters using a small number of samples collected through simulation, which will inform researchers the potential optima and guide the selection of points for further simulation trials. The new samples will improve the estimation of the response surface (system performance as a function of traffic operation controls), and iteratively lead to the optima. This simulation-based optimization framework keeps the transportation simulation models to track detailed traffic system dynamics to the maximum extent possible, while limited the number of trials by exploring the analytical features of the response surface. This guided searching approach helps to reduce the total number of samples (i.e. the number of times to run the simulation model), representing a great improvement compared to the scenario-based analysis or the grid-based brutal force searching.

Although this simulation-based optimization method has been widely applied in other fields, it has not been sufficiently explored in the context of traffic operations and management. The pros and concerns of many candidate algorithms under this framework, and the performance of the overall framework have not been well studied. Three Surrogate Model algorithms will be tested and compared in this research. These models include: Quadratic Polynomial Regression (QPR), Radial Basis Function (RBF) and Bayesian Optimization (BO). To better understand the performance of the proposed framework, an analytical traffic assignment network was adopted to replace the simulation model for evaluating the system performance in response to traffic controls. This analytical network allows us to identify the global optimal solution and use it as the benchmark for better comparison. Findings from this analysis based on a small network would help us better understand the performance of the proposed framework, and confidently extend the framework to investigate the optimal control strategies on a large network.

Relevant Studies

Although application of SBO methods in incident response studies is still preliminary, it has been widely used in many other fields. In transportation related studies, Pal & Sinha (2000) are among the first researchers who used SBO method to solve the incident response optimization problem. This research used nested partitioning method to optimize the performance of the system. Pizano (2010) used Surrogate Model to optimize regional signal timing strategies. Ciuffo et al. (2011) used kriging to improve micro-simulation calibration. Osorio & Bierlaire (2013) and Osorio & Chong (2015) proposed a SBO method that combines QPR with a queuing model to optimize traffic signal control plans. He (2014) pioneered the research efforts to implement SBO methods on large-scale transportation networks to evaluate the unknown objective function of the toll road system and used GA to explore the optimal toll rate. Osorio & Nanduri (2015) also used SBO algorithm to optimize traffic signal control strategies with the objective of minimizing total fuel consumption. Additionally, other general introductions and applications also showed that SBO methods were powerful to tackle the problems with unknown objective functions, which are ideal for addressing the shortcomings of simulation model in optimizing incident response strategies (Fu, 1994; Carson & Maria 1997; Ammeri et al., 2010; Rios & Sahinidis, 2013; Amaran et al., 2014; Xu et al., 2015). This research employs Surrogate Models (also known as Response Surface Model or Meta-Model), a subcategory of SBO methods to solve the optimization problem with an unknown objective function. These models use system performance information from sampling points (i.e. simulation outputs) to approximate the statistical structure or response surface that represents the relationship between simulation inputs and outputs. A general overview of the models can be obtained from Fu (2002); Søndergaard (2003); Barton & Meckesheimer (2006); Kleijnen (2008); Ammeri et al. (2010); and Amaran et al. (2014). The three most mainstream Surrogate Models include: Quadratic Polynomial Regression, Radial Basis Function and Bayesian Optimization (Mockus et al., 1978; Jones et al., 1998; Björkman & Holmström 2000; Gutmann, 2001; Wright, 2003; Regis, & Shoemaker 2005; and Gelbart, 2015) all of which were tested and evaluated in this study.

Methodology

The primary objective of this study is to develop optimal traffic operation strategies or the best combination of traffic operation actions that lead to the best network performance (i.e. minimum total travel time in the system) under certain incident scenarios. To maintain the tractability of the network model and generate insights on the performance of proposed Surrogate Models to be tested, we used a simple User-Equilibrium (UE) based traffic assignment model instead of large-scale traffic simulation model to simulate network dynamics in the form of O-D flows and turn movements. By very nature, the outputs of the assignment model are deterministic for each input scenario and we don't have to do multiple simulation runs to get replications as are required for simulation models. Thus, we are able to obtain the theoretical optimal traffic operation solution as a benchmark within a reasonable timeframe and compare it with the outputs from the proposed Surrogate Models. Due to the complexity and nonlinearity of the traffic assignment model, which follow UE as a constraint (a bi-level structural optimization problem). The optimal solution is hard to identify under a bi-level structure by converting to single-level relaxations (Christiansen et al., 2001; Ma 2016). Therefore, this theoretical optimal solution is obtained by exhaustive search.

The objective of the optimization problem is to minimize total travel time, $F(X)$. For a given data set Φ, $\{(X_1\ F(X_1)), (X_2\ F(X_2)), (X_3\ F(X_3)), \ldots \ldots\ (X_i\ F(X_i))\} \in \Phi$, where X_i is a multi-dimensional decision variable vector representing incident response strategies (i.e. ramp metering rate, opening of shoulder lanes, functioning of variable message signs, etc.), $F(X_i)$ is the objective evaluation value at location X_i, where i is the number of observed data point. We use different sizes of sample points derived from simulation function evaluation to estimate the underlying response surface using Surrogate Models. A simulated annealing algorithm will then be used to locate the global optima of the estimated response surface. The traffic assignment network model will be used to evaluate the system performance at the newly identified optima. The estimated surface will then be updated using information from this new point, dubbed as the interpolation procedure. This process will be repeated iteratively until: 1) the algorithm reaches the maximum number of iterations allowed, or 2) the improvement at the current iteration is less than a predetermined stopping criterion. The performance of each model will then be analyzed and compared in next section. The framework of this research is shown in Figure 2.

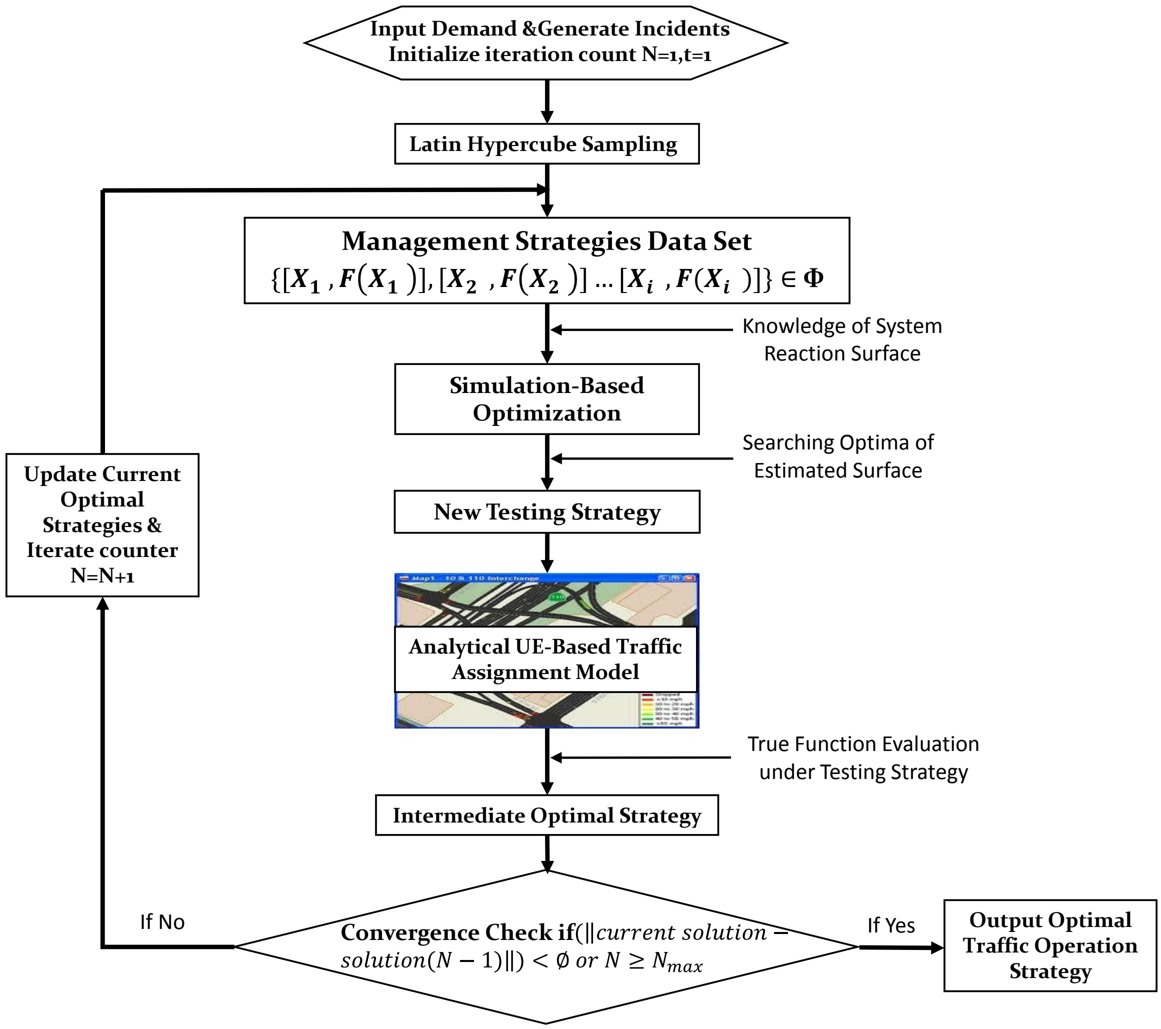

Figure 2. Framework of Surrogate Models.

Quadratic Polynomial Regression

This approach uses the sampled data points to regress the unknown surface. Least square minimization then can be used to determine the coefficient of each variable. QPR function can be expressed as Equation (1):

$$f(x, \beta) = \beta_1 + \sum_{m=1}^{d} \beta_{m+1} x_m + \sum_{m=1}^{d} \beta_{m+d+1} x_m^2, \quad x_m \in Data\ set\ (x_1, x_2, x_3, .. x_d) \qquad (1)$$

Where β_1 is the intercept, β_{m+1} and β_{m+d+1} are the coefficients to be estimated, d is the dimension of the decision variable, m is the m^{th} dimension. For more information about QPR, please refer to Bhattacharyya and Johnson (1997).

Radial Basis Function

Radial Basis Function has gained increasing attention on optimizing problems with unknown objective function. The basic idea of RBF is to use the summation of weighted basis functions to approximate a model (or a surface) that fits the observed data points. It assumes that the value of RBF only depends on the Euclidian distance of two data points. The response value of two points would behave more similar as these points closer to each other. RBF then treats every sampled data point as center and uses basis function to build approximated response surface around them. The analytical expression of RBF can be written as in Equation (2). The basis function that the most commonly used in RBF is Gaussian Basis Function as shown in Equation (3).

$$\boldsymbol{f}(\boldsymbol{x}) = \sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{n}} \boldsymbol{w}_{\boldsymbol{n}} * \boldsymbol{exp}(-\boldsymbol{\gamma}\|\boldsymbol{x} - \boldsymbol{X}_{\boldsymbol{i}}\|^{\boldsymbol{2}}) \tag{2}$$

$$\boldsymbol{\varphi} = \boldsymbol{exp}(-\boldsymbol{\gamma}\|\boldsymbol{x} - \boldsymbol{X}_{\boldsymbol{i}}\|^{\boldsymbol{2}}) \tag{3}$$

Where i is the index of observed data points; n is the number of observed data points; w_n is the weighting parameter for each observed data point; γ is the shape parameter; φ is Gaussian Basis Function. For more mathematical details about RBF algorithm, please refer to Björkman & Holmström (2000); Gutmann (2001); Wright (2003); Regis & Shoemaker (2005).

Since RBF is to use the summation of weighted basis functions to estimate a surface that fits the observed data points, RBF then turns nonlinear problems into a linear combination of basis functions with weighting parameters. Therefore, the problem is similar to simple regression, in which weighting parameters can be estimated by using least squared method or maximum likelihood estimation method.

Bayesian Optimization

Bayesian Optimization (BO) uses Gaussian Process (GP) to build the posterior distribution of the unknown objective function. Given the observed data set Φ, $\{(X_1, F(X_1)), (X_2, F(X_2)) \ldots (X_i, F(X_i))\} \in \Phi$, the posterior distribution of the function prediction at location $\boldsymbol{x}$ can be expressed as Equation (4):

$$\boldsymbol{p(f(x)|x, X_i, F(X_i)) = \int p(f(x)|x, F) * p(F|X_i, F(X_i))\, dF = GP(f(x)|\mu_*(x), \sigma_*{}^2(x))}$$

(4)

Where $F(X_\mathrm{i})$is the observed simulated system performance under strategy X_i, $f(\boldsymbol{x})$ is the estimated performance function prediction at strategy vector $\boldsymbol{x}$, $p(\boldsymbol{F}|X_i, F(X_i))$ is the posterior function distribution by given observed data, $\mu_*(\boldsymbol{x})$ and $\sigma_*{}^2(\boldsymbol{x})$ are the mean and variance of the posterior distribution to be estimated through the Gaussian Process. An important assumption of GP is that, every variable is jointly Gaussian-distributed. Thus, every observed data point from prior and predicted data points from posterior can be drawn from a multivariate Gaussian distribution with certain mean and variance. Also, GP is an approximation of a distribution over functions by given prior data points. Therefore, an estimation of the unknown function can be drawn from the posterior distribution (Rasmussen & Williams, 2006). GP can be represented in a joint distribution as Equation (5):

$$\begin{bmatrix} F(X_i) \\ f(\boldsymbol{x}) \end{bmatrix} \sim GP(\begin{bmatrix} \mu(X_i) \\ \mu_*(\boldsymbol{x}) \end{bmatrix}, \begin{bmatrix} K & k \\ k^{Transpose} & k^* \end{bmatrix}) \tag{5}$$

Where $\mu(X_i)$ is mean of the prior distribution of the simulated system performance function (which depends on the observed data points), K is a covariance matrix that depends on the observed data points only, k is a vector of the covariance matrix that is related to both the observed data points and a new testing data point, k^* is the covariance matrix that is only related to the testing data point.

Covariance matrix is consisted with covariance functions, also known as basis function or kernel function. There are several covariance functions available to use. We use Squared Exponential covariance function in this paper as Equation (6). For more information of selecting the covariance functions, please see Rasmussen & Williams (2006) and Gelbart (2015).

$$k(X_i, \boldsymbol{x}) = \sigma_f^2 * exp(-\sum_{m=1}^{d} \frac{1}{2l_m^2} /\!/X_i^m - \boldsymbol{x}^m/\!/^2) \tag{6}$$

Where i is the index of data point; d is the dimension of the decision variable, m is the m^{th} dimension; σ_f^2 is signal variance; l_m is length-scale parameter for the m^{th} dimension; $\left\|X_i^m - \boldsymbol{x}^m\right\|$ is the distance between two points. The shape of the posterior distribution is determined by signal variance σ_f^2 and length-scale parameter l_m in Equation (4). Given sampled data points, these hyperparameters σ_f^2 and l_m can be estimated by maximizing the marginal log likelihood showing in Equation (7):

$$Logp(F(X_i)|\, X_i) = -\frac{1}{2} F(X_i)^{Transpose} K^{-1} F(X_i) - \frac{1}{2} Log|K| - \frac{N}{2} Log2\pi \tag{7}$$

After finding the optimal hyper-parameters that best fit the observed data, the mean and variance of the posterior distribution of the objective function can be expressed as Equation (8) and Equation (9). For more mathematical detail information about GP, please refer to Rasmussen & Williams (2006).

$$\mu_*(\boldsymbol{x}) = k^{Transpose} * [K]^{-1} * f(X) \tag{8}$$

$$\sigma_*{}^2(\boldsymbol{x}) = k^* - k^{Transpose} * [K]^{-1} * k \tag{9}$$

After building the posterior distribution by given data points. BO algorithm then uses an acquisition function to determine at which point to be evaluated. One of the advantages of BO is that acquisition function can balance the global search between the aggressive exploitation for better solution and the exploration for uncertainty. This mechanism successfully reduces the possibility of stuck into the local optima. In this paper, we will use Expected Improvement (EI) to determine the next inquiry point. EI was initially introduced by Mockus et al. (1978) and extended by Jones et al. (1998). The point at which EI is the maximum is evaluated in the simulation for the true objective function value. This point with its function value is added to the current data set. BO uses the updated data set to revise the response surface again by using Gaussian Process iteratively.

Numerical Test

Analytical Testing Bed

Sioux Falls Network (Figure 3), which has been used often in the literature as a benchmark continuous network design problem (Bar-Gera, 2016; LeBlanc et al., 1975) was selected as testing traffic assignment network for this research. The Sioux Falls

Network has been widely used for network equilibrium traffic assignment problem. It has ground truth OD demand and well validated equilibrium link flows, which make this testing network is good for code debugging. The testing network consists of 76 links, 24 centroids and 360,600 trips. This network was digitized in MATLAB. The computation time for evaluating the single objective user equilibrium function for Sioux Falls Network is much shorter than the computation times required for simulation models. Furthermore, the output of this test network is deterministic which makes the model fully tractable. Most importantly, we used this test traffic assignment network over simulation model or a DTA model in this research because a theoretical and tractable solution is needed as a benchmark within feasible computational time budget. Traffic signals were not taken into consideration for the same reason. Thus, we will have a better understanding of the performance of the proposed framework and Surrogate Models. With that knowledge as confidence, in Chapter Four and Chapter Five, we will be applying the proposed optimization model on a simulation model which is more capable of capturing real world traffic dynamics.

BPR function was used to represent link travel time of each link as a function of free-flow time and volume-to-capacity ratio. Frank-Wolfe decomposition of the UE assignment problem was implemented to derive optimal solution for the given demand conditions. The assigned link flows and optimal objective function value under UE that resulted from the MATLAB implementation of Frank-Wolfe algorithm were compared to the results from the benchmark implementation of the Sioux Falls Network in prior studies. The relative errors of assigned volumes on all 76 links are below 1% and the optimal objective function value from the modeled network is 42.338816012 compared to the objective function value (42.31335287107) for the benchmark implementation. Thus, the Sioux Falls analytical network was deemed to be capable of being the test bed for this research.

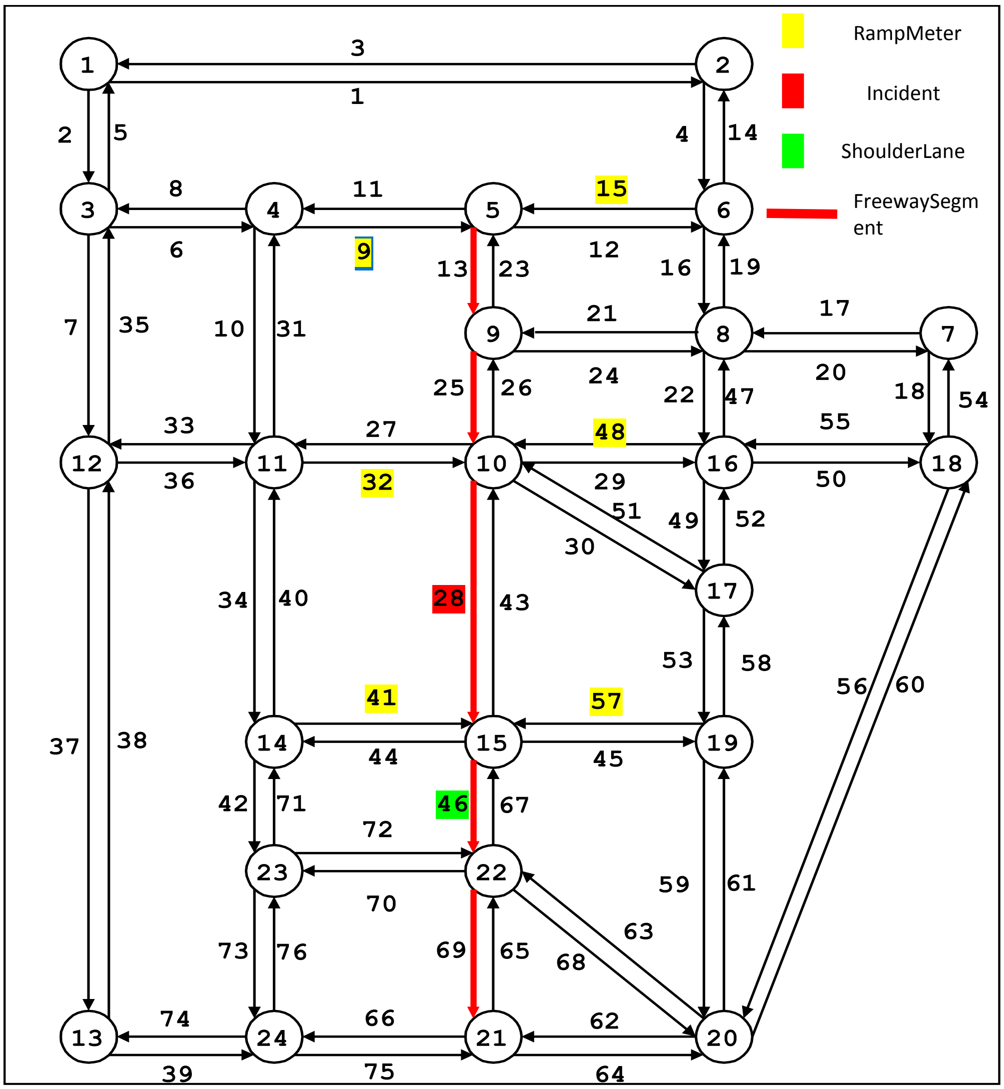

Figure 3. Sioux Falls Traffic Assignment Network.

It was assumed the links that are marked in red are freeway segments where have several on-ramp links (marked in yellow numbers) connected with the rest of the network. It should be noted that the links highlighted in red are not freeway links in the real Sioux Falls network. Likewise, we also ignore the real world network features because this test bed is purely hypothetical in nature. An incident was hypothetically installed on link 28 with half capacity reduction. Given the occurrence of an incident, multiple traffic operation strategies would be activated as incident responses to efficiently operate the traffic and reduce the total travel time for the entire network. We tested ramp metering and use of shoulder lane as two incident response strategies. Six ramp meters were installed on the adjacent links (Link 9, 15, 32, 41, 48 and 57) around incident link. The activation of ramp meters will add up a certain amount of headway (minuet) to the free flow travel time of the ramp link, and therefore dis-attract traffic flow flushing into the freeway link and therefore redirect traffic using other routes. The input of each ramp metering is from 0 minuet to 0.5 minuet. Furthermore, we installed a shoulder lane on Link46 (marked in green), which is located at the downstream link of incident location. The shoulder-lane will then be decided to open one lane as to 0.25 capacity increasing, or stay close as to 0 capacity increasing. The ramp meters and shoulder lane will then be working simultaneously to response the incident. A huge number of possible incident response strategy combinations will then be generated when 6 ramp-metering and 1 shoulder lane work cooperatively at the same time. Surrogate Models were then employed to determine optimal inputs for combination of these incident response actions.

The Surrogate Models output the minimum total network travel time, or the most improved one toward the theoretical optimal benchmark.

To start with a well-designed numerical test, the observed data points for training need to be uniformly sampled from the input-searching domain. In this study, we used Latin Hypercube Sampling (LHS) to generate training data points. LHS is a random sampling method for multi-dimensional input variables, and also a space-filling experimental design that uniformly draws sample data points from an equally stratified multi-dimensional search space without overlapping samples (McKay et al., 1979). The model has 6+1 input variables that are based on LHS, and 70 observation data points were sampled as the initial design (Jones et al., 1998). To investigate the performance of the Surrogate Models, this research then used LHS to generate data points with sample size 100, 150, 200, 250 and 300. All the proposed Surrogate Models were programmed in MATLAB.

Numerical Results Analysis

Due to the complex nature of the bi-level traffic assignment problem as discussed in the previous section, an exhaustive search was conducted to identify the theoretical benchmark traffic operation combination under incident condition. The test bed provides effective computation time for a single objective function evaluation of each incident response input combination. Since we have 7 decision variables, input for the six ramp metering scenarios are from 0~0.5 minute with an interval of 0.08 minute, and input for shoulder lane capacity increasing is either 1 or 1.25. We believe that the input of ramp metering with a 0.08 minute input resolution would give us a close enough approximation toward the true global optima. The objective function value would not jump too much within 0.08 minute interval because the BPR function has the nature of smooth curve. Thus, 232,598 different input combinations were evaluated on the test network and outputted total travel time for each input combination. These 232,598 data points were also used as testing data points to evaluate the surface fitting performance of the proposed Surrogate Models. The performance of the three proposed Surrogate Models were compared to their Root Mean Squared Errors (RMSEs) under different size of training data. RMSE is a very commonly used measurement for numerical validation of predictions via-a-vis true observations. The expression of RMSE can be written as Equation (10).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(F(\boldsymbol{x}_i) - f(\boldsymbol{x}_i))^2} \qquad (10)$$

Where n is number of observed data points, $F(\boldsymbol{x})$ is the observed data points, $f(\boldsymbol{x})$ is the model predictions. Figure 4 shows that BO and QPR both have much smaller RMSEs than RBF, which indicates that both models have good performance on surface fitting. RMSE of RBF is good and it keeps improving as the sample size increasing.

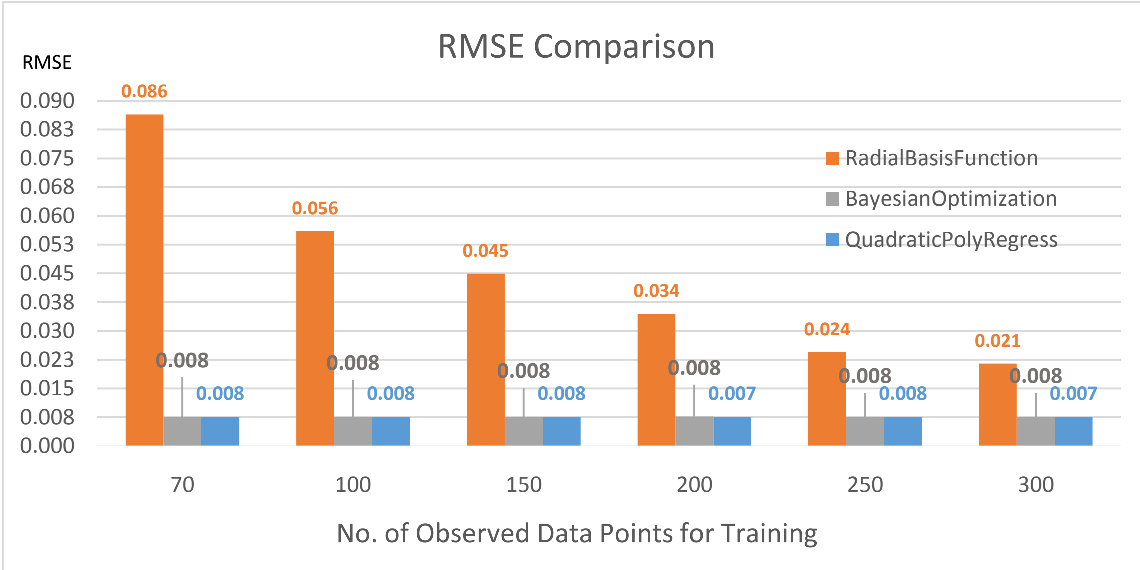

Figure 4. RMSE Comparisons.

After testing the surface fitting performance of each of the Surrogate Models under different training data size, we believe that all of these models fit the surface well. We then evaluated the capability of searching optimal solution. With the proposed framework that we developed in Figure 2, we fed the Surrogate Models with different size of the sampled data generated by LHS. Surrogate Models will then approximate an underlying surface that fits the given training data. An interpolation procedure will then be conducted. For RBF and QPR, simulated annealing algorithm will be directly applied for searching on the approximated response surface. For BO, EI will be used to balance the search between the locations where have lower means and locations where have higher uncertainty (variances). The optimal solution of the current estimated surface will be evaluated in the test bed and added into the current data set as an interpolant. Figure 5 indicates the comparison of performance on searching optima. All Surrogate Models obtained the improved optimal solution as sample size increasing. Furthermore, BO has superior performance on exploring better solution than the other two algorithms. This is because BO has the mechanism that EI offers, which explores the locations with larger uncertainty in predicting points not visited yet. BO performs the best when sample size is 200 or larger as indicated in Table 1. Based on the exhausted search, the minimum network total travel time identified is 8.1600E+06 minutes with an incident response strategy that is ramp metering setting up for 30,20,30,25,20 and 0 seconds for ramp metered Link 9, 15, 32, 41, 48 and 57, and 1 shoulder lane opening. The optimal solution BO searched are 8.1596E+06 mins, 8.1600E+06 mins and 8.1599E+06 mins with training data size 200, 250 and 300. The optimal solutions that BO obtains with sample size 200 and 300 are improved compare to the exhausted search. The incident response strategy that BO obtained with sample size 200 is ramp metering setting up for 0, 29.6, 29.6, 29.6, 30 and 29.7 seconds for ramp metered Link 9, 15, 32, 41, 48 and 57, and 1 shoulder lane opening. The incident response strategy that Bayesian Optimization obtains with sample size 300 is ramp metering setting up for 30, 20, 29, 24.9, 20 and 0 seconds for ramp metered Link 9, 15, 32, 41, 48 and 57, and 1 shoulder lane opening. As Figure 5 indicates, RBF has significantly improved when sample size is large. QPR model has the worst capability on optimal solution searching.

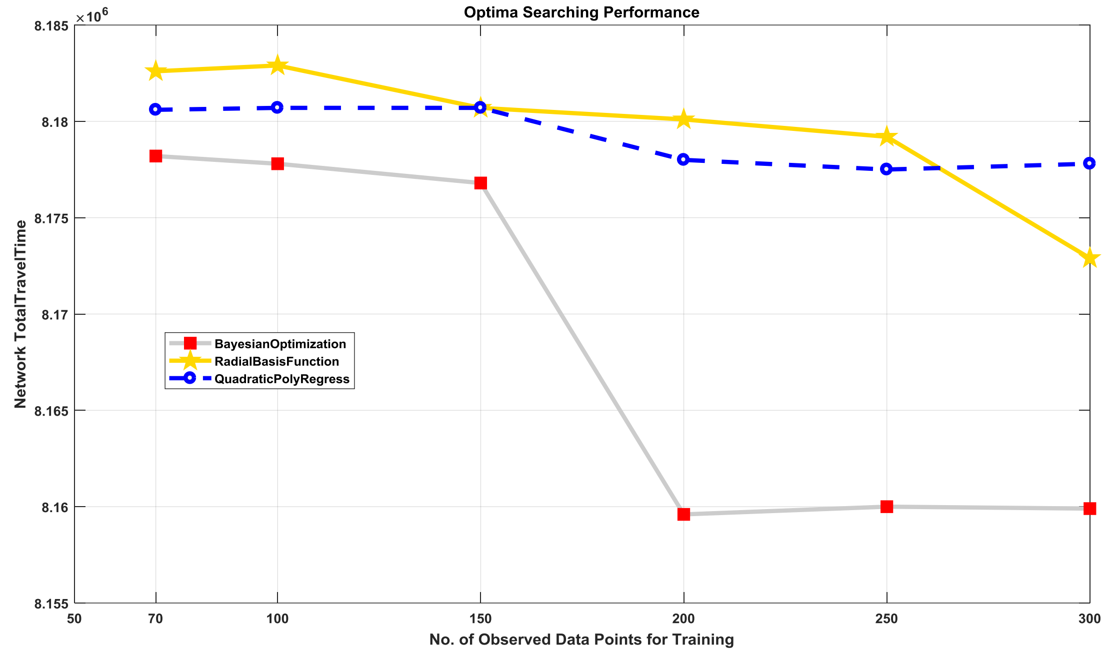

Figure 5. Optima Searching Performance

Table 1. Optimal Search Performance

	Optimal TotalTravelTime Each Surroagate Model Searched (min)					
No. of Observed Data Points for Training	70	100	150	200	250	300
QuadraticPolyRegress	8.1806E+06	8.1807E+06	8.1807E+06	8.1780E+06	8.1775E+06	8.1778E+06
RadialBasisFunction	8.1826E+06	8.1829E+06	8.1807E+06	8.1801E+06	8.1792E+06	8.1729E+06
BayesianOptimization	8.1782E+06	8.1778E+06	8.1768E+06	8.1596E+06	8.1600E+06	8.1599E+06

As Figure 6 shows, computation time for each Surrogate Models performed differently. QPR has a very time efficient computation performance. The superior performance of BO on optimal solution searching is traded off by the long computation time. That is because BO needs to calculate an N*N covariance matrix for predicting posterior distribution. The computation time would increase when we use large sample size for training surface. Also, computational time of BO with a sample size of 70 is higher than the requirement for sample sizes of 100 and 150 mainly because the model needs more iterations for interpolation when the response surface is sketchy. However, computation time of BO has greatly improved compare to the computation time of exhausted search, which is 188,238.4 seconds by more than 80%. RBF also requires large matrix calculation while approximating the response surface. Figure 7 shows a comprehensive performance comparison of all the proposed Surrogate Models. Among all the models, BO yields the best performance yet with relatively high computation time. We consider the great computation time of BO is a good trade-off for obtaining the most improved solutions within several hours. It is always good to get a better solution rather than blindly searching where limitless possible input scenarios are always impossible to be evaluated under limit time budget even on an analytical network model. It would be a huge computation advantage if we look for the optimal traffic operation solution on the realistic simulation roadway network. RBF has moderate performance on both optimal solution finding and computational efficiency. QPR model shows poor optimal solution searching yet very quick computation time. This trait can be valuable when a short-term performance is needed for incident response management.

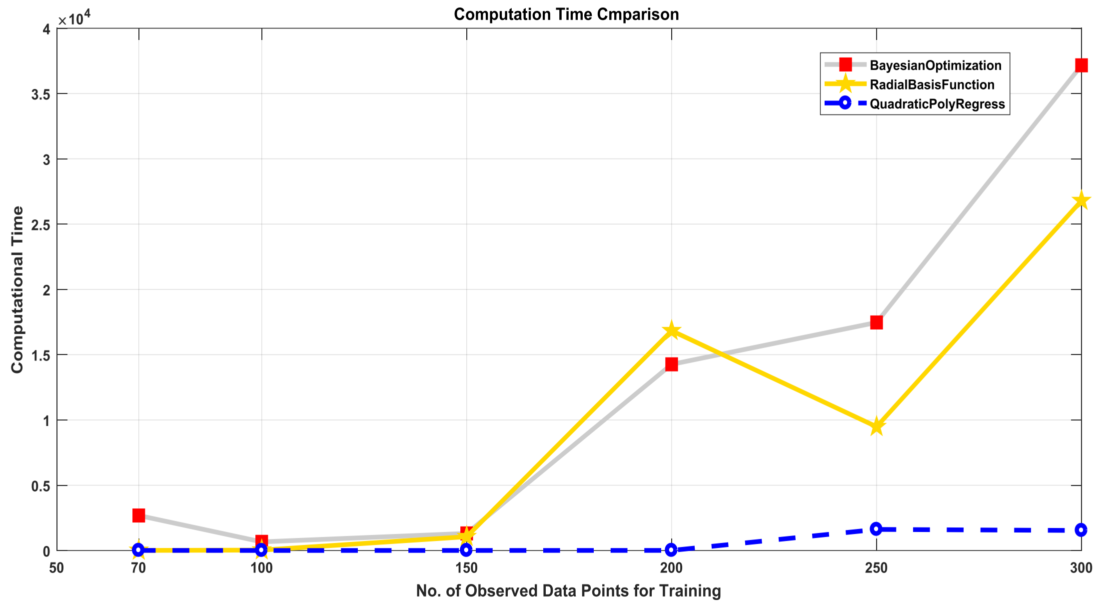

Figure 6. Computation Time Comparison

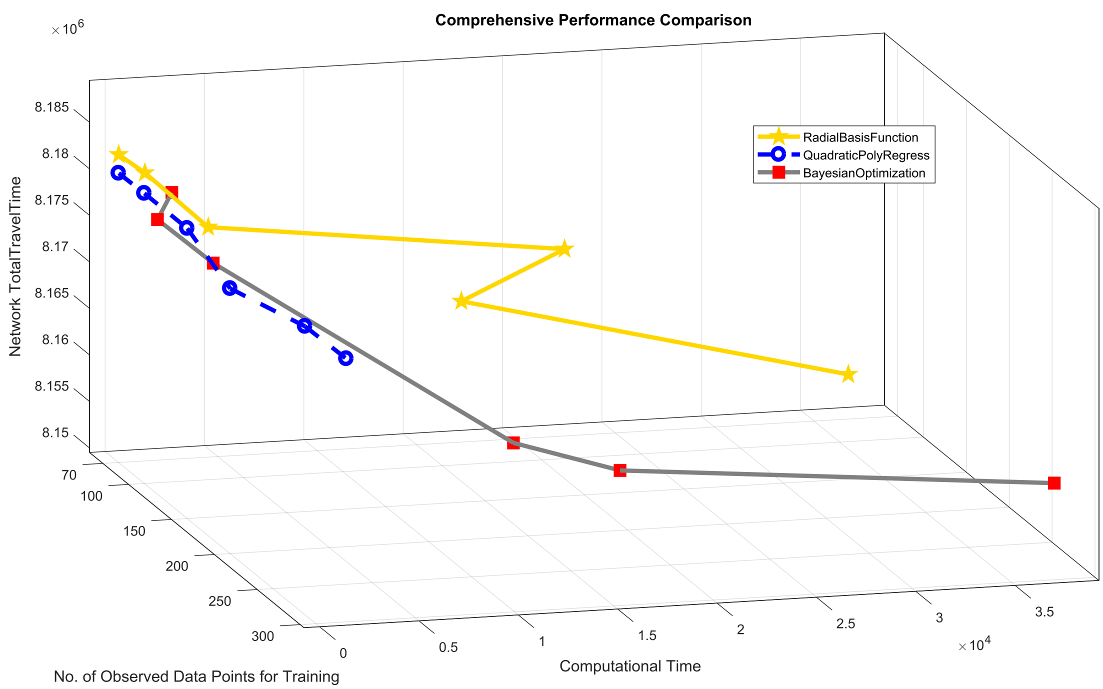

Figure 7. Comprehensive Performance Comparison

Conclusions

Traffic incidents are the most frequent contributors to traffic congestion. Using all traffic operation tools effectively in response to traffic incidents becomes critical for mitigating traffic congestion and is beneficial to the urban transportation systems. This research addressed the optimization response strategy problem that has no explicit objective function and derivative information. In this paper, a framework that integrates the analytical traffic assignment model and Simulation Based Optimization methods was developed to identify the optimal traffic operation strategy combinations under incident condition. Based on the numerical tests on an analytical network, QPR model has weak performance on optimal solution searching. However, the computational efficiency of QPR model has outperformed other Surrogate Models. This finding can be valuable when a short-term performance is needed for incident response management. RBF has average capability on identifying improved solution when sample size gets larger. BO has the best capability on searching the most improved solutions yet needs to trade off some computation time. BO required more computing time, but is much more superior in finding the optimal solution. In our example, it even outperformed the exhaustive search on a discrete grid that used significant longer computing time. It offers huge benefits since it is impractical to search every possible input for a high-dimensional domain in the real world.

For the future research, the proposed Surrogate Models will be applied on a larger scale microscopic simulation network, which has untraceable underlying structural information and expensive computational burden. The microscopic simulation model will capture the traffic dynamics and vehicle behavior more realistic and offers even more traffic operation options for incident response. Surrogate Models that search for the most improved incident response strategy in a reasonable time duration for a microscopic model would be considerably beneficial for the incident response management.

CHAPTER FOUR IDENTIFYING OPTIMAL TRAFFIC OPERATION STRATEGY ON A LARGE-SCALE SIMULATION NETWORK

Introduction

Chapter Three showed that it was feasible to use the simulation-based optimization framework to efficiently identify the optimal control strategies that minimize the system total travel time. This framework improves the computing efficiency by exploring the shape of the system response surface as a function of control strategies and reducing the number of trials through a guided process. In real world, the system performance as a function of control strategies is usually highly non-linear due to complex system dynamics. Large-scale traffic simulation model can capture the complex traffic dynamics and help predict the system performance, as shown in many existing traffic simulation studies. In this study, we will replace the analytical model which used in Chapter Three with a large-scale time-dependent traffic simulation model, and investigate the performance of the proposed method in real-world problems. The implementation of the traffic operation controls in a time-dependent network at multiple location will lead to a high dimension problem. Therefore, in this chapter, we will introduce a sequential approach as an extension of the proposed optimization framework to maintain the scalability of the proposed framework, which will reduce the searching space in the temporal dimension.

The next section briefly reviews previous research efforts on improving traffic operations using simulation models. The proposed optimization framework will then be presented. The proposed method will be demonstrated through a case study on a regional network in Northern Virginia. This case study will show how the proposed framework gradually constructs the response surface of various combinations of traffic operation strategies, while at the same time guides the search for optimal solutions. Two other commonly used SBO methods, Quadratic Polynomial Regression (QPR) and Radial Basis Function (RBF) will also be tested and compared with the proposed Bayesian Optimization (BO) model. Although the proposed framework is applied for incident responses in this study, it could also be applied to investigate optimal traffic operations strategies in other cases.

Relevant Studies

To investigate the performance of different traffic operation management tool for reducing incident related congestion, traffic simulation models are the most commonly and widely adopted approach. Sufficient of previous studies used simulation models to evaluate the performance of pre-determined traffic operation strategies under various scenarios, many of which are in the context of incident responses (e.g. ramp control, lane control, traffic signal) under various scenarios, many of which are in the context of incident responses (Hasan, 1999; Pulugurtha, 2002; Ben-Akiva et al., 2003; Zhou, 2008; Jiang and Chung, 2014). In some of the early research on real time incident response, Southworth et al. (1992) proposed a dynamic traffic assignment (DTA) modeling framework for responding mass evacuations in real time. Although many other researchers used various simulation models (macroscopic, mesoscopic, and microscopic) to evaluate the effectiveness of a wide range of traffic operation strategies (e.g. shoulder lane, ramp metering, variable message sign, toll road, etc.) (Chauhan, 2003; Chu et al., 2004; Wirtz et al., 2005; Chou and Miller-Hooks, 2011; Papayannoulis et al., 2011; Vadde et al., 2012; Dion et al., 2012; Xiong et al., 2015; Xiong et al., 2017; Lee et al., 2017), most of these efforts were still limited to scenario-based evaluation of a single strategy at a time and did not search for the optimal solution, especially in the context of combining multiple strategies. The problem becomes intractable because of the large number of strategy combinations and the long computing time for simulating a large network.

To find the optimal operation strategies quickly, a clear and well-behaved objective function (such as those adopted in analytic models) is needed. However, such an objective function is usually unavailable if we want to capture complex traffic dynamics through simulation models. Therefore, we need to develop an efficient method to estimate the objective function and the optimal solutions simultaneously. Surrogate Models (also known as Response Surface Model or Meta-Model), a subcategory under SBO, is designed for solving the optimization problem with unknown objective function (Fu, 2002; Søndergaard, 2003; Barton and Meckesheimer, 2006; Kleijnen, 2008; Ammeri et al., 2010; Amaran et al., 2014 and Xu et al., 2015). These models use system performance information from sampling points (i.e. simulation outputs) to approximate the statistical structure or response surface that represents the relationship between simulation inputs and outputs. Although the applications of SBO method in incident response studies are still preliminary, it has been used in traffic operation related fields. SBO approaches were proposed to optimize signal control plans on different roadway networks to reduce congestion (PIZANO, 2010; Osorio & Bierlaire, 2013; Osorio & Chong, 2015; Chong & Osorio, 2018). He (2014) used different SBO methods to evaluate the unknown objective function of a large-scale toll road system and used Genetic Algorithm to optimize toll rate. Chen et al. (2014) used kriging method to find the toll rate that minimizes the average travel time. Zhang et al. (2014) also applied kriging method to investigate the usage of HOT lanes and Dynamic message signs. Pal & Sinha (2000) might be the first who applied SBO method in incident response studies. However, their study is still limited to a single response strategy. To bridge the gap, this research will integrate mesoscopic traffic simulation models with SBO methods to solve the optimal incident response strategy problem while still considering the complex traffic dynamics on a large network. There are three mainstream Surrogate Models: Quadratic Polynomial Regression, Radial Basis Function and Bayesian Optimization models. Please refer to Mockus et al. (1978); Jones et al. (1998); Björkman & Holmström (2000); Gutmann (2001); Wright (2003); Regis & Shoemaker (2005); and Gelbart (2015) for more information. All of these three Surrogate Models will be integrated into the optimization framework and tested.

Methodology

Integrated Framework

To mitigate traffic congestion under different incident scenarios, traffic management agencies may consider a wide range of strategies. These strategies may include decision variables that are either discrete (e.g. whether to open a managed lane or not), or continuous (e.g. optimal rate for ramp meters). Considering these strategies simultaneously will lead to a large number of strategy combinations. Implementing these strategies will lead to complex traffic dynamics on the network, which makes the system performance (e.g. measured through the total travel time) hard to estimate or predict in advance. The large number of scenarios and significant computing time even for evaluating a single scenario for a large-scale simulation model makes it infeasible to exhaustively test all possible scenario combinations, while the complexity of system dynamics makes analytical model inapplicable. To address these challenges, this research proposes an optimization framework that integrates mesoscopic traffic simulation models and simulation-based optimization methods to find the best performing (i.e. minimum total travel time) incident response strategies on a large network. The proposed framework is shown in Figure 8.

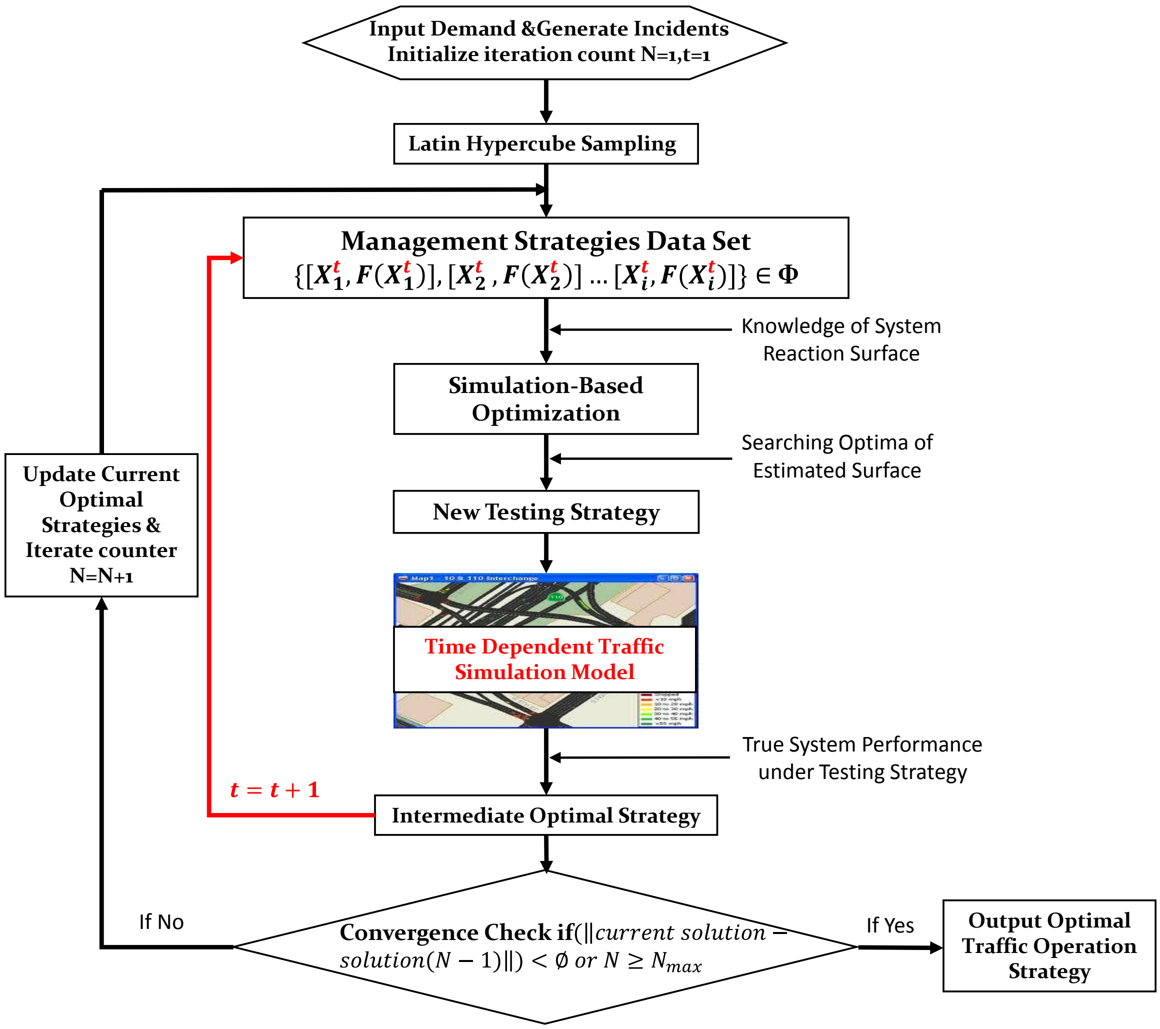

Figure 8. Framework of the Incident Response Optimization Model (Adapted from Chapter Three, New Components Highlighted in Red)

The objective of the optimization problem in this study can be generally formulated as Equation (11). It minimizes total travel time by finding the optimal traffic operation strategies.

$$\text{argmin } f(\boldsymbol{x}) = \sum_{t=1}^{T} F(x_1^t, x_2^t, \ldots x_m^t \ldots, x_d^t, O) \qquad (11)$$

st to LB$\leq x_m^t \leq$UB

$0 < O \leq$ Summation of t

Where $\boldsymbol{x}$ is multi-dimensional decision variables representing traffic operation strategies (e.g. ramp metering rate, opening of shoulder lanes, variable message signs, etc.), *m* is the *m*th active traffic operation strategy among the total number of d strategies to be considered. In practice, the decision-making period is usually discretized. For example, the metering rate or VMS will only be changed once every 15 minutes. Therefore, *t* represents the index of time interval within the study period. T is the total number of intervals that were divided in the optimization process. $F(x_m^t)$ represents the value of targeted performance measure (e.g. Total Travel Time) at time period *t* given an incident scenario *O* and a series of traffic operation strategies. Because of the complexity in traffic system dynamics, the objective function can only be evaluated through simulation models. Limited by computing resources, researchers can only evaluate the performance of a certain number of scenarios. Therefore, an algorithm is needed to help researchers to estimate the shape of the objective function (or the response surface) and identify the minimum with limited information. One significant advantage of the proposed framework based on simulation-based optimization methods is that it provides guidance on the best testing scenario and help researchers efficiently identify the optimal point through an iterative process.

To start the optimization process, we first use Latin Hypercube Sampling (LHS) method to generate training data points and store them into data set Φ. Each data point represents one scenario based on execution of the simulation model. LHS is a random sampling method for multi-dimensional input variables. It is a space-filling experimental design that uniformly draws sample data points from an equally stratified multi-dimensional search space without overlapping samples (McKay et al., 1979).

For a given data set Φ,

$$\{(X_1^t\ \mathrm{F}(X_1^t)), (X_2^t, \mathrm{F}(X_2^t)), (X_3^t, \mathrm{F}(X_3^t)), \ldots\ldots (X_\mathrm{i}^t, \mathrm{F}(X_\mathrm{i}^t))\} \in \Phi, \quad (12)$$

Where,

X_i^t is a multi-dimensional decision vector representing all operation strategies at optimization time interval *t*.

$\mathrm{F}(X_\mathrm{i}^t)$ is the value of the performance measure (e.g. Total Travel Time) under operation strategy X_i^t, *i* is the index of i^{th} strategy tested through the simulation model.

The sampled data will then be used to build the response surface of the objective function to various operation strategies using a BO Algorithm. In each iteration, the algorithm will generate an optimal solution based on current estimation of the response surface. This strategy will be tested in the simulation model and its true performance will be evaluated. The true performance may be different from the estimated performance, and will help researchers to better estimate the response surface in the next iteration. The tested strategy and the simulated performance value will become a new data point added to the existing data set Φ. This iterative process will continue until an optimum that satisfies a pre-determined criteria is found.

Depending on the chosen temporal resolution (duration of each decision period), the number of decision variables could increase rapidly. To further reduce the dimension of decision variables, we divided the entire time periods into a few sub-time intervals and consider all time periods in an iterative process. We will fix the optimal strategies that have been identified for all previous time periods and consider strategies for time interval t+1. Therefore, the proposed sequential optimization framework has a double loop structure. Inner loop searches for the best traffic operation strategies at time interval t. The outer loop will search for the optimal strategy for each time period sequentially and the algorithm will go back to t=1 after reaching the last time period. In each iteration of the optimization process, a SBO method is used to 1) estimate the response surface based on current knowledge of the simulated system and 2) identify the optimal strategy based on the current response surface. In this study, Bayesian Optimization (BO) algorithm is adopted for the SBO step because it provides a mechanism to balance the consideration of best solution under current response surface (based on the mean of the objective function) and the uncertainty within the optimization procedure (its distribution) (Rasmussen & Williams, 2006; Gelbart, 2015). One advantage of the BO method compared to other SBO methods is that BO uses acquisition function to balance exploration and exploitation. This mechanism could reduce the possibility of stuck in local optima and output the most promising trial point to test in the next iteration. In this paper, we will use Expected Improvement (EI) method to determine the next trial point (Mockus et al., 1978; Jones et al., 1998). The EI method selects the point that would maximize the expected improvement of the objective function as the testing point for the next iteration instead of the point that maximize the improvement over the mean of the objective function. Although other algorithms could also be used, this study uses the simulated annealing algorithm to search the maximum of EI. The proposed framework will run the simulation model using the strategy that maximizes the EI of the objective function, and repeat the same process until the stopping criteria is met.

As a comparison, this study will also implement two alternative SBO algorithms, Quadratic Polynomial Regression (QPR) and Radial Basis Function (RBF), in the proposed framework. This comparison would show pros and concerns of different SBO algorithms and provide additional guidance on algorithm selection to researchers and practitioners. QPR estimates the unknown surface using regression methods. RBF uses the weighted summation of a series of basis functions (e.g. Gaussian Basis Function) to approximate a model (or a surface) that fits the observed data. In this way, RBF turns nonlinear problems into a linear combination of basis functions with weighting parameters, which makes it popular in solving optimization problems that are highly nonlinear and with high-dimensional decision variables. The weighting parameters in RBF can be estimated using either least squared method or maximum likelihood estimation method.

Experimental Performance of Integrated Model

Simulation Model

In this section, we investigated the performance of the proposed optimization framework on a large network in Northern Virginia. The network was developed in a DTALite, a mesoscopic simulation model (Lu et al., 2013). The study area is shown in Figure 9.

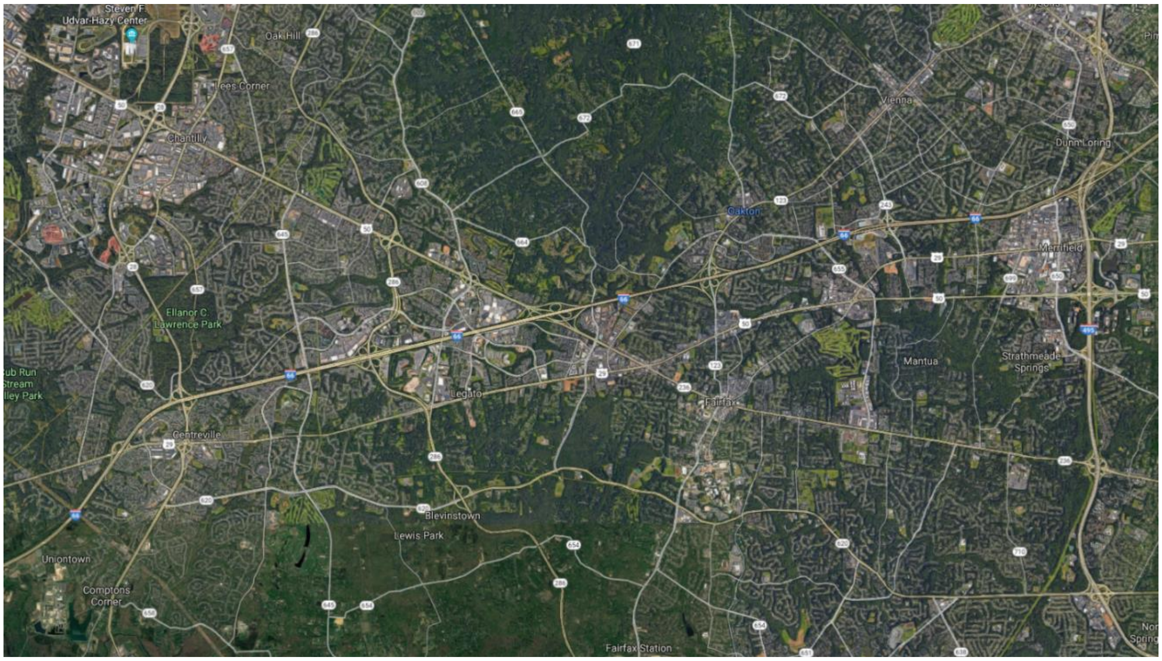

Figure 9. Study Area in Northern Virginia

The simulation network includes all freeways, and major and minor arterials in the study area, which are shown in Figure 10. It includes 1,937 links, 794 nodes, 189 traffic analysis zones (TAZs). This study focuses on afternoon peak hours from 16:00-19:00 for simplicity, although it could be extended to other periods using the same method. The initial OD for the study period was estimated from the Metropolitan Washington Council of Government (MWCOG) planning model through sub-area analysis, while additional estimation and calibration efforts were conducted to sub-divide the peak period OD tables into more detailed OD matrices.

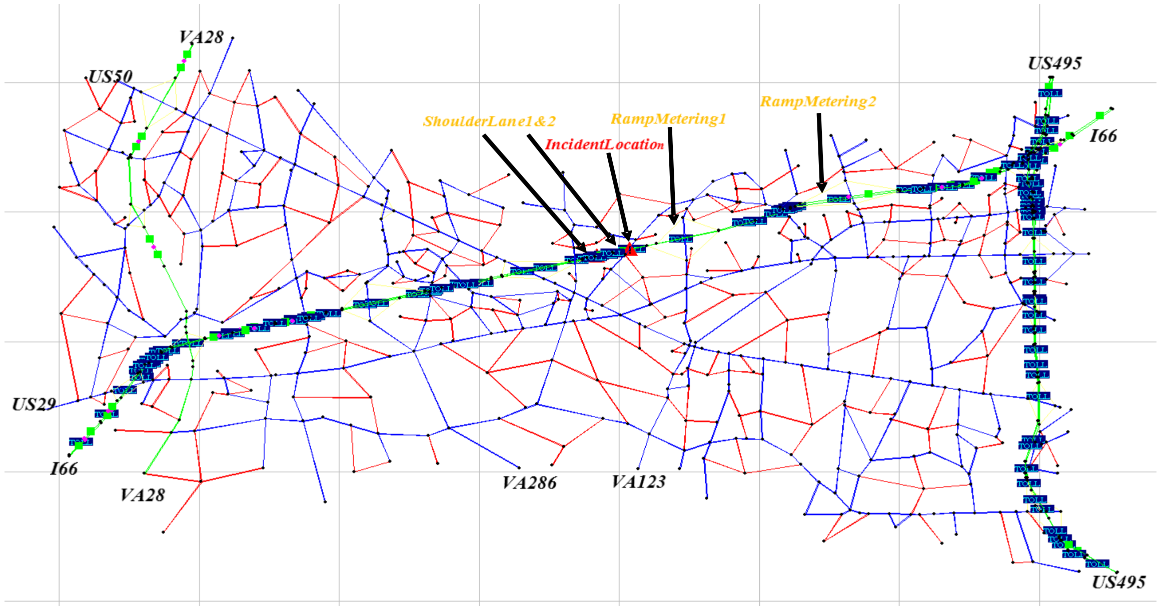

Figure 10. Simulation Network in DTALite

Model Calibration and Validation

The main goal for calibration and validation step is to replicate the observed traffic flow rate and corridor travel time as close as possible. In this study, speed is also considered to make calibration more reliable. The field traffic flow counts data and speed data were obtained from the VDOT Performance Measurement System (PeMS). Traffic data of all weekdays in May of 2015 was selected for calibration because the data quality from most sensors within the study area was good during this period. The locations of traffic count and speed detectors of observed traffic data from PeMS were marked in the green boxes and pink dots respectively in Figure 10. There are totally 47 traffic counts and 42 speed sensors in the study area. We used 15-minute time intervals as the temporal resolution, which leads to 540 average traffic volume and 480 speed observations for a period between 16:00 and 19:00. The static OD matrix estimated from MWCOG planning model was also divided into a series of 12 time-dependent OD matrices based on the aggregated percentage of traffic observed during each 15-minute period over the entire study period. A dynamic OD estimation process was then conducted to estimate the 12 OD matrices in order to match the traffic volumes and average speeds observed during the each period at all locations with sensors. This process was developed in previous studies focusing on the same metropolitan areas (Xiong et al., 2015; Zhang et al., 2013). The main idea is to identify the OD pairs that use links with sensors and adjust the time-dependent OD demands based on the gap between the observed traffic flows and simulated OD flows, the path flow patterns, and the time it takes for demands from a particular OD to arrive at the link with sensors. More information on dynamic OD estimation process can be found in Xiong et al. (2015) and Zhang et al. (2013).

presents the calibration results and compares the observed traffic flow and speed with the simulated volume counts and speed. The calibration results show that the Mean Absolute Error (MAE) for traffic counts and speed are 50.5 vehicles and 3.2 miles/hour, respectively. Relative Root Mean Square Error (RRMSE) for traffic counts and speed are 11% and 9.1% respectively. Mean Absolute Percentage Error (MAPE) is 9.56% and 9.39%, while Weighted Mean Squared Error (WMSE) is 10% and 8.6% for volume and speed, respectively. These calibration results are within the range of accuracy achieved in previous studies for a large-scale mesoscopic simulation model.

We also conducted an additional model validation by comparing the corridor travel time along I-66 westbound collected with probe cars with simulated travel time. Since this study focuses on afternoon peak hour, I-66 westbound is the direction of major congestion. The probe car travel time data was collected by driving from the east end to west end of I-66 within the study area. Ten days of probe car travel time data were collected for each 15-minute time interval. Figure 12 compares the average corridor travel time measured using probe cars and the simulated corridor travel time after model calibration.

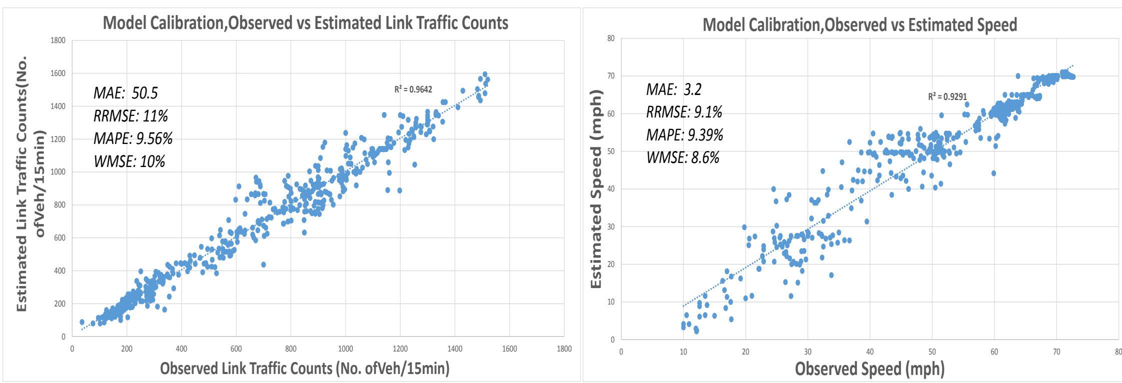

Figure 11.Comparison between Observed and Estimated Link Counts and Speed

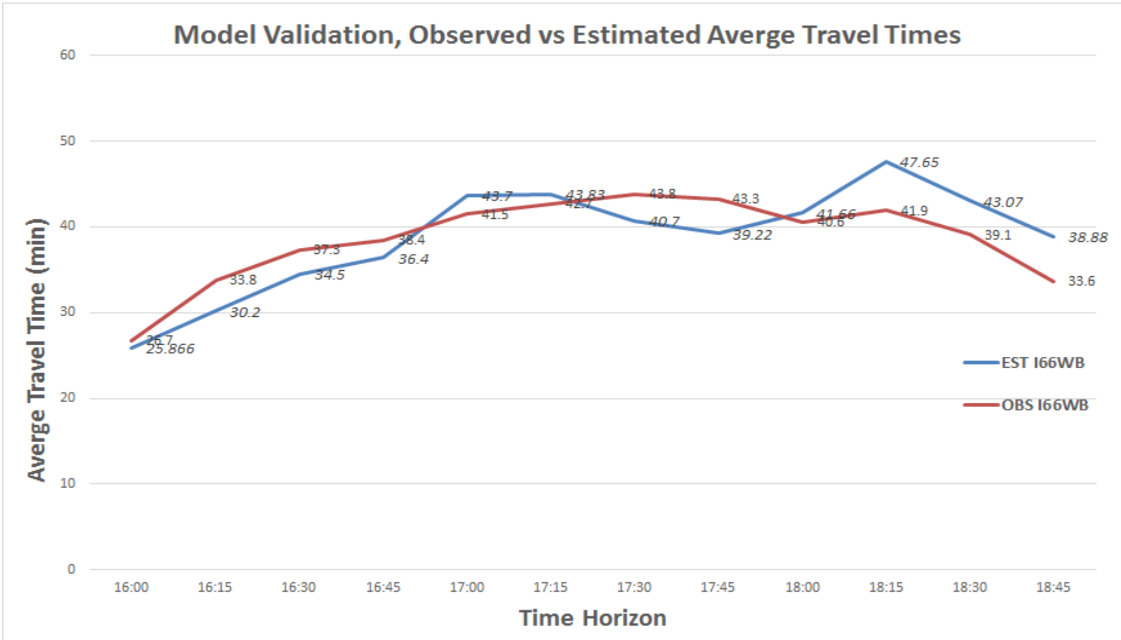

Figure 12. Observed vs. Estimated Corridor Travel Time on I-66WB

Integrated Optimization Model

This section demonstrates the proposed optimization framework that integrates the simulation model with SBO algorithms through a case study based on the calibrated model for the study area. As Figure 10 shows, an incident closes one General Purpose lane on I-66WB at 16:30. To mitigate the non-recurrent congestion, we consider strategies that include opening of a shoulder lane at the incident location and the freeway segment immediately downstream of the incident location, and 2 ramp meters at on-ramps upstream of the incident location. Shoulder lane will increase the capacity around incident location where the capacity drops due to incident. Ramp meters control the traffic flows entering freeway segment to further control the demand. The decision variables for should lanes are discrete, while the ones for ramp meters could be continuous. Both decisions include a series of decision variables for all periods following the incident. These combinations (multiple type of strategies, locations, and time periods) lead to a large number of possible scenarios, and complex interactions with traffic dynamics. An efficient algorithm to identify the optimal solution among these scenarios is needed, but has not been well studied in the literature. Previous studies on ramp meter algorithms rely on reading of local traffic volumes and focus on maintaining smooth operation of local freeway traffic flows. Thus, these algorithms do not address challenges we have in the current study.

Since we choose 15 minutes as the time resolution for OD matrices, we also assume the metering rate for ramp meters will only be adjusted once every 15 minutes for consistency. However, this choice is not a constraint and can be changed without loss of generality. We consider the adoption of response strategies after the incident occurred at 16:30. Therefore, there are totally 10 decision variables for ramp metering 1 and another 10 for ramp metering 2, one for each 15-minute period during the remaining of the simulation period. A dummy variable will represent the decision of whether or not to open the shoulder lane each of the two locations and for each time period. Therefore, there are totally 20 continuous and 20 0/1 decision variables to be determined. Therefore, we first uniformly sample 40*10=400 sample points (Jones et al., 1998) within the searching region using LHS to start the optimization process. The optimization process is time-consuming. To further reduce the searching space and improve the efficiency of the algorithm, we further divide the decision variables in three groups based on chronicle order. The proposed algorithm first optimizes the decision variables from 16:00-17:00 as

t=1 while fixing the rest of decision variables using default inputs. Then decision variables from 17:00-18:00 as t=2 are optimized while taking the optimized decision variables from 16:00-17:00 as fixed inputs. Finally, we optimized variables from 18:00-19:00 as t=3 while holding the decision variables from the two previous time intervals. The algorithm will revisit variables in the first period and repeats the process. The stopping criteria is reached after 12 iterations. The optimal traffic operation strategies include the opening shoulder lanes at both locations for all periods and a metering rate of [42,25,33,74,75,78,64,5,7,5] and [80,83,54,36,23,48,93,5,7,5] for ramp 1 and ramp 2 respectively. The number here represents the corresponding capacity reduction rate at each metering location and for the corresponding time period. The optimal network total travel time $(X) = 4.28666514037 * 10^6$ minutes.

The value of objective function (system total travel time) at each iteration is shown in Figure 13(a). It clearly shows that the optimization system based on BO algorithm converges after 9 iterations and the total travel time of the entire system becomes flat. We also compared the performance of the BO algorithm with two other commonly used SBO methods by integrating them in the proposed framework and conducting the test using the same set of sample data points. The comparison result is shown in Figure 13(b). The QPR and RPF algorithms converged with much fewer number of iterations, but they could not reduce the total travel time as much as the BO-based algorithm did. A tradeoff has to be made between better accuracy and longer computing time. Figure 14 compares the computational time of the integrated models with different SBO algorithms.

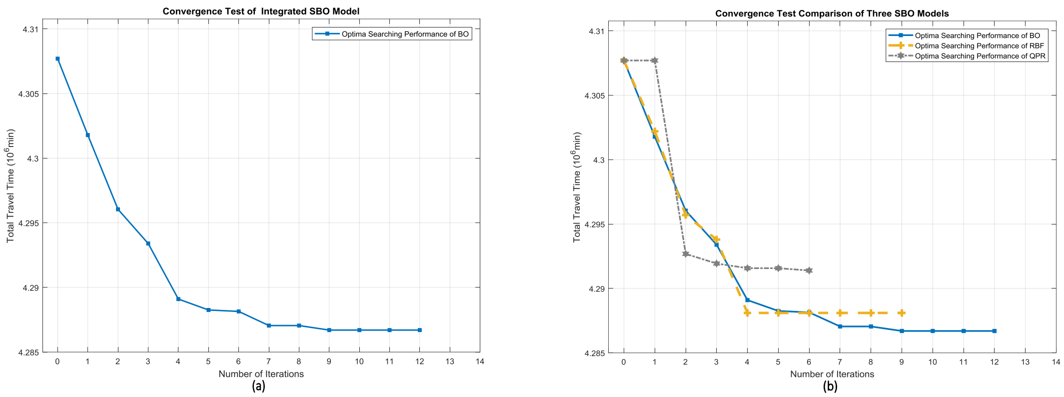

Figure 13. Convergence of System Travel Time

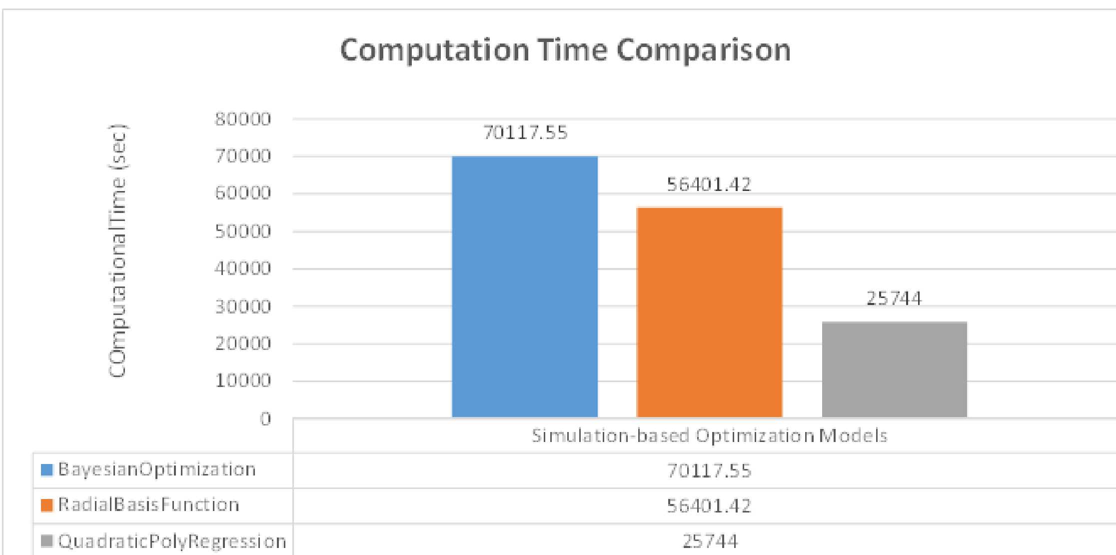

Figure 14. Computing Time Comparison of the Three SBO Algorithms

System Performance

Table 2 compares the system-wide performance between different scenarios: with the incident but no actions, activating the shoulder lanes only, and activating both shoulder lanes and ramp meters following the control strategies optimized using the three different algorithms. As a comparison, we also report the total travel time under the scenario with neither accidents nor controls to show the relative magnitude of improvements under different scenarios.

The incident closed one lane on I-66WB. Therefore, if we open the shoulder lane immediately at the affected freeway segment, we restored the capacity and the overall performance of the system is similar to the scenarios with no accidents and no controls. However, if we combine the extra capacity provided by the shoulder lane with demand control strategies implemented through ramp meters, the system would perform much better. The system travel time savings could be improved by 80% more (1.7% vs 0.94% travel time savings when compared to the incident only scenario) when meter controls were introduced.

Similar to what Figure 13 shows, the optimization framework based on the three different simulation-based optimization algorithms provide similar performance in improving the total travel time. But the difference is still observable, with BO method outperforming the RBF and QPR methods.

Table 2. Comparison of System-wide Total Travel Time under Different Scenarios

Scenario Name	TotalTravelTime (*10^6 min)	TotalTravelTime Saved (*10^6 min)	%Improvement
No Incident	4.32127832	-0.037922698	-0.9%
Incident	4.359201018	N/A	N/A
BO	4.28666514037	0.072535878	1.7%
RBF	4.288095309	0.071105709	1.63%
QPR	4.291394	0.067807	1.56%
Incident+SLOnly	4.31832854	0.040872479	0.94%

Table 3 first compares the performance of the ramp meter control strategy that was optimized under the same framework, but in a scenario with no incident and no shoulder lane operation. The metering rate is still adjusted every 15 minutes. The results show that we can still reduce the system-wide total travel time by activating the ramp meter system even without an incident. However, the magnitude of improvements is smaller when compared to the scenarios with incident. Therefore, the additional benefits for introducing the demand control through ramp meters are more significant when incidents occur and the system becomes more congested. And it works better when activated in coordination with other control strategies such as shoulder lanes.

Table 3. Performance of Ramp Meter System under No-Incident Scenario

Scenario Name	TotalTravelTime ($*10^6$min)	TotalTravelTime Saved ($*10^6$min)	%Improvement
No Incident	4.32127832	N/A	N/A
No Incident + Ramp Meters Only	4.290510771	0.030767549	0.71%

In addition to the system wide performance, we also compare the corridor performance using the speed contour to visualize the impact of different control strategies. Figure 15a shows the congestion pattern when an incident occurs at 16:30pm. All freeway segments on I-66WB upstream of the incident location are very congested (the color shows average speed). Figure 15b and Figure 15c show the congestion patterns when the shoulder lane is activated alone, and when both shoulder lanes and ramp meters are activated and operated according to the optimal strategies identified through the proposed framework. The congestion on freeway segments upstream of the incident location is clearly improved in both scenarios, and the improvements under the scenario with both strategies implemented are more significant. Because strategies implemented in Figure 15b and Figure 15c improved the traffic flow and throughput at the incident location, the segments downstream become slightly more congested. Therefore, these figures need to be combined with system-wide performance analysis as shown in Table 2 to provide a comprehensive evaluation of the performance of different strategies. In both cases, the advantage of the proposed optimization framework is clearly illustrated.

As more large-scale traffic simulation models are introduced in practice to analyze the transportation system performance in practice, it is important to develop tools that could be integrated with such models to improve traffic operations and management effectively. The long computing time associated with such models (even with mesoscopic ones) is always a challenge. The optimization framework developed in this study could be applied to improve operation strategies outside to incident response field. To illustrate, Figure 15d and Figure 15e compare the corridor congestion patterns without incident, but with ramp meters activated. The improvement is also clearly visible in this scenario. The proposed framework could be further extended to consider other operation strategies, such as VMS and road pricing. Future research will address those questions.

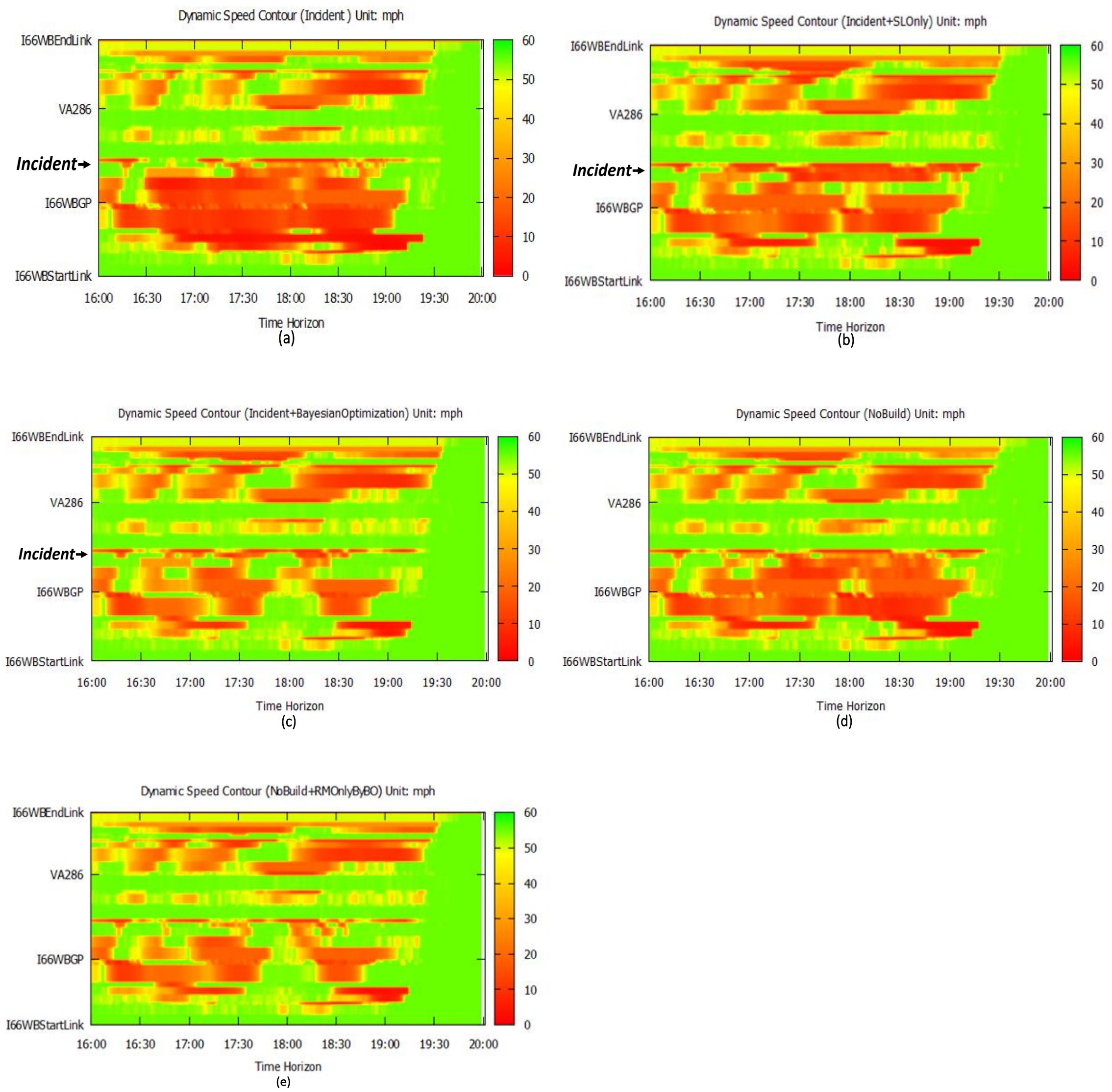

Figure 15. Corridor Travel Speed Heat-map under Different Incident and Control Scenarios

Conclusions

Traffic incidents are major contributors to traffic congestion in urban areas. Traffic operators and management agencies could mitigate the non-recurrent congestion using a wide range of tools and strategies. These strategies could act in synergy or counter-productively. It is important to quickly evaluate them and identify the best response strategies. However, given the number of potential strategies, the possible choices of control parameters (discrete and continuous), the temporal resolution, and multiple control locations, the number of scenarios to be considered will increase exponentially. Traffic simulation model could better capture complex traffic dynamics compared to the analytical traffic flow models, but the long-computing time combined with the large number of possible scenarios prevent it from widely applied in incident response strategy optimization studies. This study proposed an optimization framework by integrating a mesoscopic traffic simulation model with simulation-based optimization algorithms to address these challenges.

The application in a case study on a large network in Northern Virginia area demonstrates that the proposed framework could effectively reduce the dimension of the searching space and identify the best operation strategies within reasonable time. The case study shows that the Bayesian Optimization-based system outperforms the framework based on Radial Basis Function and Quadratic Polynomial Regression methods in improving the objective function, with slightly longer computing time. Both analyses based on corridor speed contours and system total travel time show that the optimized operation strategies could effectively mitigate traffic congestion during the incident. This research considers the metering rate of a ramp meter system in the context of a regional network, which significantly differs from previous studies that focused exclusively on local traffic patterns. This study provides a new perspective for traffic operators and management agencies.

The proposed framework is flexible to deal with both discrete and continuous control variables. The proposed sequential method could effectively reduce the search space in the temporal dimension, and the case study demonstrates that it is effective in application. The optimization framework developed in this study could be extended to consider other control strategies such as VMS and road pricing.

Current study only considered fixed, but time-dependent OD matrices. Previous studies by the authors Xiong et al. (2015) and Zhang et al. (2013) have considered integrating traffic simulation models with agent-based demand models to consider other travel choice dimensions such as the adjustment of departure time. Therefore, the proposed optimization framework could be extended to consider the elastic time-dependent travel demand, which could be important when strategies such as dynamic road pricing are considered. The optimization framework developed in the current study provides an important tool to address these future research needs.

CHAPTER FIVE OPTIMIZATION OF MULTIVARIATE TRAFFIC OPERATION STRATEGY

Introduction

Chapter Three demonstrated the feasibility of the proposed framework on a small analytical network, while Chapter Four extended the proposed optimization framework to consider optimal traffic operation strategies on a large scale time-dependent simulation network. Two control strategies (ramp meters and shoulder lanes) were considered and neither of them involved complex behavioral reactions to control strategies. In contrast, strategies such as Variable Message Signs (VMS) may cause more complicated behavioral reactions. For example, when travelers see VMS, some of them may change routes. Although we may monitor the rerouting rate by observing the portion of traffic that exits the freeway after seeing the VMS, there is no empirical studies in the literature to show who will and who will not switch route. And if they do switch routes, it is still a question which alternative route each driver exit the freeway as a result of VMS will take. A reasonable assumption could be that these travelers will take the shortest path (either travel time or broadly defined if more data become available) from the freeway exits to their final destination. Without further information, we may also assume that all potential users of the freeway segment affected by the incident are equally likely to respond to VMS and switch routes. In this chapter, we will explore the performance of the proposed simulation-based optimization framework in addressing more complicated control strategies such as VMS. As the first attempt in this direction, we will first adopt the assumptions presented above, while adopting more complicated behavioral assumptions in the future when more empirical data become available.

Methodology

Integrated Framework

This chapter will apply the same integrated optimization framework which was presented in Chapter Three by substituting the analytical assignment model with the mesoscopic traffic simulation model that was used in Chapter Four. This chapter will introduce more traffic operation controls into the simulation model and use the optimization framework to explore the optimal the best performing (i.e. minimum total travel time) incident response strategies on a large network when multiple traffic operation controls (Shoulder Lane, Ramp Meter, VMS) work together. The optimization framework for this chapter is shown in Figure 16.

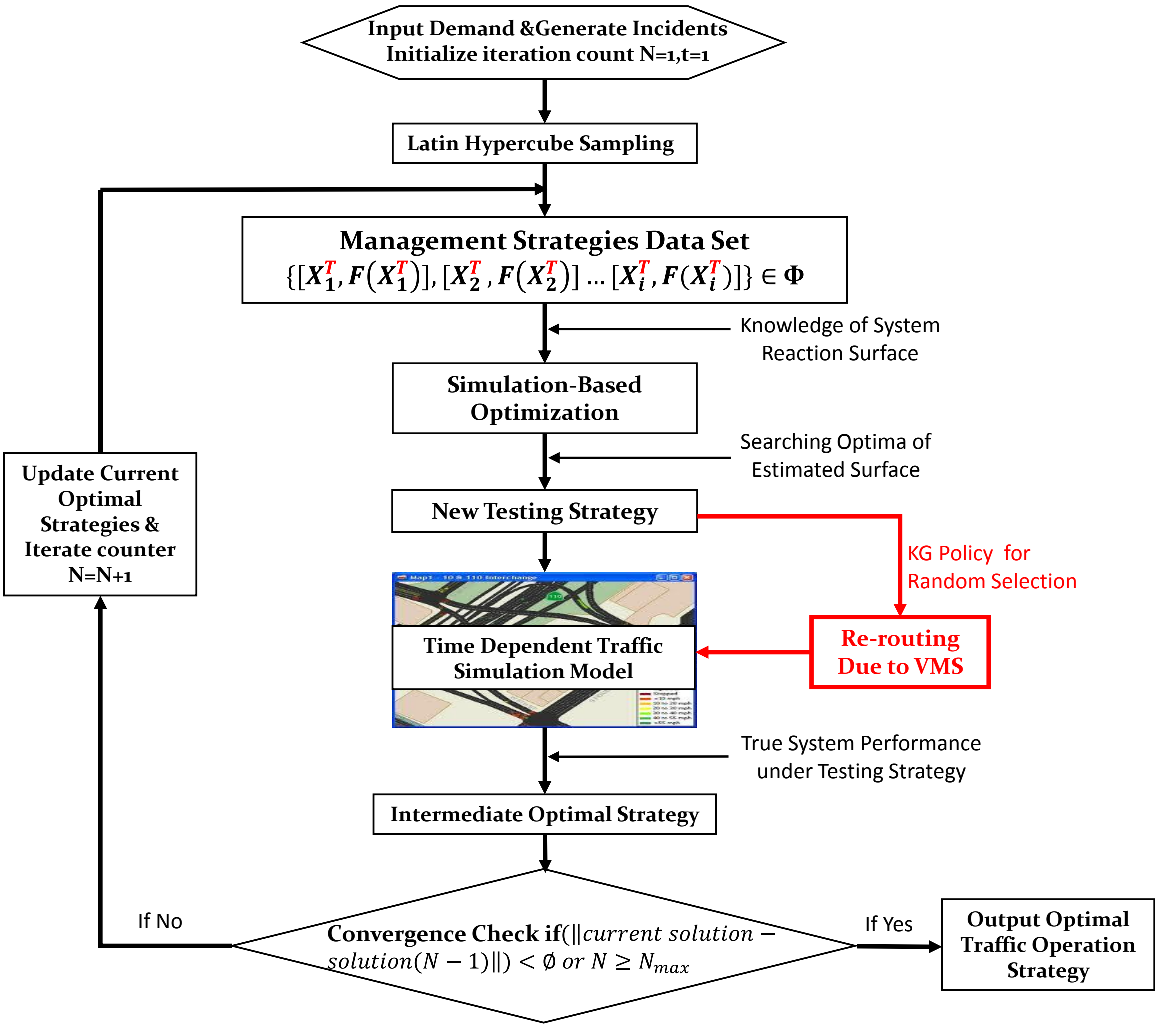

Figure 16. Framework of the Incident Response Optimization Model (Adapted from Chapter Four, New Components Highlighted in Red).

The objective of the optimization problem in this study can also be generally formulated as Equation (12). We first use Latin Hypercube Sampling (LHS) method to generate training data points and store them into data set Φ to start off the optimization procedure. Each data point represents a single traffic operation strategy design and its corresponding true function evaluation value that output by running the simulation model. LHS is a random sampling method that uniformly draws sample data points from an equally stratified multi-dimensional search space without overlapping samples (McKay et al., 1979).

For a given data set Φ,

$$\{(X_1^T, \mathrm{F}(X_1^T)),\ (X_2^T, \mathrm{F}(X_2^T)),\ (X_3^T, \mathrm{F}(X_3^T)), \ldots\ldots (X_\mathrm{i}^T, \mathrm{F}(X_\mathrm{i}^T))\} \in \Phi, \tag{13}$$

Where,

X_i^T represents a multi-dimensional decision vector of all operation strategies within study period *T*.

$\mathrm{F}(X_\mathrm{i}^T)$ is the value of the performance evaluation value (e.g. Total Travel Time) under traffic operation strategy X_i^T, *i* is the index of *i*th traffic operation combinational strategy evaluated by running the simulation model.

BO Algorithm will then approximate the posterior distribution of the unknown objective function by using the sampled data. In each iteration, the approximated objective function will be optimized by using the embedded simulated annealing method. The optimal solution at current iteration is the most promising strategy for next simulation evaluation. Hence, this optimal solution will then be tested in the simulation and its true performance will be evaluated. The true performance may be different from the estimated performance, and will help researchers to better estimate the response surface in the next iteration by interpolating this tested strategy and the simulated performance value into the existing data set Φ as a new data point. This iterative process will continue until: 1) the number of iterations reaches the maximum number of iterations$\boldsymbol{N_{max}}$, or 2) the improvement at the current iteration is less than a predetermined stopping criterion $\emptyset$.

Design of Experiment

In this section, we applied the proposed optimization framework on the same study area as we did in Chapter Four. The study area and simulation network that developed for this section is shown in Figure 9. The simulation network used in this section is also based on the current simulation network that we showed in the Chapter Four.

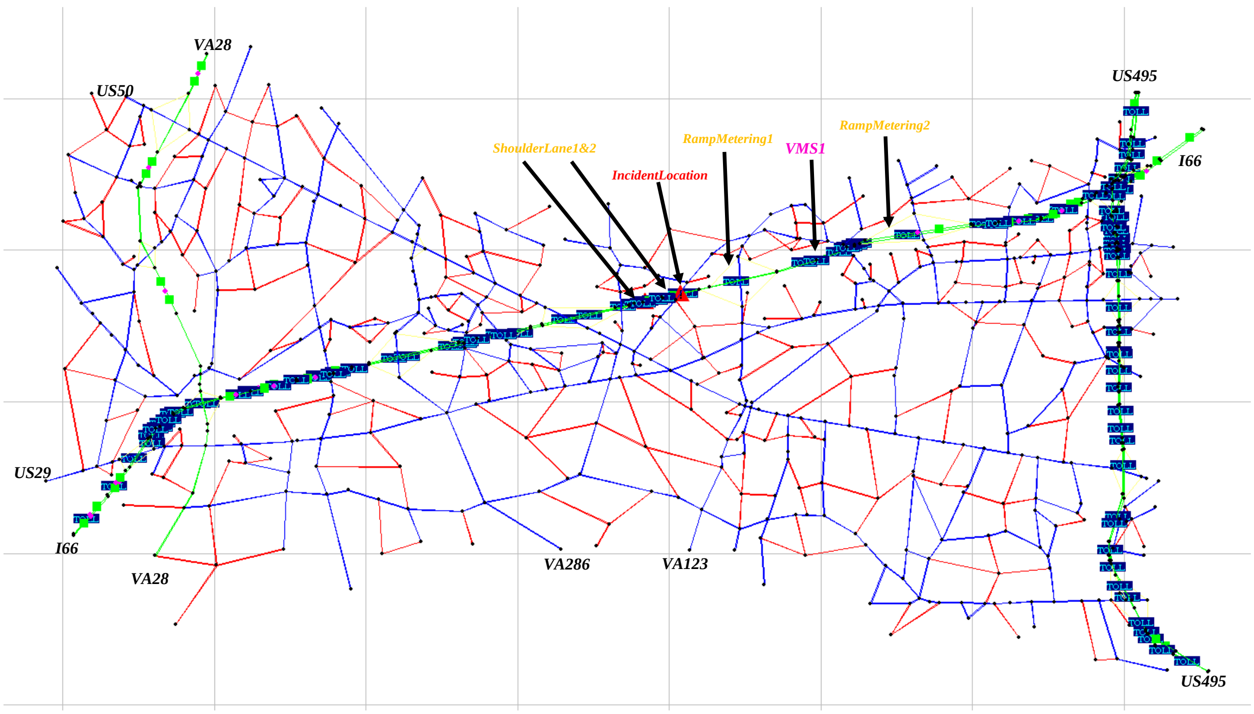

Figure 17. Simulation Network in DTALite with Multiple Traffic Operation Controls

This section demonstrates the proposed optimization framework that integrates the simulation model with BO algorithm through a case study based on the calibrated model for the study area. As Figure 17 shows, an incident closes one general purpose lane on I-66WB at 16:30. To mitigate this non-recurrent congestion, we consider strategies that include opening of a shoulder lane at the incident location and the freeway segment immediately downstream of the incident location, 2 ramp meters at on-ramps upstream of the incident location, and 1 Variable Message Sign (VMS) at the middle of 2 ramp meters. Shoulder lane will increase the capacity around incident location where the capacity drops due to incident. Ramp meters control the traffic flows entering freeway segment to further control the demand. VMS indicates the downstream traffic condition to passing-by drivers and potentially makes some percentage of the drivers to detour. The decision variables for should lanes are discrete, while the ones for ramp meters and VMS could be continuous. Both decisions include a series of decision variables for the following 30-minute time period following the incident. These combinations (multiple type of strategies, locations, and time periods) lead to a large number of possible scenarios, and complex interactions with traffic dynamics. An efficient algorithm to identify the optimal solution among these scenarios is needed (especially when VMS detouring is introduced), but has not been well studied in the literature. Previous studies on ramp meter algorithms rely on reading of local traffic volumes and focus on analytically maintaining smooth operation of local freeway traffic flows. There are few studies focus on optimizing VMS compliance rates. It is challenging to obtain the re-routing percentage rate that is due to the incident. Let alone when multiple traffic operation strategies work cooperatively, it will make the problem even more challenging. Thus, this research is looking for a method to address these challenges we have in the current study.

Since we choose 15 minutes as the time resolution for OD matrices, we also assume the metering rate and compliance percentage rate for ramp meters and VMS will only be adjusted once every 15 minutes for consistency. However, this choice is not a constraint and can be changed without loss of generality. We consider the adoption of response strategies after the incident occurred at 16:30. Since we take VMS into the consideration, we have to randomly select a portion of the through vehicles that used to take incident link to switch to a new path from their original routes. This requires us to conduct multiple random selection for a single VMS rate. The same compliance percentage rate could be due to different set of vehicles that used to drive through incident location. This will be leading to different network performances (e.g. total travel time). Therefore, multiple drawn and simulation run are required for each VMS compliance percentage rate. This will make our model need more computation time. For simplicity, we only active the adoption of response strategies for two consecutive 15 minute time intervals after the incident occurred at 16:30. We will develop various scenarios (e.g. ramp meter only, VMS only, shoulder lane only, ramp meter + VMS, ramp meter + VMS + shoulder lane) that apply different set of traffic operation strategies and then optimize those scenarios by using the proposed optimization model. We will investigate and compare the performance of the proposed optimization model under each scenario. Therefore, there are totally 2 decision variables for ramp metering 1 and another 2 for ramp metering 2, one for each 15-minute period during the remaining of the simulation period. Since we only implement one VMS in the model, thus there are 2 decision variables for VMS in the two 15-minute time intervals. A dummy variable will represent the decision of whether or not to open the shoulder lane each of the two locations and for each time period. Therefore, there are totally 6 continuous and 4 0/1 decision variables to be determined. Therefore, for the scenario that activates all adopted traffic operation strategies, we uniformly sample (6+4)*10=100 sample points (Jones et al., 1998) within the searching region using LHS to start up the optimization process. The number of initial sample points to start up the optimization by using LHS is determined by the number of active decision variables, the basic properties of different scenarios are shown in Table 4. Scenarios that include with VMS will be randomly drawing 20 times for a certain VMS percentage rate of detouring vehicles, and therefore we have to run 20 simulation evaluations for each sample data point.

Table 4. Properties of Different Scenarios for Experimental Design

Scenario Name	Time Duration	Number of Decision Variables	Number of Starting Sample Points
No Incident	N/A	N/A	N/A
Incident	16:30-17:00	N/A	N/A
Incident + Ramp Meter	16:30-17:00	4	40
Incident + VMS	16:30-17:00	2	20
Incident + Shoulder Lane	16:30-17:00	4	4
Incident + RampMeter + VMS	16:30-17:00	6	60
Incident + RampMeter + VMS + ShoulerLane	16:30-17:00	10	100

Integrated Optimization Model

In this research, we developed 7 different scenarios shown in Table 4. These traffic operation strategies will either working individually or cooperatively. The performance on congestion mitigation by these scenarios will be investigated and compared with Incident scenario and No Incident scenario. After setting up incident and candidate traffic operation strategies in simulation network, we integrated the simulation model with BO algorithm following the proposed optimization framework. For scenario with ramp meter only, the optimal traffic operation strategy with ramp metering rate of [5, 95] and [36, 95] for ramp 1 and ramp 2 respectively. The number here represents the corresponding capacity reduction rate at each metering location and for the corresponding time period. The optimal network total travel time $f(X) = 4.3338072402 * 10^6$ minutes. For scenario with VMS works individually, the optimal VMS detouring rate are [11%, 9%] which represents the compliance rate that vehicles detour from their original route at each time interval. Those vehicles used to drive through incident location and decided to switch the paths after seeing the VMS. Scenario with shoulder lane only is optimized very efficiently with opening shoulder lanes at both locations for all examined periods. The optimal traffic operation strategies include ramp meter and VMS have the metering and VMS rate of [5, 95, 9%] and [36, 95, 7%] for ramp 1, ramp 2 and VMS respectively. The optimal network total travel time is $f(X) = 4.33057405918 * 10^6$ minutes. For the scenario that all candidate traffic operation strategies working together, the optimal traffic operation strategies include the opening shoulder lanes at both locations for all periods and ramp meter, VMS rate of [5, 95, 5%] and [5, 95, 5%] for ramp 1, ramp 2 and VMS respectively. The optimal network total travel time is $f(X) = 4.30374910591 * 10^6$ minutes.

The value of objective function (system total travel time) at each iteration is shown in Figure 18. It clearly shows that the optimization system based on BO algorithm converges and the total travel time of the entire system becomes flat. We can see the all traffic operational strategies have improvements on total travel time reduction. VMS has the least total travel time reduction and ramp meter works better individually than VMS only, shown as the gray line and brown line in Figure 18. Among those scenarios with the traffic operation strategy work individually, shoulder lane has the best congestion mitigation performance. Scenario with ramp meter and VMS work together has the mediocre performance on total travel time reduction. Among all the tested scenarios, it clearly shows that the scenario with all traffic operation strategies activated together has the best performance on total network travel time reduction, which is shown as green line in Figure 18. Since there are three scenarios with VMS implemented, which required us to run multiple times for each sample point being evaluated by simulation model. We ran 20 replications for each training data point during sampling stage. In this study, for the new testing data point that optimized by maximized acquisition function within BO algorithm, we adopted knowledge-gradient policy (KG policy) as the method to conduct the random selection. The common way to conduct KG policy is to keep increasing the number of replications by an increment as BO algorithm iterates. We set the initial number of replication run is 10, and increment goes by 1 at each BO iteration. The reason we adopted KG policy is that this method is guaranteed convergence while remaining efficiency when dealing with random selection problems (Frazier et al., 2008). Figure 19Figure 19 compares the computational time of the integrated models with different scenarios. It indicates that scenarios with VMS will generally take great computing time, this is also the reason we choose to apply the traffic operation strategies only on 2 time intervals to keep the number of decision variables small. The number of decision variables will increase as we apply traffic operation strategies for more time intervals. The number of replications required will be increased, which will make the computing time increase exponentially.

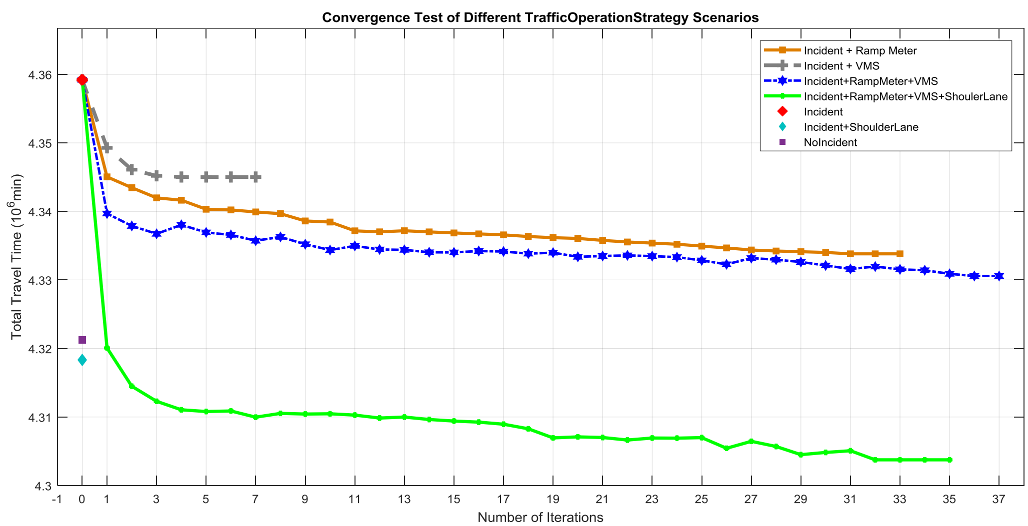

Figure 18. Convergence of System Travel Time

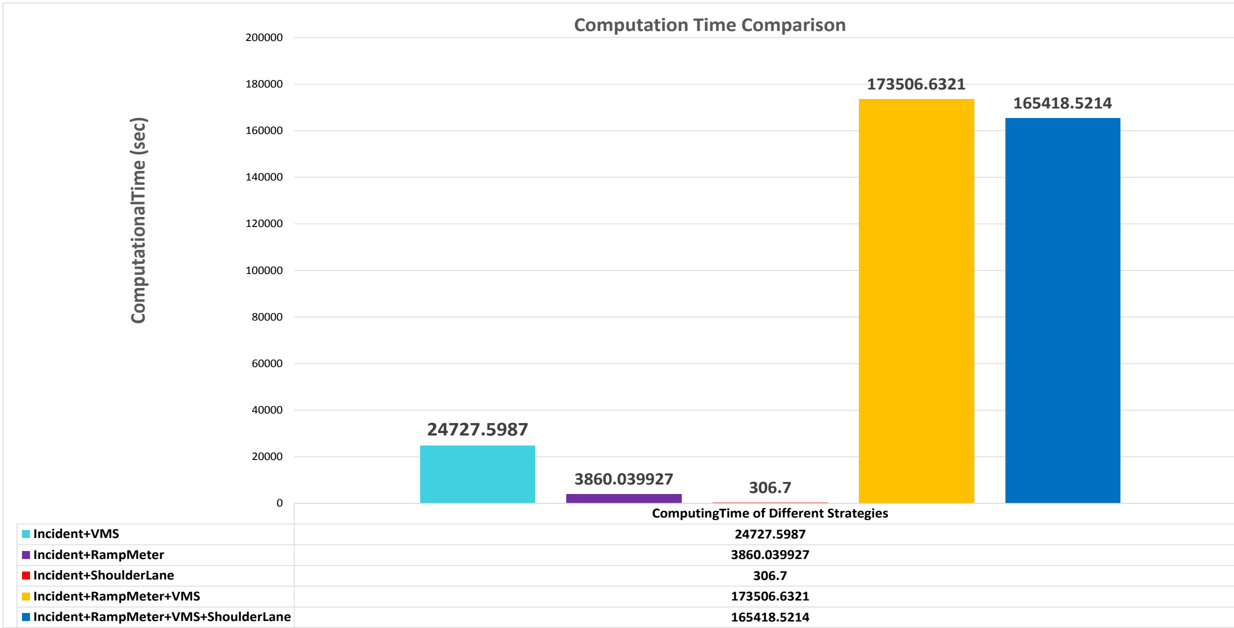

Figure 19. Computing Time Comparison

System Performance

Table 5 compares the system-wide performance between different scenarios: with the incident but no actions, activating the ramp meters, VMS and shoulder lanes individually, and activating shoulder lanes and ramp meters following the control strategies optimized using the proposed optimization model. As a comparison, we also report the total travel time under the scenario with neither accidents nor controls to show the relative magnitude of improvements under different scenarios.

The incident closed one lane on I-66WB. Therefore, if we open the shoulder lane immediately at the affected freeway segment, we restored the capacity and the overall performance of the system is similar to the scenarios with no accidents and no controls. However, if we combine the extra capacity provided by the shoulder lane with control strategies implemented through ramp meters and VMS, the system would perform much better. The system travel time savings could be improved by more than 41% (1.27% vs 0.9% travel time savings when compared to the incident only scenario) when meter controls were introduced.

Table 5. Comparison of System-wide Total Travel Time under Different Scenarios

Scenario Name	TotalTravelTime (*10^6min)	TotalTravelTime Saved (*10^6min)	%Improvement
No Incident	4.32127832	-0.037922698	-0.9%
Incident	4.359201018	N/A	N/A
Incident + Ramp Meter	4.3338072402	0.025393778	0.6%
Incident + VMS	4.34502013205	0.014180886	0.33%
Incident + Shoulder Lane	4.31832853951	0.040872478	0.9%
Incident + RampMeter + VMS	4.33057405918	0.028627	0.7%
Incident + RampMeter + VMS + ShoulerLane	4.30374910591	0.055451912	1.27%

In addition to the system wide performance, we also compare the corridor performance using the speed contour to visualize the impact of different control strategies under different examined scenarios from Table 5. Figure 20a shows the congestion pattern of I66WB corridor when no incident occurs. Figure 20b shows the congestion pattern when an incident occurs at 16:30pm. All freeway segments on I-66WB upstream of the incident location are very congested (the color shows average speed). Figure 20c to Figure 20e show the congestion patterns when the ramp meter, VMS and shoulder lane is activated alone. It indicates that VMS alone improves corridor traffic condition slightly, while ramp meter performances better than VMS works individually. Opening shoulder lanes yields the best improvements on corridor traffic congestion among scenarios where traffic control strategy operate alone. The improvement in Figure 20e improved the traffic flow and throughput at the incident location, the segments downstream become slightly more congested. Figure 20f shows that when both ramp meters and VMS are activated and operated according to the optimal strategies identified through the proposed framework. The congestion on freeway segments upstream of the incident location at examined time periods is clearly improved. As Figure 20g shows, the improvements under the scenario with all of the strategies implemented are the most significant. These figures need to be combined with system-wide performance analysis as shown in Table 5 to provide a comprehensive evaluation of the performance of different strategies. In all cases, the advantage of the proposed optimization framework is clearly illustrated.

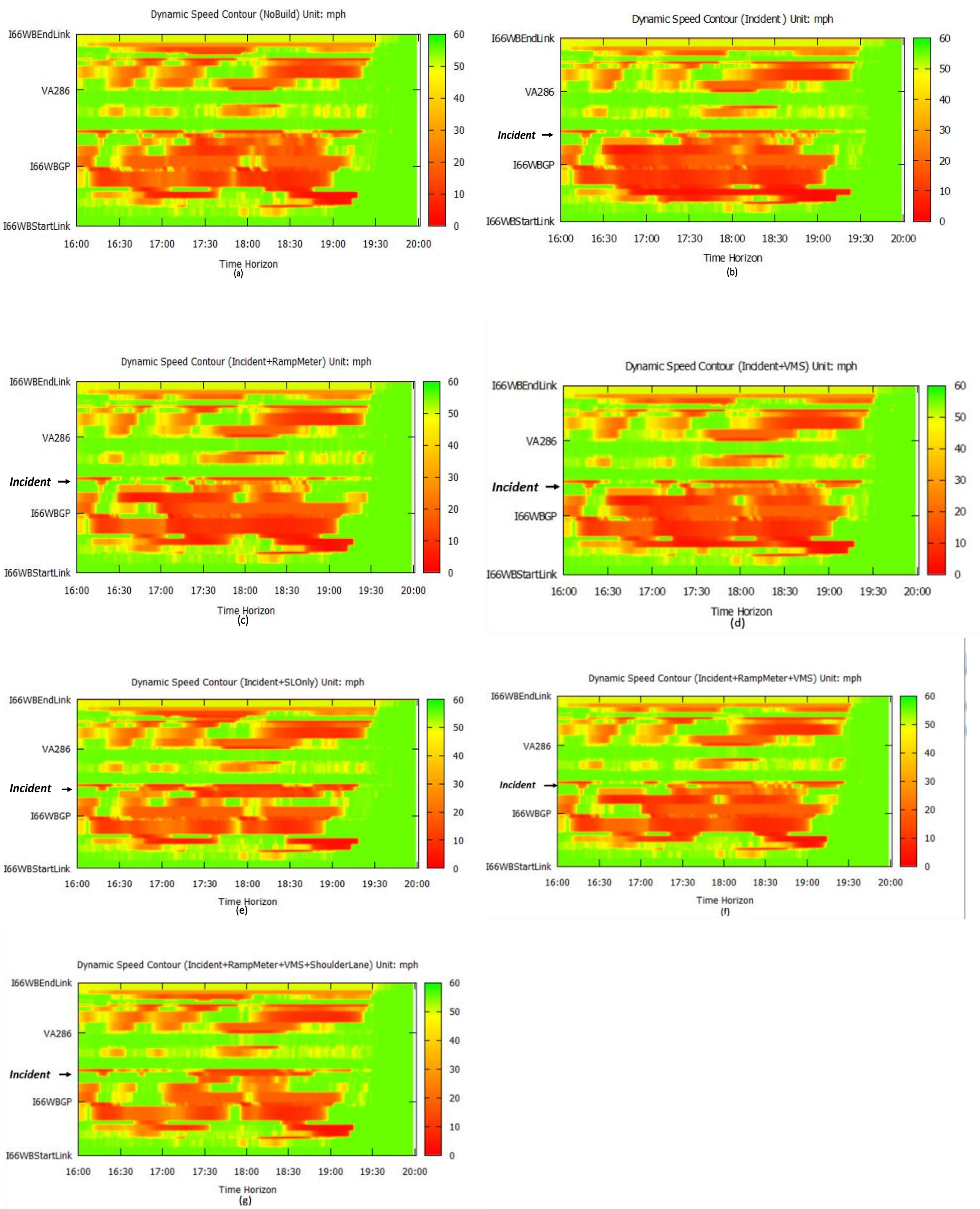

Figure 20. Corridor Travel Speed Heat-map under Different Incident and Control Scenarios

Conclusions

Traffic incident is one of the most important contributors for traffic congestion in urban areas. There is a wide range of traffic operation tools for traffic operators and engineers to choose to mitigate the incident related congestion. These traffic operation controls mostly working cooperatively, therefore, it is important to efficiently evaluate them and identify the best incident response strategies. However, given multiple choices of traffic operation tools, the possible choices of control parameters (discrete and continuous), the temporal resolution, and multiple deployment locations, the number of operation design scenarios to be considered will increase exponentially. Traffic simulation models are typically used to evaluate the performance and effectiveness of traffic operation controls because simulation models could better capture complex traffic dynamics compared to the analytical traffic flow models. However, simulation models usually need long-computing time combined with the large number of possible scenarios prevent it from widely applied in incident response strategy optimization studies. Analytical methods are normally used for obtaining the theoretical optimal solution relatively efficiently. Nevertheless, this approach has limitations in realistically capture the traffic dynamics and model to complex behaviors between traffic operation controls and traffic. Thus, there is few research applied analytical models to optimize the traffic operation management for better responding incident congestion. This study proposed an integrated optimization framework by combining a mesoscopic traffic simulation model with multiple traffic operation controls embedded and simulation-based optimization algorithm to address these challenges.

This study integrates three traffic operation strategies into a simulation-based optimization model. Previous research barely include such many of different traffic control choices all together. The application in a case study on a large network in Northern Virginia area demonstrates that the proposed framework could effectively identify the optimal traffic operation strategies with multiple strategy choices working cooperatively. Although it could not be implemented in real time in its current form, with more computing resources that are commercially available, it could be implemented in near-real time to address incident response problems in real time.

The case study shows that the scenario with all three traffic control strategies activated together outperforms all the other scenarios in improving the objective function significantly. However, it has to trade off by the longer computing time due to the VMS random selection procedure. Both analyses based on corridor speed contours and system total travel time show that the optimized operation strategies could effectively mitigate traffic congestion during the incident. This research considers the metering rate of a ramp meter system in the context of a regional network, which significantly differs from previous studies that focused exclusively on local traffic patterns. This study provides a new perspective for traffic operators and management agencies. The proposed framework is flexible to deal with both discrete and continuous control variables. Current study only considered fixed, but time-dependent OD matrices. The proposed optimization framework could also be extended to consider the elastic time-dependent travel demand, which could be important when strategy such as dynamic road pricing is considered. The optimization framework developed in the current study provides an important tool to address these future research needs.

CHAPTER SIX CONCLUSIONS

Traffic congestion keeps deteriorating worldwide. It not only compromises the societal productivity due to wasted time, but also cause other issues such as pollutions, energy efficiency, safety, and even psychological problems. Traffic incident is one of the major factors for traffic congestions. Due to the budget limitation and political constraints, there is not much room for new roadways. Traffic operation strategies became an important alternative for congestion relief, as demonstrated by nation-wide deployment of managed lanes and active traffic management systems. New management tools (e.g. Shoulder Lane, Ramp Meter, Variable Message Signs, Speed Limit Signs, HOV/HOT, etc.) can be used to mitigate traffic congestion, particularly during traffic incidents. A simulation model is typically used to test those traffic operation controls before an actual action is implemented in real world because field deployment is usually expensive and non-reversible. Previous research had made huge efforts on evaluating traffic operation strategies for congestion mitigation during incidents. Simulation models are able to capturing the complex traffic dynamics, yet have limitation on searching for the optimal traffic operation strategies when multiple traffic operation controls are being deployed (e.g. multiple deployment locations, time-dependent network, multiple operation controls). Analytical methods are always limited by their capability on reflecting real world traffic phenomena. This dissertation developed an integrated optimization model which combines Simulation-Based Optimization method with large-scale time-dependent simulation models. This study showed that the proposed method can efficiently identify the system optimal through guided search supported by an estimated response surface.

There are many methods for estimating such a response surface. This research tested three commonly used Surrogate Models (QPR, RBF, BO), each of which has been integrated into the optimization framework and tested on an analytical model. The performance of these Surrogate Models were compared with the theoretical optima which was identified through exhaustive search. Based on the numerical results, QPR model requires less computing time among all three Surrogate Models. However, it has poor performance on identifying the optimal solution. Thus, QPR may be utilized in situations when a quick solution is needed for incident response. BO has the best optima searching capability yet needs longer computational time. RBF stands in the middle for both computational efficiency and optima searching capability.

This dissertation then integrated the optimization framework with a large scale time-dependent simulation model. A case study was conducted on a regional network in Northern Virginia area. Two types of traffic operation controls (Shoulder Lane and Ramp Meter) were implemented into a mesoscopic simulation model at multiple locations along I-66WB corridor, which led to a high-dimensional optimization problem with 40 decision variables. The proposed sequential method could effectively reduce the searching space in the temporal dimension, and the case study demonstrated that it was effective in application. The optimization results showed that BO algorithm outperformed QPR and

RBF methods in searching for optimal solutions, with required slightly longer computing time. Based on the results from both system wide travel time analysis and corridor speed contour analysis, optimized operation strategies identified by the proposed framework could effectively mitigate traffic congestion during the incident. This study represents a major improvement compared to previous studies in the number of different traffic control variables considered.

This dissertation then introduced one more traffic operation strategy, VMS, which involved more complicated behavioral reactions and required more sophisticated treatment in traffic simulation into the proposed integrated optimization framework. The optimization results on the system wide travel time analysis and corridor speed contour analysis of the case study both indicated that the scenario with all three traffic control strategies activated together outperformed the scenarios where each control strategies were invoked in isolation. This result showed the importance of considering all control strategies in an integrated optimization framework and being considered simultaneously. This study also showed the importance and the complication of considering different travel behavior, such as compliance rate to VMS, choice of alternative path, and randomness in who may comply with the VMS control. Findings in these areas may guide further studies in active traffic management and incident responses.

This study provided a new tool for traffic operators and management agencies who need to react quickly in real world traffic operations. The proposed optimization approach will provide researchers and decision makers better insights in incident response management and traffic operations in general. Travel time savings from better traffic incident management strategies will significant improve system efficiency, environment, and safety.

REFERENCES

AASHTO - National Traffic Incident Management Coalition - Home. (2016). Retrieved from http://ntimc.transportation.org/

Adeli, H., & Karim, A. (2000). Fuzzy-wavelet RBFNN model for freeway incident detection. Journal of Transportation Engineering, 126(6), 464-471.

Amaran, S., Sahinidis, N. V., Sharda, B., & Bury, S. J. (2014). Simulation optimization: a review of algorithms and applications. 4OR, 12(4), 301-333.

Ammeri, A., Chabchoub, H., Hachicha, W., & Masmoudi, F. (2010). A Comprehensive literature classification of simulation–optimization methods. In The 9th International Conference on Multiple Objective Programming and Goal Programming.

Bar-Gera, H. (2016). Transportation Network Test Problems. http://www.bgu.ac.il/~bargera/tntp/

Barton, R. R., & Meckesheimer, M. (2006). Chapter 18 Metamodel-Based Simulation Optimization. Handbooks in Operations Research and Management Science, 13(C), 535-574. https://doi.org/10.1016/S0927-0507(06)13018-2

Ben-Akiva, M., Cuneo, D., Hasan, M., Jha, M., & Yang, Q. (2003). Evaluation of freeway control using a microscopic simulation laboratory. Transportation

Research Part C: Emerging Technologies, 11(1), 29-50. doi: 10.1016/s0968-090x(02)00020-7

Bhattacharyya, G., & Johnson, R. (1997). Statistical Concepts and Methods (1st ed.). Wiley.

Björkman, M., & Holmström, K. (2000). Global optimization of costly nonconvex functions using radial basis functions. Optimization and Engineering, 1(4), 373-397.

Brennan, T. R.A. Gurriell, A.J. Bechtel and M.M. Venigalla. (2019). Visualizing and Evaluating Interdependent Regional Traffic Congestion and System Resiliency, a Case Study Using Big Data from Probe Vehicles. Journal of Big Data Analytics in Transportation. Springer. http://doi.org/10.1007/s42421-019-00002-y

Brennan, T.M., M.M. Venigalla, A. Hyde, and A. LaRegina. (2018). Performance Measures for Characterizing Regional Congestion Using Aggregated Multiyear Probe Vehicle Data Transportation Research Record: Journal of the Transportation Research Board. http://journals.sagepub.com/doi/10.1177/0361198118797190

Brennan, T., and M.M. Venigalla. A Constructability Assessment Method (CAM) for Sustainable Division of Land Parcels. *Land Use Policy*. Vol. 56, November 2016, pp. 47-57. http://dx.doi.org/10.1016/j.landusepol.2016.04.031.

Carson, Y., & Maria, A. (1997). Simulation optimization: methods and applications. In Proceedings of the 29th conference on Winter simulation (pp. 118-126). IEEE Computer Society.

Carlson, R., Papamichail, I., Papageorgiou, M., & Messmer, A. (2010). Optimal mainstream traffic flow control of large-scale motorway networks. Transportation Research Part C: Emerging Technologies, 18(2), 193-212. doi: 10.1016/j.trc.2009.05.014

Carlson, R., Papamichail, I., Papageorgiou, M., & Messmer, A. (2010). Optimal Motorway Traffic Flow Control Involving Variable Speed Limits and Ramp Metering. Transportation Science, 44(2), 238-253. doi: 10.1287/trsc.1090.0314

Chauhan, A. S. (2003). Development and evaluation of diversion strategies under incident response using dynamic traffic assignment system. Doctoral dissertation, Northeastern University.

Chen, X. M., Zhang, L., He, X., Xiong, C., & Li, Z. (2014). Surrogate-Based Optimization of Expensive ‐ to ‐ Evaluate Objective for Optimal Highway Toll Charges in Transportation Network. Computer ‐ Aided Civil and Infrastructure Engineering, 29(5), 359-381.

Chen, X., Zhang, L., He, X., Xiong, C., & Zhu, Z. (2017). Simulation-based pricing optimization for improving network-wide travel time reliability. Transportmetrica A: Transport Science, 14(1-2), 155-176. doi: 10.1080/23249935.2017.1379038

Chen, O., Hotz, A., & Ben-Akiva, M. (1997). Development and Evaluation of a Dynamic Ramp Metering Control Model. IFAC Proceedings Volumes, 30(8), 1089-1095. doi: 10.1016/s1474-6670(17)43966-8

Chong, L., & Osorio, C. (2018). A Simulation-Based Optimization Algorithm for Dynamic Large-Scale Urban Transportation Problems. Transportation Science, 52(3), 637-656. doi: 10.1287/trsc.2016.0717

Chou, C. S., & Miller-Hooks, E. (2011). Exploiting the capacity of managed lanes in diverting traffic around an incident. Transportation Research Record: Journal of the Transportation Research Board, 2229, 75-84.

Christiansen, S., Patriksson, M. & Wynter, L. (2001). Struct Multidisc Optim. *21: 361*. *https://doi.org/10.1007/s001580100115*

Chu, L., & Yang, X. (2003). Optimization of the Alinea ramp-metering control using genetic algorithm with micro-simulation. TRB Annual Meeting, (p. 6).

Chu, L., Liu, H., & Recker, W. (2004). Using microscopic simulation to evaluate potential intelligent transportation system strategies under nonrecurrent congestion. Transportation Research Record: Journal of the Transportation Research Board, 1886, 76-84.

Ciuffo, B., Punzo, V., & Quaglietta, E. (2011). Kriging meta-modeling to verify traffic micro-simulation calibration methods. In 90th TRB Annual Meeting. TRANSPORTATION RESEARCH BOARD.

Clark, J., Neuner, M., Sethi, S., Bauer, J., Bedsole, L., & Cheema, A. (2017). Transportation Systems Management and Operations in Action. Retrieved from https://ops.fhwa.dot.gov/publications/fhwahop17025/index.htm#ch1

Dion, F., Sivakumaran, K., & Ban, X. (2012). Evaluation of traffic simulation model use in the development of Corridor System Management Plans (CSMPs).

Fellendorf, M., & Vortisch, P. (2010). Microscopic Traffic Flow Simulator VISSIM. In In Fundamentals of Traffic Simulation (pp. 63-93). New York: Springer.

Frazier, P., Powell, W., & Dayanik, S. (2008). A Knowledge-Gradient Policy for Sequential Information Collection. SIAM Journal On Control And Optimization, 47(5), 2410-2439. doi: 10.1137/070693424

Fu, M. C. (1994). Optimization via simulation: A review. Annals of Operations Research, 53(1), 199-247.

Fu, M. C. (2002). Optimization for simulation: Theory vs. practice. *INFORMS Journal on Computing, 14(3), 192-215.*

Fu, M. C., & Howell, W. C. (2003). Application of perturbation analysis to traffic light signal timing. 42nd IEEE Conference. 5, pp. 4837-4840. IEEE.

Gelbart, M. A. (2015). Constrained Bayesian Optimization and Applications. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences. Retrieved from http://nrs.harvard.edu/urn-3:HUL.InstRepos:17467236

Gutmann, H. M. (2001). A radial basis function method for global optimization. Journal of Global Optimization, 19(3), 201-227.

Hajibabai, L., & Saha, D. (2018). Patrol Route Planning for Incident Response Vehicles under Dispatching Station Scenarios. Computer-Aided Civil And Infrastructure Engineering, 34(1), 58-70. doi: 10.1111/mice.12384

Hasan, M. (1999). *Evaluation of ramp control algorithms using a microscopic traffic simulation laboratory, MITSIM*.

He, X. (2014). Simulation-based optimization of transportation systems: Theory, surrogate models, and applications. Retrieved from http://hdl.handle.net/1903/16065

Henchey, M. J., Batta, R., Blatt, A., Flanigan, M., & Majka, K. (2014). A simulation approach to study emergency response. Journal of Simulation, 8(2), 115-128.

Huaguo, Z. (2008). Evaluation of route diversion strategies using computer simulation. Journal of Transportation Systems Engineering and Information Technology, 8(1), 61-67.

Hegyi, A., De Schutter, B., & Hellendoorn, H. (2005). Model predictive control for optimal coordination of ramp metering and variable speed limits. Transportation Research Part C: Emerging Technologies, 13(3), 185-209. doi: 10.1016/j.trc.2004.08.001

INRIXINRIX. (2017). Los Angeles Tops INRIX Global Congestion Ranking - Retrieved from http://inrix.com/press-releases/scorecard-2017/

Jiang, R., & Chung, E. (2014). A Ramp Metering Strategy for Rapid Congestion Recovery. Computer-Aided Civil And Infrastructure Engineering, 30(3), 202-216. doi: 10.1111/mice.12101

Jonas Mockus, Vytautas Tiesis, & Zilinskas., a. A. (1978). The application of Bayesian methods for seeking the extremum. Towards Global Optimization.

Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. Journal of Global optimization, 13(4), 455-492.

Kaviti, S., M.M. Venigalla, S. Zhu, K. Lucas, and S. Brodie. (2018). Impact of Pricing and Transit Disruptions on Bikeshare Ridership and Revenue. Transportation. Springer Publications. http://doi.org/10.1007/s11116-018-9904-5.

Kleijnen JPC (2008) Response surface methodology for constrained simulation optimization: An overview. Simulation Model. Practice and Theory 16(1):50–64.

Kotsialos, A., Papageorgiou, M., & Middelham, F. (2001). Optimal Coordinated Ramp Metering with Advanced Motorway Optimal Control. Transportation Research Record: Journal Of The Transportation Research Board, 1748(1), 55-65. doi: 10.3141/1748-07

Kotsialos, A., Papageorgiou, M., Mangeas, M., & Haj-Salem, H. (2002). Coordinated and integrated control of motorway networks via non-linear optimal control. Transportation Research Part C: Emerging Technologies, 10(1), 65-84. doi: 10.1016/s0968-090x(01)00005-5

LeBlanc, L.J., Morlok, E.K., Pierskalla, W.P. (1975). An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research Vol. 9, pp. 309-318.*

Lee, M., Zhu, Z., Xiong, C., & Zhang, L. (2017). Dynamic traffic assignment integration with real-time ramp metering for large-scale network management. *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), 463-468. 2017.*

Levinson, D. (2003). Effectiveness of Variable Message Signs. SSRN Electronic Journal. doi: 10.2139/ssrn.1748577

Li, Z., Madanu, S., Zhou, B., Wang, Y., & Abbas, M. (2010). A heuristic approach for selecting highway investment alternatives. Computer-Aided Civil and Infrastructure Engineering, 25(6), 427-439.

Lu, 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. doi: 10.1016/j.trc.2013.05.006

Ma, S. (2016). A nonlinear bi-level programming approach for product portfolio management. *SpringerPlus, 5(1), 727. http://doi.org/10.1186/s40064-016-2421-0*

Margiotta, R.; H. Cohen; G. Elkins; A. Rathi; and M.M. Venigalla. 1994. Generic Vehicle Speed Model Based on Traffic Simulation: Development and Application. Proceedings of the 74th Annual Meeting of the Transportation Research Board. 1994. 10p.

McKay, M.D., Conover, W.J. and Beckman, R.J. (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics 21: 239–245*.

Melouk, S. H., Keskin, B. B., Armbrester, C., & Anderson, M. (2011). A simulation optimization-based decision support tool for mitigating traffic congestion. Journal of the Operational Research Society, 62(11), 1971-1982.

Nitsche, P., Olstam, J., Taylor, N., Reinthaler, M., Ponweiser, W., Bernhardsson, V., & van Dam, E. (2016). Pro-Active Management of Traffic Incidents using Novel Technologies. Transportation Research Procedia, 14, 3360-3369.

Osorio, C., & Bierlaire, M. (2013). A simulation-based optimization framework for urban transportation problems. Operations Research, 61(6), 1333-1345.

Osorio, C., & Chong, L. (2015). A Computationally Efficient Simulation-Based Optimization Algorithm for Large-Scale Urban Transportation Problems. Transportation Science, 49(3), 623-636. doi: 10.1287/trsc.2014.0550

Osorio, C., & Nanduri, K. (2015). Energy-Efficient Urban Traffic Management: A Microscopic Simulation-Based Approach. Transportation Science, 49(3), 637-651. doi: 10.1287/trsc.2014.0554

Owens, N. (2010). *Traffic Incident Management Handbook. U.S. Department of Transportation Federal Highway Administration*.

Ozbay, K. M., Xiao, W., Jaiswal, G., Bartin, B., Kachroo, P., & Baykal-Gursoy, M. (2009). Evaluation of incident management strategies and technologies using an integrated traffic/incident management simulation. World Review of Intermodal Transportation Research, 2(2-3), 155-186.

Pal, R., & Sinha, K. (2002). Simulation Model for Evaluating and Improving Effectiveness of Freeway Service Patrol Programs. Journal Of Transportation Engineering, 128(4), 355-365. doi: 10.1061/(asce)0733-947x(2002)128:4(355)

Pal, R., & Sinha, K. (2000). Optimal design of freeway incident response systems. West Lafayette, Ind.: Purdue University [Joint Transportation Research Program].

Papayannoulis V, Marsico M, Maguire T, Strasser J, Scalici S (2011) An integrated travel demand, mesoscopic and microscopic modeling platform to assess traffic operations for Manhattan, New York. Proc. Transportation Res. Board (TRB) Conf. (TRB, Washington, DC).

PIZANO, C. O. (2010). Mitigating Network Congestion: Analytical Models, Optimization Methods and their Applications. Doctoral dissertation, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE.

Pulugurtha, S., Nambisan, S., Dangeti, M., & Kaseko, M. (2002). Simulating and analyzing incidents using CORSIM and VISSIM traffic simulation software. In 7th International conference on: applications of advanced technology in transportation.

Rasmussen, C., & Williams, C. (2006). Gaussian processes for machine learning. Cambridge, Mass.: MIT Press.

Regis, R. G., & Shoemaker, C. A. (2005). Constrained global optimization of expensive black box functions using radial basis functions. *Journal of Global optimization, 31(1), 153-171*.

Rios, L. M., & Sahinidis, N. V. (2013). Derivative-free optimization: a review of algorithms and comparison of software implementations. Journal of Global Optimization, 56(3), 1247-1293.

Schrank, D., Lomax, T., and Eisele, B. (2011). *TTI's 2011 Urban Mobility Report. Texas A&M Transportation Institute*

Schrank, D., Elisele, B., Lomax, T., & Bak, J. (2015). *TTI's 2015 Urban Mobility Report. Texas A&M Transportation Institute*.

Sisiopiku, V. P. (2007). Application of traffic simulation modeling for improved emergency preparedness planning. Journal of urban planning and development, 133(1), 51-60.

Skabardonis, A. V. (2003). *Measuring recurrent and nonrecurrent traffic congestion. Transportation Research Record: Journal of the Transportation Research Board, 118-124.*

Søndergaard, J. (2003). Optimization using surrogate models—By the space mapping technique. *Ph.D. thesis, Technical University of Denmark, Lyngby, Denmark.*

Southworth F., B. Janson, and M.M. Venigalla. (1992) DYMOD: Towards Real Time, Dynamic Traffic Routing During Mass Evacuations. Proceedings of Simulation Multi-Conference, Orlando, April 1992.

Stevanovic, J., Stevanovic, A., Martin, P. T., & Bauer, T. (2008). Stochastic optimization of traffic control and transit priority settings in VISSIM. Transportation Research Part C: Emerging Technologies, 16(3), 332-349.

Tao, X., & Schonfeld, P. (2005). Lagrangian relaxation heuristic for selecting interdependent transportation projects under cost uncertainty. Transportation Research Record: Journal of the Transportation Research Board, 1931, 74-80.

Tasic, I. (2012). Using microsimulation to evaluate traffic incident responses for Traffic Operations Center decision making. Master Thesis, The University of Utah.

Teklu, F., Sumalee, A., & Watling, D. (2007). A genetic algorithm approach for optimizing traffic control signals considering routing. Computer ‐ Aided Civil and Infrastructure Engineering, 22(1), 31-43.

Ting, C. J., & Schonfeld, P. (1998). Optimization through simulation of waterway transportation investments. Transportation Research Record: Journal of the Transportation Research Board, 1620, 11-16.

Vadde, R., Sun, D., Sai, J., Faruqi, M., & Leelani, P. (2012). *A simulation study of using active traffic management strategies on congested freeways. Journal Of Modern Transportation, 20(3), 178-184. doi: 10.1007/bf03325796*

Venigalla, M. M. (1996). A network assignment based approach to modeling mobile source emissions. *Transportation Research Part A, 1*(30), 56.

Venigalla, M. M. (2004, November). Household travel survey data fusion issues. In Resource Paper, National Household Travel Survey Conference: Understanding Our Nation's Travel (Vol. 1, No. 2).

Venigalla, M.M., X. Zhou, and S. Zhu. The Psychology of Route Choice in Familiar Networks: Minimizing Turns and Embracing Signals. *Journal of Urban Planning and Development*, Vol. 142(3), September 2016, pp. 1-14. http://dx.doi.org/10.1061/(ASCE)UP.1943-5444.0000364

Wang, S. L., & Schonfeld, P. (2005). Scheduling interdependent waterway projects through simulation and genetic optimization. Journal of waterway, port, coastal, and ocean engineering, 131(3), 89-97.

Wirtz, J., Schofer, J., & Schulz, D. (2005). Using simulation to test traffic incident management strategies: The benefits of preplanning. Transportation Research Record: Journal of the Transportation Research Board, 1923, 82-90.

Wright, G. B. (2003). Radial basis function interpolation: numerical and analytical developments. *Boulder Boulder, CO: University of Colorado.*

Xiong, C., Pan, Y., Lee, M., Zhu, Z., & Zhang, L. (2017). Integrated Agent-Based Travel Behavior and Dynamic Traffic Microsimulation for Ramp-Metering Analysis. Transportation Research Record: Journal Of The Transportation Research Board, 2665(1), 11-20. doi: 10.3141/2665-02

Xiong, C., Zhu, Z., He, X., Chen, X., Zhu, S., & Mahapatra, S. et al. (2015). Developing a 24-Hour Large-Scale Microscopic Traffic Simulation Model for the Before-and-After Study of a New Tolled Freeway in the Washington, DC–Baltimore Region. Journal Of Transportation Engineering, 141(6), 05015001. doi: 10.1061/(asce)te.1943-5436.0000767

Xu, J., Huang, E., Chen, C. H., & Lee, L. H. (2015). Simulation optimization: a review and exploration in the new era of cloud computing and big data. Asia-Pacific Journal of Operational Research, 32(03), 1550019.

Zhao, J., Ma, W., & Xu, H. (2017). Increasing the Capacity of the Intersection Downstream of the Freeway Off-Ramp Using Presignals. Computer-Aided Civil And Infrastructure Engineering, 32(8), 674-690. doi: 10.1111/mice.12281

Zhang, H., Jayakrishnan, R., & Recker, W. (2001). An Optimization Algorithm for Freeway Traffic Control. In In Proceedings of 4th IEEE Conference on Intelligent Transportation Systems: 102-107.

Zhang, L., Chang, G., Zhu, S., Xiong, C., Du, L., & Mollanejad, M. et al. (2013). Integrating an Agent-Based Travel Behavior Model with Large-Scale

Microscopic Traffic Simulation for Corridor-Level and Subarea Transportation Operations and Planning Applications. Journal Of Urban Planning And Development, 139(2), 94-103. doi: 10.1061/(asce)up.1943-5444.0000139

Zhang, L., Chen, X., He, X., Xiong, C., & Zhu, Z. (2014). Bayesian stochastic kriging metamodel for active traffic management of corridors. In IIE Annual Conference (pp. 1790-1799).

Zhou, H. (2008). Evaluation of Route Diversion Strategies Using Computer Simulation. Journal Of Transportation Systems Engineering And Information Technology, 8(1), 61-67. doi: 10.1016/s1570-6672(08)60010-0

Zhu, S., Kim, W., & Chang, G. (2012). Design and Benefit–Cost Analysis of Deploying Freeway Incident Response Units. Transportation Research Record: Journal Of The Transportation Research Board, 2278(1), 104-114. doi: 10.3141/2278-12

Zhu, S., Kim, W., Chang, G., & Rochon, S. (2014). Design and Evaluation of Operational Strategies for Deploying Emergency Response Teams: Dispatching or Patrolling. Journal Of Transportation Engineering, 140(6), 04014021. doi: 10.1061/(asce)te.1943-5436.0000670

Zhou, X., M.M. Venigalla, and S. Zhu. (2017). A Bounding Box Approach to Network Pruning for Efficient Path Search Through Large Networks. *Journal of Computing in Civil Engineering*. Vol. 31(5), Sept. 2017. ASCE. http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000675.

BIOGRAPHY

Guanqi Liu graduated from the Middle School attach to Jilin University and No.11 High School, Changchun, China in 2003 and 2006. He received the Bachelor Degree in Changchun University of Science and Technology, Changchun, China and Master Degree of Transportation Engineering from Texas Southern University, Houston, USA, in 2010 and 2014.

After getting his master degrees, he then joined a transportation consulting company, CLR Analytics, Inc., working as a Traffic Engineer in Irvine, California. His major responsibility in CLR is developing transportation simulation models, collecting traffic data, analyzing MOEs from simulation model and drafting technical reports. He had been fully involved in 2 consulting projects during the time he worked at CLR Analytics. One is LA Metro's I5/I405 HOV Direct Connector project which is sponsored by LA Metro. He took fully involved in this project by developing simulation models, analyzing and pre-processing traffic data, analyzing MOEs from simulation model by evaluating various alternatives, and drafting technical memos and meeting with clients. The other one is SHPR2 consulting project to test and investigate the reliability of a VBA based software by using the real world traffic data. He then enrolled in George Mason University for his Ph.D. degree and hired as Research Assistant. During his time in George Mason University, he has been involved in projects funded by TransInfo University Transportation Center, and Virginia Department of Transportation. His expertise included traffic simulation model development at various levels (Macro/Meso/Micro), bid data processing, machine learning, and transportation planning models.

Guanqi Liu received his Doctorate in Civil and Infrastructure Engineering (with specialization in Transportation Engineering) from George Mason University in spring, 2019.