ABSTRACT

Title of Dissertation:	SIMULATION-BASED OPTIMIZATION OF TRANSPORTATION SYSTEMS: THEORY, SURROGATE MODELS, AND APPLICATIONS
	Xiang He, Doctor of Philosophy, 2014
Dissertation directed by:	Lei Zhang, Associate Professor Department of Civil & Environmental Engineering

The construction of new highway infrastructure has not kept pace with the growth of travel, mainly due to the limitation of land and funding availability. To improve the mobility, safety, reliability and sustainability of the transportation system, various transportation planning and traffic operations policies have been developed in the past few decades. On the other hand, simulation is widely used to evaluate the impacts of those policies, due to its advantages in capturing network and behavior details and capability of analyzing various combinations of policies. A simulation-based optimization (SBO) method, which combines the strength of simulation evaluation and mathematical optimization, is imperative for supporting decision making in practice.

The objective of this dissertation is to develop SBO methods that can be efficiently applied to transportation planning and operations problems. Surrogate-based methods are selected as the research focus after reviewing various existing SBO methods. A systematic framework for applying the surrogate-based optimization methods in transportation research is then developed. The performance of different forms of

surrogate models is compared through a numerical example, and regressing Kriging is identified as the best model in approximating the unknown response surface when no information regarding the simulation noise is available. Accompanied with an expected improvement global infill strategy, regressing Kriging is successfully applied in a real world application of optimizing the dynamic pricing for a toll road in the Inter-County Connector (ICC) regional network in the State of Maryland. To further explore its capability in dealing with problems that are of more interest to planners and operators of the transportation system, this method is then extended to solve constrained and multi-objective optimization problems.

Due to the observation of heteroscedasticity in transportation simulation outputs, two surrogate models that can be adapted for heteroscedastic data are developed: a heteroscedastic support vector regression (SVR) model and a Bayesian stochastic Kriging model. These two models deal with the heteroscedasticity in simulation noise in different ways, and their superiority in approximating the response surface of simulations with heteroscedastic noise over regressing Kriging is verified through both numerical studies and real world applications. Furthermore, a distribution-based SVR model which takes into account the statistical distribution of simulation noise is developed. By utilizing the bootstrapping method, a global search scheme can be incorporated into this model. The value of taking into account the statistical distribution of simulation noise in improving the convergence rate for optimization is then verified through numerical examples and a real world application of integrated corridor traffic management.

This research is one of the first to introduce simulation-based optimization methods into large-scale transportation network research. Various types of practical

problems (with single-objective, with multi-objective or with complex constraints) can be resolved. Meanwhile, the developed optimization methods are general and can be applied to analyze all types of policies using any simulator. Methodological improvements to the surrogate models are made to take into account the statistical characteristics of simulation noise. These improvements are shown to enhance the prediction accuracy of the surrogate models, and further enhance the efficiency of optimization. Generally, compared to traditional surrogate models, fewer simulation evaluations would be needed to find the optimal solution when these improved models are applied.

SIMULATION-BASED OPTIMIZATION OF TRANSPORTATION SYSTEMS:
THEORY, SURROGATE MODELS, AND APPLICATIONS

By

Xiang He

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Lei Zhang, Chair/Advisor
Professor Michael Fu
Professor Ali Haghani
Professor Elise Miller-Hooks
Professor Paul Schonfeld

Acknowledgements

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

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

I would like to extend my thanks to the rest of my dissertation committee: Professor Michael Fu, Professor Ali Haghani, Professor Elise Miller-Hooks and Professor Paul Schonfeld, for their encouragement, insightful comments and valuable suggestions to my research.

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

I thank all my colleagues in the Transportation Systems Research Lab at the University of Maryland. The ideas and suggestions they shared with me during our group meetings have greatly contributed to the completion of this research. I would also like to thank my fellow graduate students from other groups in the transportation

program. I appreciate the opportunities to take class, cooperate on projects and have discussions with them.

Finally, I would like to express my gratitude towards my family and dedicate this dissertation to them. They have been encouraging me, supporting me and loving me all the time. I am especially thankful to my wife, Zheng Zhang, for her understanding and love, and for coming to the US to accompany me during those days I worked on this dissertation.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AEOV	Absolute Error of Optimal Value
AMS	Anisotropic Mesoscopic Simulation
ANOVA	Analysis of Variance
BSK	Bayesian Stochastic Kriging
CCD	Central Composite Design
CI	Confidence Interval
CV	Cross Validation
DMS	Dynamic Message Signs
DoE	Design of Experiments
D-SVR	Distribution-based Support Vector Regression
DTA	Dynamic Traffic Assignment
DynusT	Dynamic Urban Systems in Transportation
EGO	Efficient Global Optimization
EGOp	Estimated Global Optimum
EI	Expected Improvement
FDSA	Finite Differences Stochastic Approximation
FIFO	First-In, First-Out
GA	Genetic Algorithm
GFV	Gap Function Vehicle
HOT	High-Occupancy Toll
HOV	High-Occupancy Vehicle

H-SVR	Heteroscedastic Support Vector Regression
IAGO	Informational Approach to Global Optimization
LHS	Latin Hypercube Sampling
LR/SF	Likelihood Ratio/Score Function
MAE	Maximum Absolute Error
MCP	Multiple Comparison Procedures
MLE	Maximum Likelihood Estimation
MoEs	Measures of Effectiveness
MPEC	Mathematical Program with Equilibrium Constraints
NMAE	Normalized Maximum Absolute Error
NRMSE	Normalized Root Mean Squared Error
OO	Ordinal Optimization
PA	Perturbation Analysis
PCC	Pearson Correlation Coefficient
PDF	Probability Density Function
PI	Probability of Improvement
RBF	Radial Basis Function
RBFNN	Radial Basis Function Neural Network
RMSE	Root Mean Square Deviation
R&S	Ranking and Selection
RS	Radom Search
SA	Stochastic Approximation
SBO	Simulation-based Optimization

SKO   Sequential Kriging Optimization

SNOBFIT Stable Noisy Optimization by Branch and FIT

SPSA   Simultaneous Perturbation Stochastic Approximation

SUE   Stochastic User Equilibrium

SVM   Support Vector Machine

SVR   Support Vector Regression

TTI   Texas Transportation Institute

UE   User Equilibrium

VSL   Variable Speed Limit

List of Mathematical Symbols

b	Index of bootstrap sample
B	Bootstrap sample size
$\mathbf{B}$	Covariance matrix of the prior distribution of $\boldsymbol{\beta}$
C	Coefficient in SVR governing the tradeoff between model complexity and prediction accuracy
$f(\bullet)$	Objective function
$\hat{f}(\bullet)$	Estimated objective function
$F(\bullet)$	Sample response function
$\mathbf{g}_{\min}$, $\mathbf{g}_{\max}$	Lower and upper bound for constraints
$G(\bullet)$	Sample constraint function
i	Index of input sample point
$\mathbf{I}$	Identity matrix
j	Index of regression basis functions
k	Dimension of input variables
$L(\bullet)$	Likelihood function
$\tilde{L}(\bullet)$	Negative log-likelihood function
m	Dimension of the vector of regression basis functions
M	Number of points in the Pareto set
n	Sample size
$p(\bullet)$	Probability density function
$\mathbf{q}(\bullet)$	Vector of regression basis functions
$\mathbf{Q}$	Matrix of regression basis functions

r	Index of replication of simulation
R	Number of replications of simulations
$\mathbb{R}$	The real set
$\hat{s}^2(\bullet)$	Estimated predictor's variance
$s_\varepsilon^2(\bullet)$	Sample variance in Bayesian stochastic Kriging
$\mathbf{w}$	Vector of weighted coefficients of basis function
$\mathbf{x}$	Vector of input variables
$\mathbf{x}_{min}$, $\mathbf{x}_{max}$	Lower and upper bound for input variables
$\mathbf{X}$	Set of input vectors
y, $\hat{y}$	Simulation response and its estimation
$\mathbf{y}$	Set of simulation responses
$Y(\bullet)$	Random variable corresponding to simulation response
$\mathbf{Z}(\bullet)$	Random field in Bayesian stochastic Kriging
α^+, α^-, η^+, η^-	Dual variables in SVR
β_{Gumbel}	Scale parameter for Gumbel distribution
$\boldsymbol{\beta}$	Weight parameters of regression basis functions
$\boldsymbol{\beta}_0$	Mean of the prior distribution of $\boldsymbol{\beta}$
ε	SVR insensitive margin
$\varepsilon(\bullet)$	Random field in Kriging
$\varepsilon_r(\bullet)$	Simulation random noise in Bayesian stochastic Kriging
$\phi(\bullet)$	Probability density function
$\phi_{P-Gaussian}(\bullet)$	Product of the probability density functions of two Gaussian distribution

$\mathbf{\Phi}$	Gram matrix
$\varphi(\bullet)$	Basis function
γ	Euler's constant
$\mu,\ \hat{\mu}$	Mean of the objective function and its MLE estimate
μ_{Gumbel}	Location parameter for Gumbel distribution
θ	Scaling coefficients for Kriging
$\sigma^2,\ \hat{\sigma}^2$	Variance of the random field in Kriging and its MLE estimate
$\mathbf{\Sigma}_\varepsilon$	Intrinsic covariance matrix in Bayesian stochastic Kriging
ξ	SVR prediction error
$\left|\xi\right|_\varepsilon$	Prediction loss for traditional SVR
$\xi^+,\ \xi^-$	Prediction loss for under-prediction and over-prediction in SVR
$\psi(\bullet)$	Kriging basis function
$\mathbf{\psi}$	Vector of Kriging basis functions
$\mathbf{\Psi}$	Correlation matrix for Kriging
$\tilde{\mathbf{\Psi}}$	Correlation matrix for regressing Kriging

Chapter 1: Introduction

1.1 Background

Transportation systems are core public infrastructure that is closely related to the everyday life of personal people. It does not only directly serve people's travel demand, but also supports living needs indirectly through the transport of goods. Efficient transportation can benefit the society in various aspects, including stimulating the economy, improving the environment, reducing energy consumption, etc. Keeping the transportation system in the best possible condition with the available resources is the primary goal for transportation system planners and operators.

Transportation planning is defined as the process of developing strategies for operating, managing, maintaining, and financing the area's transportation system in such a way that it will advance the area's long-term goals (FHWA and FTA, 2007). The most widely used method for transportation planning is the rational planning model (Friedmann, 1987), which is composed of five major modules: defining goals and objectives, identifying problems, generating alternatives, evaluating alternatives, and developing plans. Figure 1 illustrates the typical steps involved in the rational planning process. It can be noticed that the optimal decision is not derived through searching thoroughly over the entire feasible domain. Instead, several alternatives are generated and evaluated first, and the optimal strategy is selected through pairwise comparison of the system performance of those alternatives. In practice, only a finite

number of alternatives can be analyzed for this what-if scenario-based analysis, which could be a major disadvantage of the rational planning model.
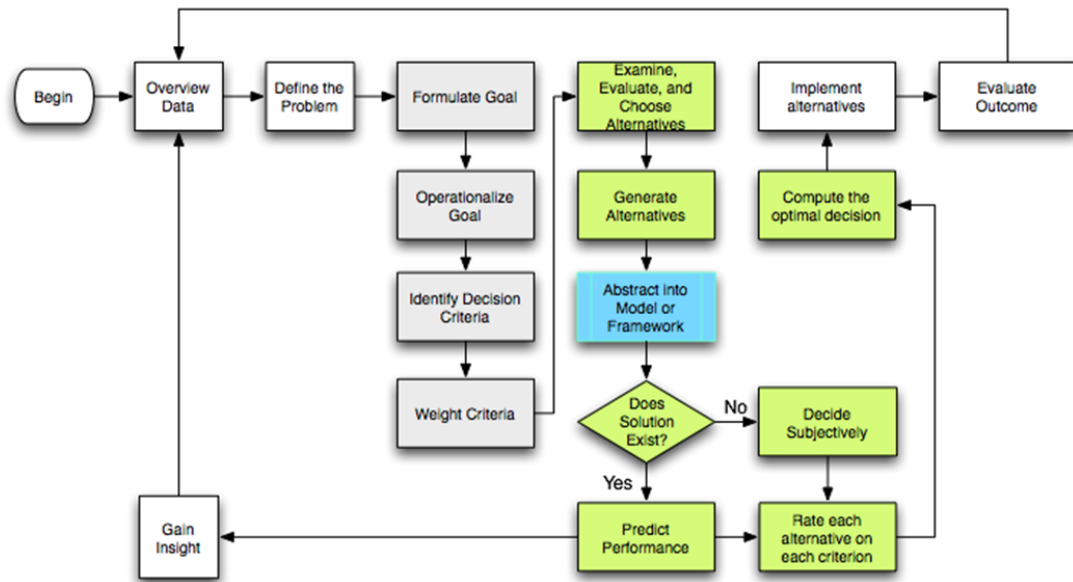
Figure 1-1: Rational Planning Model Flowchart. (Source: Levinson et al., 2013)

For decisions on infrastructure construction, the assumption of limited alternatives may be appropriate because this type of transportation planning strategies faces strict constraints in terms of both budget and land use. There are usually very few choices available for consideration. However, for decisions on travel demand management policies and traffic operations strategies, the limited alternative assumption can hardly hold. Unlike the discrete nature of construction decisions (e.g. mode to investment, number of lanes, alignment, etc.), the optimization of travel demand management policies and traffic operations strategies are usually continuous or approximately continuous problems (e.g. pricing for tolled facilities, ramp metering strategy, traffic signal plan, etc.). Thus unlimited alternatives should be considered. The goal of this type of problem becomes seeking improved settings of

policy instruments from all possible options instead of selecting the best settings from a fixed set of alternatives. In this situation, scenario-based analysis is no longer a sufficient method for searching for the optimal solution. A routine that retrieves information from evaluated alternatives and intelligently generates new promising alternatives is needed.

There is already a large number of optimization methods developed in the operations research field. However, none of those methods can solve all types of problems effectively. A prerequisite of the success in resolving a specific optimization problem is the judicious choice of appropriate methods. The key factors that influence the choice of optimization methods include the structure of the objective functions, constraints, dimension of the problems, etc. Of all these factors, the structure of the objective functions plays a rather crucial role as it determines what information is available for the optimization module. These include whether the problem is a linear problem, whether gradient information is available, how much time it takes to evaluate one decision variables combination and so on. In the following subsection, the major forms of objective functions for transportation planning and operations decision problems will be introduced.

1.2 Objective Evaluation

The performance of the transportation network is the result of complex travelers' behaviors and their interactions with the network infrastructure. To evaluate the transportation system performance under certain planning and operations

3

strategies, the influence of implemented measures on the network characteristics and the resulting impact on travelers' behavior need to be modeled. At the macro level, traveler's behavior is usually modeled by mathematical equations with various assumptions (e.g. Wardrops's first principle, utility maximization, etc.). These mathematical equations have been applied in the widely-used four-step model as well as the more advanced activity-based travel demand models. On the other hand, simulation models are developed for the meso or micro level analysis of travel behavior. By intelligently modelling the true behavior of individual traveler/group of travelers, simulation can introduce more stochasticity into the system. At the same time, simulation has a better capability in describing the dynamics of travel behavior than typical analytical models. Both mathematical equations and simulation models have been widely utilized to evaluate the performance of transportation systems. In the following two subsections, a brief review of transportation planning and operations decision problems whose objectives are evaluated through these two methods will be presented. Moreover, the advantages and disadvantages of the two evaluation methods are also summarized.

1.2.1 Mathematical Methods

The analysis of transportation planning and management problems is usually conducted for a relatively large network. Decision makers are interested in the overall performance of the regional network under a certain planning and management strategy. In this case, the behavior of each traveler in response to the policy stimulus should be modeled in the first stage, and the aggregate performance can then be

summarized based on the individual behaviors. Various decision making problems of transportation planning and management have been formulated mathematically as bi-level optimization, with the upper-level formulating the objective function while the lower-level modeling individual travelers' behavior in response to the policy. The main outputs of the lower-level problem are the travel path for each traveler and the associated attributes of each travel path. Usually, travelers are assumed to make their route choices in a user optimal manner, and hence the lower-level problem can be formulated as a user equilibrium (UE) traffic assignment model. To relax the assumption that travelers have complete knowledge of the network conditions, stochastic user equilibrium (SUE) models can also be applied for the lower level problem. As for the objective defined by the upper-level problem, objectives which have been frequently considered include minimizing total travel cost, minimizing average trip travel time, minimizing total vehicle miles traveled, maximizing total revenue, maximizing consumers' surplus, etc.

Optimization for transportation planning and management problems with objectives evaluated through mathematical equations has been extensively studied in the existing literature. LeBlanc and Boyce (1986) investigated the optimal road network design in achieving the lowest summation of total travel cost and network improvement cost. In this paper, the problem is formulated following the bi-level structure, and both upper and lower problems are approximated with piecewise linear functions for the convenience of solution derivation. The optimal congestion pricing under different network conditions in terms of maximizing economic benefit was studies in the 1990's. Cases with the lower level optimization specified as a

deterministic user equilibrium problem (Yang and Huang, 1998) or a stochastic user equilibrium problem (Yang, 1999) were discussed separately. Chan and Lam (2002) proposed a bi-level programming model to determine the best density for speed detector deployment in a network with travel time information provided to travelers. The lower-level problem was a probit-based traffic assignment model, and the upper-level problem searched for the speed detector density that minimized the measured travel time error variance as well as the social cost of the speed detectors. More recent applications of bi-level optimization structure for transportation planning and management include optimizing shelter location for hurricane events (Li et al., 2011), deriving optimal land use development plan to improve network reliability (Yim et al., 2011), finding the optimal combination of exclusive transit lanes on a network basis (Mesbah, 2011), and others.

Decision making on traffic operations measures has also been formulated as optimization problems with objective functions expressed by mathematical equations. For urban road traffic control, isolated intersection control strategies with either fixed-time or traffic-response strategies were investigated, respectively (Improta and Cantarella, 1984; Vincent and Young, 1986). Later, both strategies for coordinated control of signal at multiple intersections were studied (Stamatiadis and Gartner, 1996; Kessaci et al., 1990; Gartner et al., 2001). These research efforts generally formulate the problem as a constrained optimization, and model the traffic dynamics with different types of assumptions either at the link level or at a more detailed level in a way such as cell-transmission representation. Regarding freeway traffic control, the most direct measure ramp metering has been studied. The optimization of ramp

metering strategies is formulated as mathematical programs with traffic dynamics described by different traffic flow models, among which both fixed-time ramp metering strategies and reactive ramp metering strategies analyzed (Papageorgiou, 1980; Chen et al., 1997; Zhang et al., 2001). In addition, mathematical methods have been applied to evaluate the performance of other types of traffic control strategies, such as dynamic message signs (DMS) (Messmer and Papageorgiou, 1995; Ben-Akiva et al., 1997), variable speed limit (VSL) (Yang et al., 2013), etc. More advanced integrated optimization of combined traffic operations strategies has also been investigated (Kotsialos et al., 2002; Hegyi et al., 2005; Carlson et al., 2010).

The mathematical methods are of interest to researchers mainly due to their tractability. In this case, the computation of the objective values would be very convenient, and traditional optimization approaches can be easily applied. However, this method suffers from the limitation that strong assumptions on travel behavior and network dynamics need to be set up in order to build up the analytical form for the objective functions. For instance, one major behavior dimension needs to be considered for evaluating transportation planning policies is route choice. The travelers' route choice is usually described by UE or SUE models, in which link travel time is assumed to always decrease along with the increase of link flow. Link flow is supposed to be able to increase without limit. Meanwhile, the influence of traffic control measures (e.g. traffic signal, ramp metering, etc.) on route choice cannot be captured, and some real world phenomenon such as queuing and the First-In, First-Out (FIFO) constraint are not taken into account. All these features limit the ability of the mathematical methods to reflect the traffic realism accurately. For the

evaluation of traffic operations strategies, the major concern is the traffic dynamics. Common assumptions associated with mathematical methods include constant link speed, ignorance of route choice behavior in response to traffic control, etc. Moreover, it is more appropriate to apply the mathematical methods for analysis in small regions (e.g. a single intersection or a single corridor). In the case of utilizing mathematical methods evaluating the impact of traffic control measures implemented in large-scale networks (i.e. regional network or even larger), further assumptions such as the store-and-forward concept need to be made to reduce the computational complexity. As the analytical models for the impact of transportation planning policies and traffic operations strategies focus on different aspects of the transportation system, it is very difficult to build a single analytical model capable of evaluating the joint impact of planning policies and operations strategies on the transportation system performance, which is another limitation of the mathematical methods.

1.2.2 Simulation Models

Simulation models mentioned in this section mainly refer to microscopic or mesoscopic models, which capture more detailed traffic dynamics than macroscopic models. Compared to analytical models, fewer assumptions at the aggregate level which obviously violate the characteristics of real traffic (e.g. unlimited link volume, ignorance of the FIFO constraint, etc.) are imposed in simulation. Thus transportation management agencies are becoming more favorable to simulation models and are prone to use simulation for informing decisions about specific investment or

management and operations policies in the transportation system. Computerized simulation models representing the transportation system were developed since the 1950's in the United States (Wang and Prevedouros, 1996), to address more and more operational needs in the planning and operations decisions. In simulation models, specific rules are assigned to individual or group of travelers in their behavior, and their response to the network infrastructure. The entire transportation system is then formed with lots of the complex interactions among travelers as well as between the travelers and the network infrastructure. Therefore, the evolution of the transportation system is essentially determined by those behavioral rules. This bottom-up design makes simulation models very good representatives of the real world transportation systems, and thus more reliable evaluation tools for agencies in informing decisions.

Various simulation models at different analysis levels for the transportation system have been developed, such as CORridor SIMulation (Halati et al., 1997), TransModeler (Balakrishna et al., 2009), Vissim (Fellendorf and Vortisch, 2010), Aimsun (Barceló and Casas, 2005), DynusT (Chiu et al., 2010a), DynaMIT (Ben-Akiva et al., 2002), DYNASMART (Mahmassani, 2002), Dynameq (Tian et al., 2007), etc. In terms of application, simulation models have been widely used in evaluating the effect of transportation planning and operations policies. Examples include studies on congestion pricing (De Palma et al., 2005), high-occupancy toll (HOT) lane pricing (Murray et al., 2001), traffic signal (Mosseri et al., 2004), dynamic route guidance (Gao et al., 2008), ramp metering (Bellemans et al., 2006), etc.

In general, simulation models exhibit strong advantages in capturing network and behavior details, which is of great benefit to transportation planning and management agencies. In addition, as fewer aggregate-level strong assumptions on traffic dynamics are imposed in simulation models, the evaluation through simulation is believed to perform better in representing the real traffic than mathematical methods. Moreover, a simulation model is suitable for evaluating the impact of combinations of transportation planning and operations strategies, since different levels of traffic details are taken into account simultaneously. The major disadvantage associated with simulation models is that the evaluation of objective functions through simulation is usually computationally expensive. It may cost hours to days for a single run of the simulation on a large network (e.g. the entire road network for a metropolitan area). This would make optimization based on simulation evaluation rather time consuming. Meanwhile, as multiple random processes are involved in the simulation, there are usually no closed forms for the objective functions. In this case, it would be difficult to retrieve information such as gradient, which further restricts the usage of traditional optimization methods (e.g. Newton's method) when simulation evaluation is applied.

1.3 Research Objectives and Methodology

Along with the increase of vehicle ownership and population, travel demand in the U.S. continues to grow, but the construction of new highway infrastructure has not kept pace with the growth of travel due to the constraints of land availability and

budget limit. According to Highway Statistics (1980, 2010), the vehicle miles traveled in the U.S. increased 94 percent between 1980 and 2010, while the miles of highways only increased 5.4 percent during the same period of time. The resulting traffic congestion problem thus becomes the major issue that transportation planners and operators need to take care of, especially in urban areas where population is densely distributed. The Texas Transportation Institute (TTI) estimated in their 2012 Urban Mobility Report (Schrank et al., 2012) that congestion in 498 metropolitan areas caused urban Americans to travel 5.5 billion hours more and to purchase an extra 2.9 billion gallons of fuel for a congestion cost of $121 billion in 2011.

Besides adding highway capacity, various transportation planning and operations strategies have been proposed to improve mobility and reduce congestion, including congestion pricing, traffic signal, access management, travel demand management, traveler information, etc. These policies may be implemented separately or jointly in a particular area. The major challenge for a transportation agency is how to select specific policy variables, with the aim of operating the transportation system in the most efficient way. The evaluation of the impact of specific policies on the transportation system performance is thus a crucial step in the decision making process.

Realizing the advantages of simulation models in representing the traffic realism, transportation agencies are more prone to choose simulation models as the evaluation tool than mathematical methods. Therefore, the objective of this dissertation is to evaluate transportation planning and operations strategies with simulation, and develop methods to optimize those strategies based on simulation

evaluation. Meanwhile, as the objective functions would be very expensive to evaluate when simulation is applied, the developed optimization method should be very efficient and able to find optimal or near-optimal solutions with a reasonable computation budget.

In order to achieve the specified objectives, the dissertation starts with a review of existing simulation-based optimization (SBO) methods implemented in both the transportation field and other disciplines. The appropriate method and implementation framework are then developed according to the characteristics of transportation simulation and the specified requirement of limited computation budget. The optimization methods for problems in different forms (i.e. single-objective optimization, multi-objective optimization and constrained optimization) are investigated. To deal with the heterogeneity of error variance and asymmetric distribution observed in transportation simulation outputs, advanced SBO methods including enhanced support vector regression (SVR) and Bayesian Stochastic Kriging (BSK) models are developed. The associating infill strategies for the iterative application of these methods are also discussed.

1.4 Outline of the Dissertation

This dissertation is composed of 8 chapters. Existing SBO methods developed in different fields are first reviewed. The appropriate method suitable for implementation in the domain of transportation research is then selected and applied. Based on the observation of specific characteristics associated with transportation

simulation, the selected method is then improved to make it fit better the particular problem. The remaining chapters are arranged as follows.

Chapter 2 reviews SBO methods developed in existing literature, including both methods discussed theoretically and those applied practically. The features associated with each method are analyzed and compared. Specifically, a review on the applications of SBO in transportation research is conducted.

Chapter 3 introduces the framework as well as the technical details of the surrogate-based optimization method, which is selected as the method to be applied in the current research. Different types of surrogate models are discussed and compared using a numerical example. The promising surrogate models identified through the numerical test are then utilized for a real world application of optimizing the dynamic pricing for a toll road.

Chapter 4 further extends the capability of the surrogate models in dealing with problems of more interest to decision makers. Constrained optimization and multi-objective optimization problems are formulated, and the appropriate infill strategies are applied accompanying with the regressing Kriging model to solve these two types of problems.

Chapter 5 describes the heterogeneity of error variance observed in the simulation outputs from the dynamic road pricing case study, and then develops an enhanced SVR model named heteroscedastic SVR for the simulation-based optimization, taking into account the heteroscedastic simulation noise.

Chapter 6 develops another type of surrogate model named Bayesian stochastic Kriging, which also takes into account the heterogeneity in simulation

outputs. Meanwhile, the uncertainty in parameter estimation is considered in this model. A different case study with integrated optimization of HOT toll rate and freeway diversion control is created for the application of this model.

Chapter 7 further investigates the distribution of simulation noise, and develops an improved surrogate model based on SVR, which is named distribution-based SVR, with prediction error penalized in a way related to the probability density function of simulation noise. In addition, the method of bootstrapping is used for the estimation of predictor's variance. The expected improvement infill strategy is then incorporated to the surrogate model for global optimization.

Chapter 8 concludes the dissertation. Major findings and contributions of the research is summarized. The dissertation ends with the proposition of future research directions as the extension of the current research.

Chapter 2: Literature Review

Simulation-based optimization is a technique that integrates optimization routines into simulation analysis. A typical SBO problem consists of decision variables serving as parameters in simulation and objective functions evaluated through simulation. Sometimes, simple box constraints to decision variables or constraints that are associating measurements of experimental simulations would be imposed to the problem. The problem can usually be formulated in the following from,

$$\min_{\mathbf{x} \in \mathbb{R}^k} f(\mathbf{x}) = E[F(\mathbf{x}, w)] \qquad (2.1)$$

$$\text{s.t. } \mathbf{x}_{\min} \le \mathbf{x} \le \mathbf{x}_{\max}$$

$$\mathbf{g}_{\min} \le E[G(\mathbf{x}, w)] \le \mathbf{g}_{\max}$$

where $\mathbf{x} \in \mathbb{R}^k$ is the k-dimensional vector of input variables, $f(\mathbf{x})$ represents the objective function, w is a sample path (simulation replication), and F is the sample response function (simulation). When the simulation model F is a stochastic simulation, a single run of the model F only provides an estimator of the true response. Thus the objective function of the problem is the expectation of the simulation output. The vector of decision variables $\mathbf{x}$ is restricted by the lower bound $\mathbf{x}_{\min}$ and upper bound $\mathbf{x}_{\max}$. The last inequality represents another constraint that is evaluated through the simulation function $G(\mathbf{x}, w)$, and the lower and upper bound for this constraint are $\mathbf{g}_{\min}$ and $\mathbf{g}_{\max}$, respectively.

There are quite a few excellent review papers on the subject of simulation-based optimization approaches. Fu (1994) reviewed methods for optimizing discrete-even systems via simulation. Both the discrete parameter and the continuous parameter cases were discussed. Techniques for optimization from a finite set including multiple-comparison procedures and ranking-and-selection procedures were introduced for the discrete parameter case, while gradient-based methods were the focus for the continuous parameter case. Andradóttir (1998) presented a review focusing on gradient-based methods for optimization of problems with continuous decision variables and random search methods for problems with discrete decision variables. Tekin and Sabuncuoglu (2004) summarized the latest development in SBO methods, and classified the existing techniques according to problem characteristics such as the shape of the response surface (global as compared to local optimization), objective functions (single or multiple objectives) and parameter spaces (discrete or continuous parameters). More recently, Hachicha et al. (2010) provided a literature survey with all classification criteria and proposed a global classification scheme of SBO methods. In their review paper, SBO problems were classified regarding their input variables (quantitative variables and qualitative variables), output variables (single-objective problem or multi-objective problem), parameter spaces (discrete or continuous parameters), the shape of the response surface (global as compared to local optimization), or by their optimization procedure.

In addition to the reviews of theoretical development of SBO methods, a discussion regarding the progress of SBO methods in application was provided in Fu (2002). The optimization modules in five commercial software —AutoStat,

16

OptQuest, OPTIMIZ, SimRunner and WITNESS Optimizer were introduced. Based on the review, the optimization procedures implemented in simulation software were all based on metaheuristics and predominantly evolutionary algorithms.

2.1 Simulation-Based Optimization Methods

On the basis of the review of SBO approaches aforementioned, we group the most popular methods into five categories according to the characteristics of the methods in this dissertation. The five categories are methods primarily for discrete optimization problems, gradient-based methods, mathematical programming-based methods, metaheuristics, and surrogate-based methods. The following subsections briefly introduce each type of the SBO methods.

2.1.1 Methods Primarily for Discrete Optimization Problems

The SBO methods suitable for discrete optimization problems are reviewed in Swisher et al. (2004). The methods are differentiated based on the size of the feasible solution set. When the feasible solution set is finite and small, ranking and selection (R&S) and multiple comparison procedures (MCP) would be appropriate. On the other hand, if the set is infinite or very large, techniques such as random search (RS) and ordinal optimization (OO) can be applied.

Both R&S and MCP methods need to perform exhaustive evaluations of all feasible solutions. This is only practical when the solution set is small. Being

different from traditional optimization methods, R&S and MCP focus more on the comparison aspect than searching.

There are two important concepts in the R&S method which need user specification: an indifference zone δ and a confidence level P. With multiple replications for each parameter setting, the sample mean and variance of the simulation output can be computed. According to the user-defined indifference zone and confidence level, the number of additional replications for each parameter setting is determined. The optimal solution is then chosen as the parameter setting with the smallest average sample mean. This derived optimal solution can be guaranteed to be within δ of the true optima at the confidence level P (Law and Kelton, 1991).

The MCP method is different from R&S in that it makes conclusions based on the confidence intervals of the difference between any pair of parameter settings instead of directly comparing the sample means. With the assumption that the simulation for different parameter settings is independent and the simulation noise follows the normal distribution, the confidence interval for the difference between any pair of input points can be computed. When all confidence intervals are formed, one would simply look for the parameter setting that the confidence interval for the difference with all other pairs is strictly negative. This parameter setting is the clear winner in the feasible solution set. If no clear winner is found, one can crudely eliminate some candidates, run more replications for the smaller set and then look for the clear winner. This process can be repeated until the conclusive inference can be made (Hsu and Nelson, 1998; Goldsman et al., 1991).

For problems with infinite feasible solutions, it is definitely impractical to apply methods that need exhaustive evaluations of all possible solutions. The appropriate method for this type of problem should employ some mechanism to reduce the size of the effective solution set.

The random search method is of interest due to the existence of theoretical convergence proofs (Fu, 2002). The main idea of the method is to move iteratively from a current parameter setting to another one in the neighborhood of the current point. The number of visits to each design point would be counted. Within each iteration, a new design point is selected from the neighborhood of the current point according to some pre-specified probability distribution. This new point is then evaluated through simulation and compared to the performance of the current point. The counter for the point with better performance would be increased by one, and the new current point is updated. However, this algorithm may face implementation difficulty of sampling randomly from the neighborhood with the appropriate distribution (Banks et al. 2000). More details on the random search method in simulation can be found in Andradóttir (2005).

Ordinal optimization can reduce the search for optima from a very large sample to a much smaller one, by softening the goal of looking for the best to looking for a solution that is good enough. The development of this method is based on the observation that finding ordering among candidate solutions is much easier than carrying out the estimation for each solution individually in most cases. It is found that estimating the difference in performance for two design points using the sample mean is governed by the Monte Carlo convergence rate of $1/\sqrt{n}$, while deciding the

ordering of two design points based on the sample mean has an exponential convergence rate (Fu et al., 2005). More details on the ordinal optimization method can refer to Ho et al. (1992), Ho and Deng (1994) and Lee et al. (1999).

2.1.2 Gradient-Based Methods

Gradient-based methods applied in simulation-based optimization mainly refer to the stochastic approximation (SA) algorithms. Stochastic approximation is essentially the adaption of gradient search method in deterministic optimization for stochastic settings. This method iteratively updates the current solution by moving toward the gradient direction,

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \nabla \widetilde{f}(\mathbf{x}_k) \tag{2.2}$$

where $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$ are the current solution and the updated solutions, respectively. $\nabla \widetilde{f}(\mathbf{x}_k)$ represents the estimated gradient of the objective function at the current design point, and α_k is the step length.

As SA searches for the updated solution along a line, it is a local optimization method. Under proper conditions, the SA method can guarantee to converge to a local optimum. The appropriate estimation of the gradient is very crucial to the effectiveness of the method. A review of available gradient estimators that can be implemented in simulation is provided in Fu (2006). Popular techniques for gradient estimation include finite difference estimation, perturbation analysis (PA) (Spall, 1992; Bettonvil, 1989; Glasserman, 1991; Fu and Hu, 1994), likelihood ratio/score function (LR/SF) estimators (Glynn, 1987; Glynn, 1989; Rubinstein and Shapiro,

1993), and weak derivatives (Heidergott, 2001). Based on the theory of finite difference estimation, three alternative estimators are generated, which are naïve one-sided finite difference estimation, two-sided symmetric difference estimation and simultaneous perturbations. Generally, only single run of simulation is needed for the PA and LR/SF estimators, but the implementation of these two estimators require knowledge about the structure of the simulation and modifications to the simulation source code. On the other hand, application of the finite difference estimation and weak derivatives method require multiple simulation runs, and the gradient estimation is usually noisier than the PA and LR/SF estimators.

2.1.3 Mathematical Programming-Based Methods

The main idea of mathematical programming based methods is to evaluate a relatively large set of samples, so as to approximately turn the stochastic problem into a deterministic problem. In this way, the powerful deterministic optimization methods already developed can be directly applied for simulation-based optimization problems (Robinson, 1996). Sample-path optimization is the most popular method falls into this category implemented for SBO problems. It takes many simulations for each of the design point, and conducts optimization based on the sample means,

$$\bar{f}_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \tilde{f}_i(\mathbf{x}) \tag{2.3}$$

where $\tilde{f}_i(\mathbf{x})$ are the independent and identically distributed (i.i.d.) unbiased estimates of the true function f (i.e. the output from a single run of simulation), n is the

number of replications, and $\overline{f}_n(\mathbf{x})$ represents the sample mean at the design point $\mathbf{x}$ over n replications. According to the strong law of large numbers,

$$\overline{f}_n(\mathbf{x}) \to f(\mathbf{x}) \text{ with probability 1.}$$

The task of sample-path optimization is then to optimize the deterministic function $\overline{f}_n(\mathbf{x})$. If the derivatives are available for this approximated deterministic problem, the effectiveness of this approach would be enhanced significantly, as a lot of nonlinear programming packages require them. However, the major disadvantage of this approach is the requirement of large number of replications of simulation evaluation, which are usually very computationally expensive.

2.1.4 Metaheuristics

Metaheuristics have been applied successfully in a lot of real world applications. The main strength of metaheuristics is their flexibility. They can be applied in various types of problems (e.g. discrete optimization problem, continuous optimization problem, deterministic optimization problem, stochastic optimization problem, etc.), and all types of constraints can be taken into account. Moreover, researchers are interested in them because they are global optimization methods. Four most popular metaheuristics that have been applied to simulation-based optimization are genetic algorithms, tabu search, simulated annealing and scatter search.

Genetic algorithm is an evolutionary algorithm, which is inspired from the process of biological evolution. This method updates a finite set of solutions iteratively. In each iteration, new solutions are generated through crossover and

mutation processes. Only those satisfying certain criteria would be accepted, and used for the generation of their offspring in the next iteration. The algorithm would proceed to generate more favorable solutions along iterations. More detailed introduction of the genetic algorithm can be found in Goldberg (1989) and Schmitt (2001).

The most notable feature of tabu search is the memory introduced into the optimization framework. This method is based on local search that iteratively moves the current solution to a neighbor solution. Unlike other local search methods, tabu search allows moves to neighbor solutions that have worse objective values than the current solution. In this way, the procedure would not be trapped at a local optimal solution. Meanwhile, a list of forbidden moves would be memorized to avoid circling or infinite loops. The tabu list would be updated along the iterations. Glover and Laguna (1997) provided a very comprehensive reference for tabu search and its applications.

Simulated annealing (Eglese, 1990) mimicked the process of annealing in metallurgy, which involves heating and controlled cooling of a material. Similarly to the tabu search setting, simulated annealing also iteratively searches for a new solution in the neighborhood of the current solution. If a better solution is found, it replaces the current solution with probability one. On the other hand, if a worse solution is found, the current solution would be replaced by it with a probability strictly less than one, and the probability of moving to a worse solution should decrease along the iterations.

Similarly to genetic algorithm, scatter search is also an evolutionary algorithm that iteratively updates a finite set of solutions. Basically, generalized forms of linear combinations of current solutions would be generated as the optional solutions for the updated set. These new solutions are then mapped into associated feasible solutions. After the mapping, the new feasible solutions would be evaluated based on some criteria including diversity, objective function value, etc. Only those defined as good points according to the criteria are kept in the new solution set for the next iteration. An advantage of the scatter search is that information not contained separately in the original points can be captured through the combination process. A complete reference on scatter search can be found in Laguna and Marti (2002).

2.1.5 Surrogate-Based Methods

Surrogate-based optimization approach is a feasible alternative to solve continuous optimization problems with computationally costly objective functions. Surrogate modeling or metamodel-based simulation optimization aims to regress the response surface that characterizes the relationship between the inputs of decision variables and simulation outputs (Hussain et al., 2002; Queipo et al., 2005; Jakobsson et al., 2010). The surrogate simplifies simulation optimization because of its deterministic rather than stochastic relationship between the input and output (Barton and Meckesheimer, 2006). Using only an initial input dataset and corresponding output values of the objective function, the surrogate model can be developed as an approximation of the expensive-to-evaluate objective. In the surrogate-based optimization approach, an unknown function that formulates the relationship between

simulation input and output is approximated with a predefined parametric function whose coefficients can be determined via the experiment design. The most fundamental forms of surrogate models are linear and quadratic polynomial regression (Montgomery, 2008). On the basis of exploration and exploitation of the computationally efficient surrogate, the optimal solution can be obtained.

However, the expediently implemented low-order polynomials may be heavily biased when applied in complex functions of high nonlinearity. More advanced radial basis functions (RBF) (Björkman and Holmström, 2000; Gutmann, 2001; Regis and Shoemaker, 2005; Zhou et al., 2013) and Kriging models are capable of providing good predictions for the complex response surface. A radial basis function neural network (RBFNN) learns input-output mapping by covering the input space with basis functions that transform a vector from the input space to the output space (Adeli and Karim, 2000; Adeli and Jiang, 2009). A support vector regression (SVR) surrogate model provides a good compromise between prediction accuracy and robustness of other approximations (Smola and Schölkopf, 2004; Wandekokem et al., 2011; Li et al., 2013).

Forrester et al. (2006) pointed out that one smooth continuous approximation function (e.g. Gaussian RBF and ordinary Kriging) is unable to fit the discontinuous simulation output due to random noises around the true average response value. The optimization accuracy relies on how accurate the surrogate models are in capturing the performance variations.

2.2 Deterministic vs. Stochastic

In a stochastic simulation, the output from a single run of simulation is just a realization of a random number. In this case, the output is not likely to be equal to the true objective value. Directly using the stochastic output as the estimator of the true objective value in the optimization routines may result in searching in the wrong direction or premature stopping. The noise introduced by stochastic simulation should be properly controlled during the optimization process.

There are mainly three types of methods implemented in stochastic optimization to take into account the simulation noise.

The most common way is to conduct multiple simulation runs at each design point, and use the mean of the output from those replications as the estimator of the true objective value. This is an unbiased estimator which would converge to the true objective value with probability 1 if the number of replications is infinite. The advantage of this method is that it is very flexible and deterministic optimization methods can be directly applied without any revisions. However, the performance of this method is determined by the quality of the estimator, which is heavily dependent on the number of replications. For simulations with expensive-to-evaluate objective functions, raising the number of replications would significantly increase the computational burden for the entire optimization process.

Instead of requiring the accurate objective value for each design point, the second type of method acknowledges the noise in simulation outputs, and allows the objective value used in the optimization routine to be different from the simulation

output. However, this method would only accept deviations within a specific limit. If the deviation is greater than the predefined limit, the part exceeding the limit would be penalized. A typical method in this category is surrogate-based method with SVR as the metamodel. By defining an ε- insensitive loss function, the difference of the surrogate prediction and the actual simulation output at any design point that is larger than ε will be penalized when constructing the SVR response surface. The error acceptance band ε is usually selected to be the standard deviation of simulation noise. Very accurate estimation of the true objective value is not necessary when applying this type of method. Thus only a few replications at each design point are needed, which can already provide rough estimates of the true objective value as well as the variance of simulation noise. In this way, the computational cost is remarkably reduced compared to the first type of method. A problem with this type of method is that the choice of ε is rather arbitrary, and the distribution of the stochastic output is intrinsically assumed to be symmetric.

The third type of method realizes the stochastic nature of the simulation outputs, and incorporates the consideration of the distribution of the simulation noise into the optimization process. In addition to utilizing the mean of simulation outputs as the estimator of the true objective value, the method makes use of the information retrieved from the replications more comprehensively, by assuming the sample variance of simulation outputs to be the variance of simulation noise. This estimated distribution is then explicitly introduced into the optimization process. A recent surrogate-based optimization method with stochastic Kriging metamodels (Ankenman et al., 2010) falls into this category. The simulation noise is assumed to

27

be normally distributed with zero mean and variance equal to the sample variance. The parameters of the metamodel are then estimated through maximum likelihood estimation. Similar to the second type of methods, the number of replications needed for this method is much smaller than the first one. And from the theoretical point of view, this type of methods should perform better as it utilizes more information from the evaluation of design points than the second type of methods. A major disadvantage of this method is that the simulation noise is usually assumed to be normally distributed due to computational convenience. This assumption may restrict the power of the method in dealing with stochastic optimization problems when the actual simulation noise is not normally distributed.

2.3 SBO Application in Transportation Research

As many real world problems are very complex and mathematically intractable, simulation has become an appropriate tool for performance evaluation. Therefore, SBO has been applied in various research fields to help in decision making. Examples of SBO applications have been found in inventory management in supply chains (Schwartz et al., 2006), logistics management (Kochel et al., 2003: Yoo et al., 2010), production planning and scheduling (Feng and Wu, 2006; Kazemi and Zarandi, 2008), wireless sensor network design (Simon et al., 2003), building design (Tresidder et al., 2012; Gengembre et al., 2012), design of aircraft (Simpson et al., 2001), etc.

28

Specifically for the research in the transportation planning and operations field, applications of SBO are not very common yet. In the literature, transportation engineering optimization problems are characterized by computationally expensive objective functions, high dimensional decision variables, and stochastic simulation experiments. In the early stages, due to the limitation of computers' computing power, metamodels were developed to approximate expensive simulation models in order to reduce computational burden. Studies developing surrogate models and comparing the performance of surrogate models and simulation models were conducted, which include approximating delays caused by a single queue in waterway (Ramanathan and Schonfeld, 1994), approximating delays through series of waterway queues (Dai and Schonfeld, 1998) and approximating delays caused by a queuing network (Zhu et al., 1999).

In terms of optimization, SBO is most frequently applied for the calibration of simulation models, while a few other studies have utilized SBO to suggest transportation planning or operations strategies.

The dynamic traffic assignment (DTA) model calibration of O-D flows and all other simulation parameters is formulated as a large-scale iterative simulation-based optimization problem, which can be solved with several alternative approaches, such as Bayesian method, Stable Noisy Optimization by Branch and FIT (SNOBFIT), Box-Complex, Simultaneous Perturbation Stochastic Approximation (SPSA), Finite Differences Stochastic Approximation (FDSA) (Vaze et al., 2009; Sundaram et al., 2011; Flötteröd et al;, 2011; Omrani and Kattan, 2012). The performance of SPSA for the calibration of large-scale traffic simulation models has also been demonstrated.

For example, Balakrishna et al. (2007a) adapted the systematic traffic simulation model calibration methodology for the simultaneous calibration of all demand and supply models within a microscopic traffic simulation model using aggregate, time-varying traffic measurements; Balakrishna et al. (2007b) presented a systematic offline DTA calibration methodology that estimated all demand-and-supply inputs and parameters simultaneously.

Efforts have also been made to support decision making on various transportation planning and operational policies. Problems investigated include selection of a charging cordon in a general traffic network using Genetic Algorithm (GA) (Sumalee, 2004), developing a decision support tool for mitigating traffic congestion (Melouk et al., 2010), determining the traffic light signal timings for a single intersection with stochastic approximation (Fu and Howell, 2003), optimizing regional signal timing strategies using surrogate modeling (Osorio, 2010), optimizing coordinated, area-wide traffic signal control considering drivers re-routing behaviors using a metaheuristic model (Teklu et al., 2007), jointly optimizing traffic control and transit priority settings with GA (Stevanovic et al., 2008), optimizing waterway transportation investment with SPSA (Ting and Schonfeld, 1998), selection of interdependent transportation projects considering cost uncertainty and budget constraint with GA (Tao and Schonfeld, 2005), selection and scheduling of interdependent transportation projects with island models considering cost uncertainty (Tao and Schonfeld, 2007) or not (Tao and Schonfeld, 2006), scheduling of interrelated waterway projects with single budget constraint (Wang and Schonfeld, 2005), constraints on project multiplicity and precedence (Wang and Schonfeld, 2008)

or even more complicated constraints on project multiplicity, precedence and regional budget (Wang and Schonfeld, 2012) using GA, scheduling of waterway maintenance projects with constraints on budget and lock condition using GA (Wang et al., 2009), and to develop a heuristic approach for system wide highway project selection based on total benefit maximization under budget uncertainty (Li et al., 2010). Furthermore, to enhance the efficiency of simulation based optimization with heuristic methods, Yang (2010) developed a hybrid approach combining simulation and analytic methods along with parallel computing techniques. A first stage coarse search was conducted based on the analytic model, and in the second stage, a refined search based on simulation model was then performed inside the promising region provided by the first stage.

The most widely used SBO approach is heuristic methods (i.e. GA and simulated annealing), which can be conveniently applied to deal with different problems, while they are not very efficient in terms of searching for the optima. In a few other studies, stochastic approximation and surrogate modelling methods are investigated and applied.

Besides being directly utilized for solving optimization problems, surrogate-based methods are becoming popular and have been applied for other types of problems in the transportation domain in recent years. This type of methods is gaining attention mainly due to its ability of approximating the global trend of the objective function and power of significantly reducing the number of simulation runs. In particular, Ciuffo et al. (2011) utilized Kriging metamodeling to verify different micro-simulation calibration methods. Sensitivity analysis of traffic simulation

models have also been benefited from the application of surrogate models (Retherford and McDonald, 2011; Ciuffo et al., 2013).

2.4 Summary

Simulations have shown their advantages in representing the realism and various detailed aspects of the transportation system, and simulation-based optimization has shown its potential in supporting transportation planning and operations decision makings. Although various SBO methods have been developed for decades, the application of SBO in the transportation domain is still very few. According to the "no free lunch" theorems (Wolpert and MacReady, 1997; Ciuffo and Punzo, 2013) in the optimization community, there is no algorithm that outperforms all the others over the entire domain of problems, which means that the choice of the most appropriate algorithm depends on the features of the specific problem.

As transportation simulations take account of the interactions between complex travel demand patterns and network supply, it is always time consuming to evaluate the system performance. It can take up to hours to days for a single run of a regional network simulation. Thus, simulation-based optimization methods that require fewer objective function evaluations are desired. Meanwhile, due to the stochastic nature of transportation simulations, method that can deal with the simulation noise with fewer replications at design points is preferred. Although online simulation, as well as real-time traffic control, has appeared along with the

development of technology, most of transportation planning and operations decisions are still made offline due to the constraint of computational power. In this case, optimization methods that can take advantage of distributed computing are favored. Moreover, transportation simulation models are usually very complex due to the consideration of the interactions among travelers and between travelers and the network on all dimensions (e.g. departure time, mode, route, destination, etc.). The underlying models for different simulators are usually different. To reduce the potential cost of transferring the developed methods from one simulator to another, methods that do not require specific knowledge about the underlying structure of the simulation is of interest. Therefore, we favor derivative-free methods for the optimization problems based on transportation simulation.

Overall, according to all these features associated with transportation simulations, surrogate-based methods are chosen as the focus of this dissertation. In addition of applying existing surrogate-based methods, improvements will be made to better adapt the methods to specific problems and enhance the performance of the optimization procedure.

Chapter 3: Surrogate-Based Optimization Procedures and Application

3.1 Framework

A framework of surrogate-based optimization procedures using transportation simulations is illustrated in Figure 3-1.
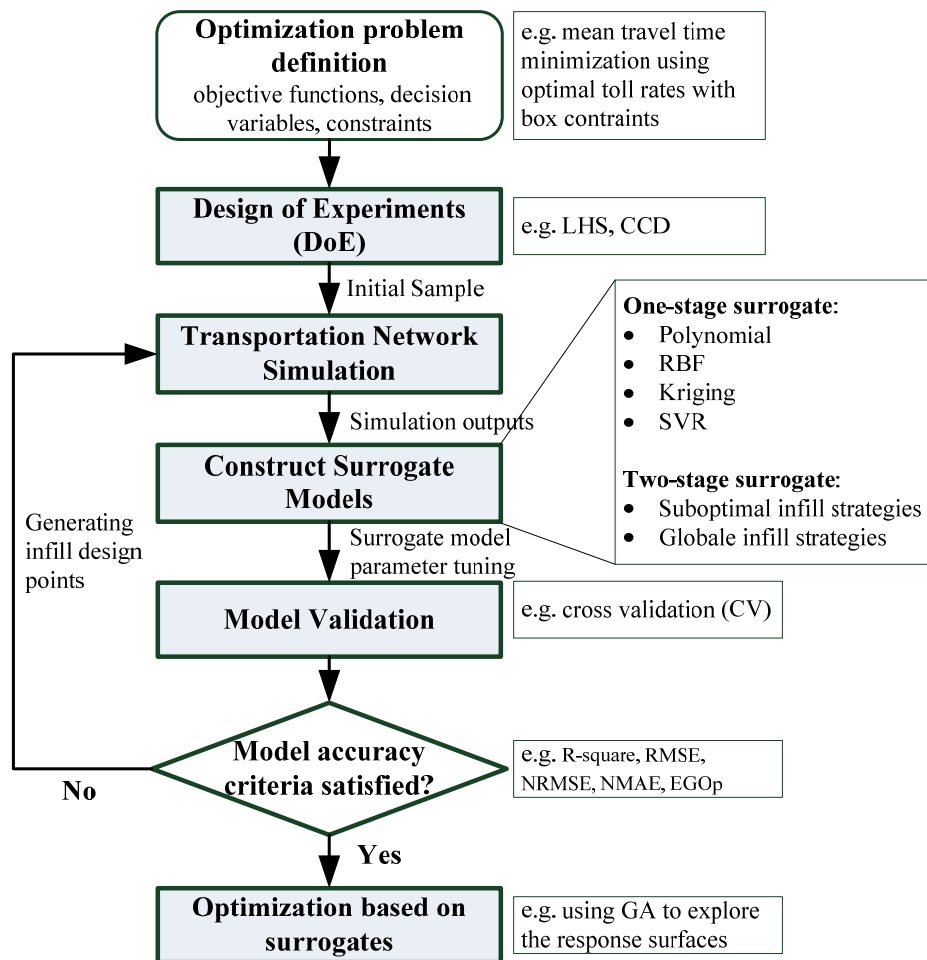
Figure 3-1: Transportation Simulation-based Optimization Procedure using the Surrogate Methodology.

The first step is to define the optimization problem. Given the objective functions that can be evaluated by simulation outputs, we generate an initial number of toll charges through a design of experiments (DoE). The DoE method used in the current study is Latin Hypercube Sampling (LHS), which would be introduced in the next subsection.

The second step is to select a transportation simulator for the specific optimization problem. Run the initial set of design points using the chosen simulator. Based on the simulation outputs, we evaluate the objective functions. Considering the generally large number of DoE cases (that aim to obtain a more accurate surrogate), the analysis could be computationally expensive. Simulation outputs of a transportation network usually involve random noises, so it is better to run each sampling point for several repetitions and estimate the mean value of objective function evaluations, if the computational burden is affordable.

The third step is to construct a response surface using surrogate models. In this study, first we adopt one-stage surrogate models, i.e. quadratic polynomial function, Gaussian RBF, ordinary Kriging and SVR. To find new design points based on the initial samples, we then consider two-stage surrogate models by infilling points to the initial set using criteria such as the probability of improvement and the expectation of improvement across the response surface. Due to simulation random errors, we may also incorporate the surrogate models that are capable of dealing with noises. Once all of the cases are analyzed, proper parameters of the surrogate models can be determined. This step is regarded as the most important component in the whole framework.

The fourth step is to assess and validate the assumed surrogate models by comparing an additional test set of objective function data with values estimated by the surrogates at points corresponding to the variables at which the independent objective function values are calculated. On the basis of the error observed with the validation dataset, the accuracy of each surrogate model is checked using certain criteria such as correlation coefficient or coefficient of determination to determine whether the initially assumed surrogate model is appropriate. If it is, the best surrogate model will be employed to explore the optimal solutions. If the evidence shows that a certain surrogate model does not achieve the required predictive performance based on the current test dataset, a proper way is to recall the two-stage surrogate models to generate infill points and run transportation simulations for new points until the accuracy criteria are reached.

Finally, we find the optimal solution using the optimized surrogate models. Though the estimated response surface may be too complex to explore its global optima using analytical techniques such as gradient descent method, we can still apply a heuristic approach, e.g. GA, to seek the global optima for the estimated response surface. The computational costs of this tuning process can be neglected compared to the burden of transportation simulations.

All key components in the framework are highlighted in shadow boxes in Figure 3-1. The details of three of the main components: DoE, surrogate model construction, and infill strategies, will be further explained in the following subsection.

3.2 Optimization Procedures

This section explains the technical steps that are necessary to apply the surrogate models to simulation-based optimization problems.

3.2.1 Design of Experiments

A space filling DoE is useful when only a few runs of simulation can be afforded within the computational budget. In the current study, LHS is employed to generate initial samples to fit surrogate models. Each design variable is stratified into an equal number of intervals according to the LHS setting. Being different from classic designs such as 2^k or 3^k fractional factorial designs and central composite designs (CCD) in (Montgomery, 2008), where each dimension of the decision variable is split into a relatively large number of equal-size bins, in which subsamples are uniformly generated. The LHS is advantageous in that the mapping of high dimensional design inputs into each dimension is uniformly distributed without overlap. Thus, such a property makes LHS one of the space-filling types of DoE.

In this paper, we recall the maximin design defined by Forrester et al. (2008). The ranking criterion function proposed by Morris and Mitchell (1995) is

$$\mathbf{X}^* = \arg \min_{\mathbf{X}} \phi_p(\mathbf{X}) = \arg \min_{\mathbf{X}} \left(\sum_{j=1}^{m} J_j d_j^{-p} \right)^{1/p} \qquad (3.1)$$

where $d = \| \mathbf{x}^{(i)} - \mathbf{x}^{(j)} \| = \sqrt{(\mathbf{x}^{(i)} - \mathbf{x}^{(j)})^{\mathrm{T}} (\mathbf{x}^{(i)} - \mathbf{x}^{(j)})}$, values of m distances are sorted in an ascending order, i.e. $d_1 \le d_2 \le, ..., \le d_m$. Let $J_1, J_2, ..., J_m$ be the number of $d_1, d_2, ..., d_m$, respectively. To illustrate the concept of LHS, we generate 100 different

2-dimensional plans. Each one contains 20 points. Then we calculate the ϕ_p values of

100 different plans. Figure 3-2 shows the evolutionary process of the sampling plan

space filling values, two of which are zoomed in and compared. The LHS plan with a

lower value of ϕ_p distributes more uniformly in the feasible domain.
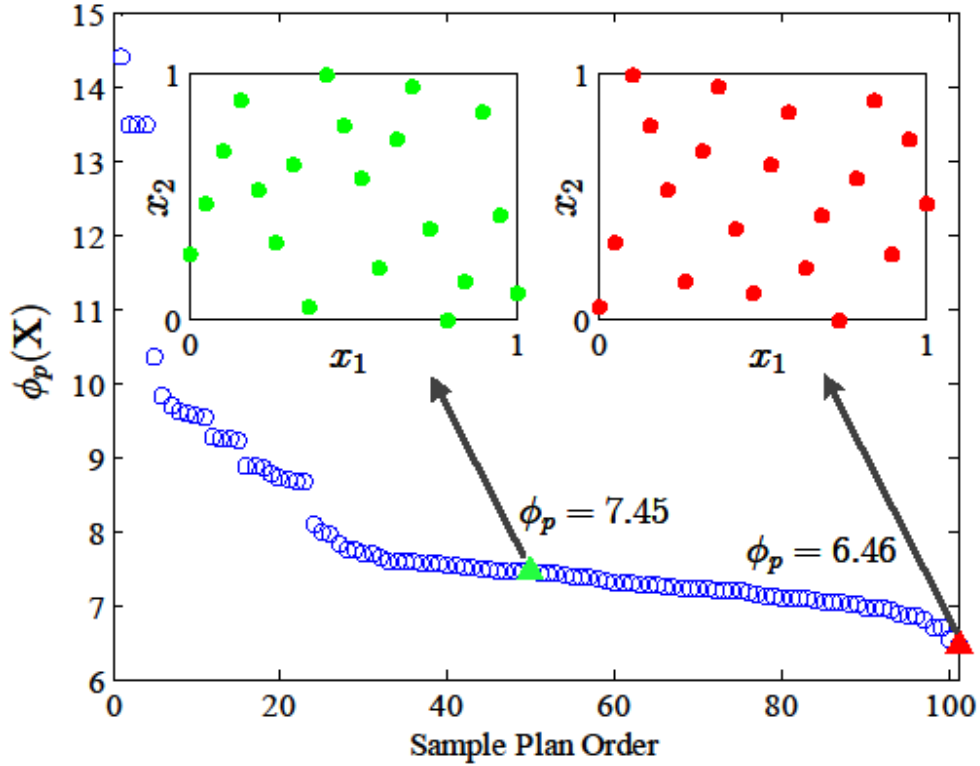
Figure 3-2: An Illustration of 100 LHS DoEs Based on the Spacing Filling Criterion $\phi_p(\mathbf{X})$.

3.2.2 Surrogate Model Construction

Surrogate models serve as the approximation of the true response surface of

expensive-to-evaluate objective functions. As the closed form of the surrogate models

is available, great computational cost can be saved when searching for the optimal

solution. Common forms of surrogate models in the literature are introduced in this section.

- *Quadratic Polynomial Function*

Using the DoE generation approach introduced above, we have an initial set of sampling plan $\mathbf{X} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}]^{\mathrm{T}}$ and responses $\mathbf{y} = [y^{(1)}, y^{(2)}, \ldots, y^{(n)}]^{\mathrm{T}}$. The most commonly used surrogate model is the quadratic polynomial function given by

$$\hat{f}(\mathbf{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j} \sum_j \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2, \quad \mathbf{x} \in \mathbb{R}^k \tag{3.2}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_k]^{\mathrm{T}}$ is a k–dimensional point to be predicted, $\hat{f}(\mathbf{x})$ is the estimate of the real objective function $f(\mathbf{x})$, β_0 is the intercept, β_i s are the linear coefficients, β_{ij} s are the coefficients of interaction terms, β_{ii} s are the quadratic coefficients.

- *Radial Basis Function (RBF)*

Compared with lower order polynomial function, RBF surrogate models can obtain better approximations to true objective functions of high nonlinearity. RBF uses the basis function $\varphi(r)$ that only depends on the radial distance r between $\mathbf{x}$ and each sample point $\mathbf{x}^{(i)}$. It assumes that the correlation of arbitrary two sample points depends only on the distance (e.g. Euclid distance) in the decision variable space. We seek a RBF approximation to $\hat{f}$ in the form

$$\hat{f}(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \boldsymbol{\varphi} = \sum_{i=1}^{n} w_i \varphi(\| \mathbf{x} - \mathbf{x}^{(i)} \|) \tag{3.3}$$

where $\mathbf{w}$ is the vector of weighted coefficients of RBF vector $\boldsymbol{\varphi}$, and $\| \mathbf{x} - \mathbf{x}^{(i)} \|$ is the Euclidean norm.

39

A Gaussian basis function is used in this paper, i.e.

$$\varphi(\| \mathbf{x} - \mathbf{x}^{(i)} \|) = \exp\left(-c \| \mathbf{x} - \mathbf{x}^{(i)} \|^2\right), \quad c > 0 \tag{3.4}$$

where c is the shape parameter that can be determined by tuning the minimization of a cross-validation (CV) error estimate in the optimization step. It is worthy to note that c can be various if we define different radial basis functions (RBF network), while normalization of input variables is not necessary for the Gaussian basis function because there are weight parameters for each function, so a universe c will be used for the Gaussian radial basis function in this paper.

The prediction at a new point is given by

$$\hat{f}(\mathbf{x}) = (\mathbf{\Phi}^{-1}\mathbf{y})^{\mathrm{T}}\boldsymbol{\varphi} \tag{3.5}$$

where $\mathbf{\Phi}$ denotes the so-called Gram matrix, each element of which is defined as $\mathbf{\Phi}_{i,j} = \varphi(\| \mathbf{x}^{(j)} - \mathbf{x}^{(i)} \|)$.

The prediction error at any $\mathbf{x}$ in the design space is given by (Gibbs, 1997)

$$\hat{s}^2(\mathbf{x}) = 1 - \boldsymbol{\varphi}^{\mathrm{T}}\mathbf{\Phi}^{-1}\boldsymbol{\varphi} \tag{3.6}$$

- *Kriging Method*

The Kriging method predicts a response by summarizing a linear model and a high frequency variation component that represents fluctuations around the trend. In this study, we will consider the ordinary Kriging model

$$f(\mathbf{x}) = \mu + \varepsilon(\mathbf{x}), \quad \mathrm{E}[\varepsilon] = 0 \tag{3.7}$$

where μ is the mean of the objective function, and ε is the estimation error with a covariance of $\mathrm{Cov}[\varepsilon(\mathbf{x}^{(i)}), \varepsilon(\mathbf{x}^{(j)})] = \sigma^2\psi(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$, σ^2 is the variance, $\psi(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ is the Kriging basis function with the following correlation form

40

$$\psi(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \exp\left(-\sum_{l=1}^{k} \theta_l (x_l^{(i)} - x_l^{(j)})^2\right) \tag{3.8}$$

where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_k]^{\mathrm{T}}$ is a vector of scaling coefficients that allow different widths of the basis function for each dimension of the k-dimensional $\mathbf{x}$ decision variable. The element of correlation matrix based on all the observed data is $\boldsymbol{\Psi}_{i,j} = \psi(\mathbf{x}^{(i)} - \mathbf{x}^{(j)})$.

Suppose all observed data are jointly Gaussian distributed, the likelihood function can be formulated as

$$L(\mathbf{y} \mid \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2} |\boldsymbol{\Psi}|^{1/2}} \exp\left[-\frac{(\mathbf{y} - \mathbf{1}\mu)^{\mathrm{T}} \boldsymbol{\Psi}^{-1} (\mathbf{y} - \mathbf{1}\mu)}{2\sigma^2}\right] \tag{3.9}$$

where $|\boldsymbol{\Psi}|$ is the determinant of $\boldsymbol{\Psi}$, and $\mathbf{y}$ is a $n \times 1$ vector representing the observation of the n sample points. $\mathbf{1}$ is a $n \times 1$ unit column. By taking the derivatives of the log-likelihood function with respect to μ and σ^2, and set them to be zero, the maximum likelihood estimates (MLEs) of μ and σ^2 are then

$$\hat{\mu} = \frac{\mathbf{1}^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \mathbf{y}}{\mathbf{1}^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \mathbf{1}} \tag{3.10}$$

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu})^{\mathrm{T}} \boldsymbol{\Psi}^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu})}{n} \tag{3.11}$$

Substituting $\hat{\mu}$ and $\hat{\sigma}^2$ into Equation 3.9, the vector of scaling coefficients $\boldsymbol{\theta}$ can be tuned through maximizing $L(\mathbf{y} \mid \hat{\mu}, \hat{\sigma}^2)$. To make predictions at a new point $\mathbf{x}^*$ using the Kriging model, suppose $f(\mathbf{x}^*)$ is the supposed function value at $\mathbf{x}^*$. We can add this new point to the observed dataset, and calculate the maximum likelihood estimation (MLE) of the mean value and variance at the new point $\mathbf{x}^*$ based on Gaussian Process (Rasmussen and Williams, 2006).

$$\hat{f}(\mathbf{x}) = \hat{\mu} + \mathbf{\psi}^{\mathrm{T}} \mathbf{\Psi}^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu}) \qquad (3.12)$$

$$\hat{s}^2(\mathbf{x}) = \hat{\sigma}^2 \left(1 - \mathbf{\psi}^{\mathrm{T}} \mathbf{\Psi}^{-1} \mathbf{\psi} + \frac{1 - \mathbf{1}^{\mathrm{T}} \mathbf{\Psi}^{-1} \mathbf{\psi}}{\mathbf{1}^{\mathrm{T}} \mathbf{\Psi}^{-1} \mathbf{1}} \right) \qquad (3.13)$$

where $\mathbf{\psi}$ is the vector of the correlation between the new point $\mathbf{x}^*$ and all the sample points, expressed as $\mathbf{\psi} = [\psi(\mathbf{x}, \mathbf{x}^{(1)}), \dots \psi(\mathbf{x}, \mathbf{x}^{(n)})]^{\mathrm{T}}$.

The aforementioned ordinary Kriging model is noise free. So the predictions at sampled points are exactly the same as observations, which may be biased when simulation noise is taken into account. As a consequence, the surrogate response surface may perform over fitting features because the estimated response surface needs to pass all sampled points in the ordinary Kriging model. A model which has been overfit will generally have poor predictive performance, as it can exaggerate minor fluctuations in the data. To get rid of this problem, a regularization constant is added into the correlation matrix to filter noise. The regressing Kriging model can be used by adding the error estimation of the observed data to the diagonal of the correlation matrix, so that the new matrix is $\tilde{\mathbf{\Psi}} = \mathbf{\Psi} + \lambda \mathbf{I}$, where λ is the regulation constant, $\mathbf{I}$ is identity matrix. λ can also be determined through MLE. Using the regressing Kriging model, the predicted mean value and estimation variance a the new point $\mathbf{x}^*$ are given by

$$\hat{f}_r(\mathbf{x}^*) = \hat{\mu}_r + \mathbf{\psi}^{\mathrm{T}} \tilde{\mathbf{\Psi}}^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu}) \qquad (3.14)$$

$$\hat{s}_r^2(\mathbf{x}^*) = \hat{\sigma}_r^2 \left(1 + \lambda - \mathbf{\psi}^{\mathrm{T}} \tilde{\mathbf{\Psi}}^{-1} \mathbf{\psi} + \frac{1 - \mathbf{1}^{\mathrm{T}} \tilde{\mathbf{\Psi}}^{-1} \mathbf{\psi}}{\mathbf{1}^{\mathrm{T}} \tilde{\mathbf{\Psi}}^{-1} \mathbf{1}} \right) \qquad (3.15)$$

And the MLEs of the mean and variance of the Gaussian Process are given by

$$\hat{\mu}_r = \frac{\mathbf{1}^{\mathrm{T}} \tilde{\boldsymbol{\Psi}}^{-1} \mathbf{y}}{\mathbf{1}^{\mathrm{T}} \tilde{\boldsymbol{\Psi}}^{-1} \mathbf{1}} \tag{3.16}$$

$$\hat{\sigma}_r^2 = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu}_r)^{\mathrm{T}} \tilde{\boldsymbol{\Psi}}^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu}_r)}{n} \tag{3.17}$$

- *Support Vector Regression (SVR)*

SVR is one of the most important applications of support vector machine (SVM). An overview of the basic ideas underlying SVR for regression and function estimation has been given in (Smola and Schölkopf, 2004). The key attribute of SVR is that it specifies and calculates the so-called ε-margin within which the sample data errors are accepted without impacts on the surrogate prediction. The prediction is determined entirely by the support vectors that lie on or outside the ε-margin (Forrester et al., 2008). In the ε-SVR, the goal is to find a surrogate that has the least ε deviation from the observations for all the training dataset, and at the same time is as flat as possible. SVR is a powerful tool for prediction given large, high-dimensional datasets. The parameter tuning time is longer than other surrogate methods due to the presence of the quadratic programming problem during the computation process. However, the additional computation time is marginal compared to the time spent on simulation. SVR is selected as an alternative surrogate-based optimization method in this study and compared with other aforementioned surrogate models. The technical details of the SVR method would be introduced later in Chapter 4.

3.2.3 Infill Strategies

To enhance the accuracy of surrogate models based on initial samples, further objective function evaluations based on certain infill or update strategies are required. This section incorporates the suboptimal exploration strategy that induces local optimization and the global exploration strategy that is promising to locate the global optimum.

- *Suboptimal infill strategy*

The local optima search strategy can be achieved by exploration over the surrogate surface estimated using the aforementioned surrogate models. In this study, we use GA to explore $\hat{f}(\mathbf{x})$ and seek its global optima. The update point is given by

$$\mathbf{x}_{\text{update}} \in \underset{\mathbf{x}_{\min} \leq \mathbf{x} \leq \mathbf{x}_{\max}}{\arg\min} \hat{f}(\mathbf{x}) \tag{3.18}$$

Then the simulation output at $\mathbf{x}_{\text{update}}$ will be evaluated by an extra simulation run, i.e. y_{update}. The infill strategy will be terminated when the maximal number of simulation runs is reached or the Euclidian norm of two adjacent update points is smaller than a predefined tolerance.

- *Global optimal infill strategy*

Global optimization can be classified into deterministic and stochastic methods. The former one generates a deterministic sequence of points converging to a globally optimal solution. Transportation simulation-based optimization problem may not belong to deterministic category because various sources of uncertainties lead to stochastic simulation outputs, e.g. random seeds in trip generation, probabilistic route choice behaviors of travelers, and DTA. The latter one randomly generates feasible

44

updating points to infill the initial samples using a number of heuristic algorithms for the optimal parameter tuning (Kim and Adeli, 2001; Baraldi et al., 2011; Chabuk et al., 2012; Sarma and Adeli, 2001; Song et al., 2013).

To obtain the global optima for expensive-to-evaluate functions, a series of two-stage procedures can be incorporated (the first stage includes DoE and surrogate model construction). The second stage conducts the exploitation process to locate promising regions. The global optimization has been investigated in quite a few existing studies. Jones et al. (1998) proposed the Efficient Global Optimization (EGO) based on Kriging basis functions, and applied the expected improvement of the surrogate to select new points. To handle noisy objective functions, Huang et al. (2006) provided the Sequential Kriging Optimization (SKO) as an extension of the EGO algorithm. Villemonteix et al. (2009) proposed the Informational Approach to Global Optimization (IAGO) that selects the infill point based on the entropy minimization. Jakobsson et al. (2010) proposed an RBF-based surrogate model for global optimization of expensive and noisy black box functions, whereas updating infill points minimize the total model uncertainty weighting. More detailed discussions on the exploration and exploitation process can be found in (Jones, 2001; Forrester and Keane, 2009; Kleijnen, 2009).

The two most common methods use estimated standard deviation information to select an infill sample with the maximum probability of improvement (PI) or expected improvement (EI), which are given by

$$\text{PI}(\mathbf{x}) = \frac{1}{\sqrt{2\pi}\hat{s}(\mathbf{x})} \int_{-\infty}^{y_{\min}} \exp\left(-\frac{(u-\hat{f}(\mathbf{x}))^2}{2\hat{s}(\mathbf{x})^2}\right) \mathrm{d}u \qquad (3.19)$$

45

$$\mathrm{EI}(\mathbf{x}) = \frac{1}{\sqrt{2\pi}\hat{s}(\mathbf{x})} \int_{-\infty}^{y_{\min}} (y_{\min} - u) \exp\left(-\frac{(u - \hat{f}(\mathbf{x}))^2}{2\hat{s}(\mathbf{x})^2}\right) du \qquad (3.20)$$

where $\mathrm{PI}(\mathbf{x})$ and $\mathrm{EI}(\mathbf{x})$ are the PI and EI estimations at the point $\mathbf{x}$, $y_{\min}$ denotes the smallest value of all outputs in the training dataset.

3.2.4 Summary

Table 3-1 shows the twelve models we investigate in the current study. Method 1 (M1) is the quadratic polynomial function based surrogate, which only recalls the 2nd-order polynomial functions with interaction terms. Methods 2, 3 and 4 are both one-stage models that estimate the response surface only using the initial samples. Methods 5 through 8 are suboptimal two-stage approaches. M9 and M10 enhance the Kriging method with the global optimal infill strategy using PI and EI maximization, respectively. To deal with noisy data, M11 is the one stage regressing Kriging model. The main difference between Kriging with and without noisy errors is that the estimated response surface would pass through the known points in M3, while M11 allows some bias to the known points to obtain a much smoother response surface. Finally, M12 is the EI infill re-interpolating Kriging method. The 12 methods will be tested and compared using a small transportation network with additive toll links in the following subsection. The selected methods from the numerical test are then applied to a real transportation network in the State of Maryland.

Abbr.	Methods	One-stage surrogate	Infill surrogate	Global optimum	Simulation noise
M1	Quadratic polynomial	√	×	×	×
M2	Gaussian RBF	√	×	×	×
M3	Kriging	√	×	×	×
M4	SVR	√	×	×	×
M5	Suboptimal updating quadratic polynomial	×	√	×	×
M6	Suboptimal updating Gaussian RBF	×	√	×	×
M7	Suboptimal updating Kriging	×	√	×	×
M8	Suboptimal updating SVR	×	√	×	×
M9	Probability of improvement infill Kriging	×	√	√	√
M10	Expected improvement infill Kriging	×	√	√	√
M11	Regressing Kriging	√	×	×	√
M12	Expected improvement infill Re-interpolating Kriging	×	√	√	√

Table 3-1: Characteristics of Surrogate Models.

3.3 Numerical Test

This section tests the surrogate models in Table 3-1 using a second-best social optima additive highway pricing with a fixed demand for a small network. The user equilibrium (UE) assignment is chosen because the true objective function can be exactly known through an analytical derivation, so we can validate the estimated response surfaces with the true response surface. Though UE and simulation are quite different in objective function evaluations, e.g. mean travel time of all travelers, the input-output mapping can be estimated and validated through surrogate models. The test based on the small network with UE assignment can provide insights on the features of different surrogate models and help to identify the most capable method that can be used to model the input-output mapping in a larger network.

The link-based pricing scheme is investigated as the second-best toll charging in a small road network, where tolls are charged only on a subset of selected links,

which can be categorized as a Mathematical Program with Equilibrium Constraints (MPEC) (Yang and Huang, 2005). The second-best road pricing problem in this paper is to choose a set of optimal toll charges to minimize the total travel time (or the average travel time due to fixed demand). The bi-level mathematical program with equilibrium constraints can be formulated. The upper level model is

$$\min_{\mathbf{z}} \ F(\mathbf{z}, \mathbf{q}^*) = \sum_{a \in A} t_a(q_a^*) q_a^* \qquad (3.21a)$$

$$\text{s.t.} \quad \mathbf{q}^* = \arg\min_{\mathbf{q}} \sum_{a \in A} \int_0^{q_a} c_a(q, \mathbf{z}) \mathrm{d}q \qquad (3.21b)$$

$$\mathbf{z}_{\min} \leq \mathbf{z} \leq \mathbf{z}_{\max} \qquad (3.21c)$$

$$q_a = \sum_{r \in R} \sum_{s \in S} \sum_{p \in P_{rs}} f_p^{rs} \delta_{ap}^{rs}, \quad a \in A \qquad (3.21d)$$

$$q_k = \sum_{r \in R} \sum_{s \in S} \sum_{p \in P_{rs}} f_p^{rs} \delta_{kp}^{rs}, \quad k \in K \qquad (3.21e)$$

$$\sum_{p \in P_{rs}} f_p^{rs} = d_{rs}, \quad r \in R, s \in S \qquad (3.21f)$$

$$f_p^{rs} \geq 0, \quad r \in R, s \in S, p \in P_{rs} \qquad (3.21g)$$

where F is the total travel time function, $\mathbf{z} = [z_1, ..., z_k]^{\mathrm{T}}$ is the link toll vector, satisfying $k \in K$, $K \subseteq A$ is a subset of tolled links, A is the whole link set, $\mathbf{q}^* = [..., q_a^*, ...]^{\mathrm{T}}$ is the equilibrium link flow vector, q_a^* is the equilibrium flow of link a, satisfying $a \in A$, t_a is the average travel time of link a, constraints (3.21b) through (3.21g) are the conservation conditions, f_p^{rs} is the path flow of OD pair (r, s), δ_{ap}^{rs} is the 0-1 indicator, R and S are origin and destination sets.

The well-known Frank–Wolfe method can be used to solve the lower level programming problem of the traffic equilibrium model (Ramadurai and Ukkusuri, 2011; Szeto et al., 2011; Aziz and Ukkusuri, 2012; Unnikrishnan and Lin, 2012). The solution of the bi-level programming problem can be obtained by using the gap function approach solved by the augmented Lagrangian algorithm (Yang and Huang, 2004; Meng and Wang, 2008).

Travelers are assumed to be homogeneous with identical values of time. From the perception of link-based cost, the generalized cost function c_a of link a can be expressed as follows

$$c_a(q_a, \mathbf{z}) = \begin{cases} z_k + \eta t_k(q_k) & a = k \in K \\ \eta t_a(q_a) & a \in A, a \notin K \end{cases} \tag{3.22}$$

where z_k and t_k are the toll charge and average travel time of link k, η is the value of time.
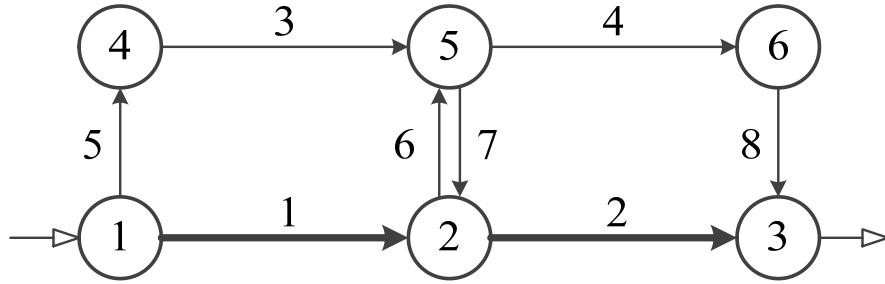
Figure 3-3: Numerical Network. (link 1 and 2 are the tolled links)

Consider a network depicted in Figure 3-3, consisting of 6 nodes and 8 links. Link 1 and 2 are road segments subject to toll charge. The BPR link performance function is applied

$$t_a(q_a) = t_a^0 \left[1.0 + \alpha \left(\frac{q_a}{C_a} \right)^{\beta} \right], \quad a \in A \tag{3.23}$$

49

where t_a^0 is the free-flow travel time, C_a is the link capacity, parameters are $\alpha = 0.15$ and $\beta = 4$, see Table 3-2.

Link	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
t_a^0	20	20	20	20	6	1	1	6
C_a	800	800	600	600	500	800	800	500
q_a^*	686	686	314	314	314	0	0	314
$t_a(q_a^*)$	21.6	21.6	20.2	20.2	6.1	1	1	6.1

Table 3-2: Input Data and Equilibrium Flow for a Small Road Network.
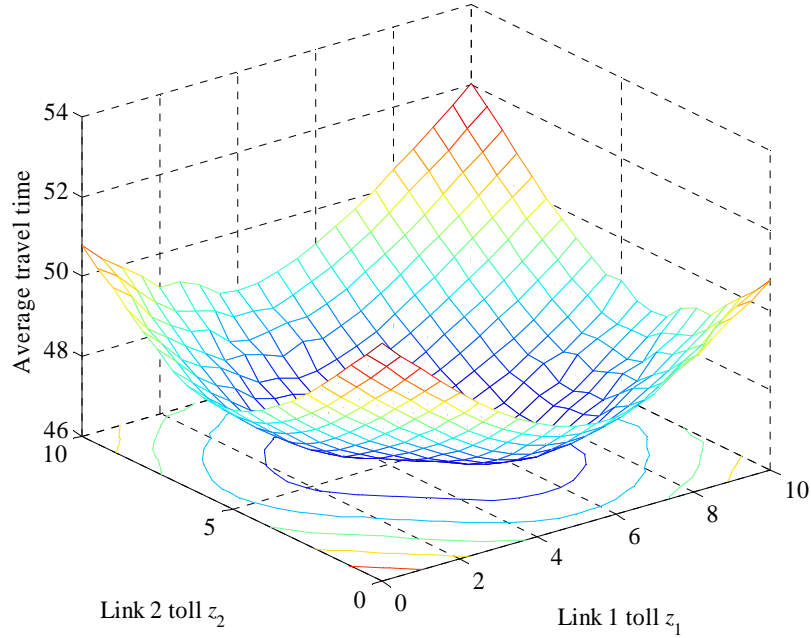
Figure 3-4: The True Response Function of $F(\mathbf{z}, \mathbf{q}^*)$.

There is only one O–D pair from node 1 to node 3 with a demand of $d = 1000$ (flow units), and the value of time is $\eta = 1$. Four paths from node 1 to node 3 using links are: 1–2, 1–6–4–8, 5–3–4–8, 5–3–7–2. One of the optimal toll charges is $\mathbf{z}^* = [5.555, 4.045]^T$. The minimized average travel time is $F_{min} = 46.22$. Optimal

path flows are $f_1^* = f_3^* = 686$ and $f_2^* = f_4^* = 0$. Figure 3-4 shows the true responses of the upper level objective function $\hat{F}(\mathbf{z}, \mathbf{q}^*)$ corresponding to z_1 and z_2. It is an interpolation surface based on a uniform 20×20 grid. The simulation random errors are because the UE link flows may be not integers.

To compare the surrogate-based optimization approaches shown in Table 3-1, we generate 10, 20, 30 and 40 initial LHS samples for the five one-stage surrogate models M1, M2, M3, M4 and M11, respectively. Then we generate other initial 8, 15, 25 and 35 LHS points for the two-stage models M5, M6, M7, M8, M9, M10 and M12, then add 2, 5, 5 and 5 infill points to the initial samples, respectively. The surrogate models are then validated by a uniform 20×20 grid sample points. The overall performance of the surrogate models is evaluated using 6 accuracy measures:

(1) Root Mean Square Error (RMSE), which provides a global error measure over the entire design domain

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (f(\mathbf{x}^{(i)}) - \hat{f}(\mathbf{x}^{(i)}))^2} \qquad (3.24)$$

(2) Maximum Absolute Error (MAE), which is indicative of local deviations

$$\text{MAE} = \max_{1 \le i \le n} | f(\mathbf{x}^{(i)}) - \hat{f}(\mathbf{x}^{(i)}) | \qquad (3.25)$$

(3) Normalized Root Mean Squared Error (NRMSE)

$$\text{NRMSE} = \sqrt{\frac{\sum_{i=1}^{n} [f(\mathbf{x}^{(i)}) - \hat{f}(\mathbf{x}^{(i)})]^2}{\sum_{i=1}^{n} (f(\mathbf{x}^{(i)}))^2}} \qquad (3.26)$$

(4) Normalized Maximum Absolute Error (NMAE)

$$\text{NMAX} = \frac{\max_{1 \le i \le n} |f(\mathbf{x}^{(i)}) - \hat{f}(\mathbf{x}^{(i)})|}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (f(\mathbf{x}^{(i)}) - \overline{f})^2}}, \quad \overline{f} = \frac{1}{n} \sum_{i=1}^{n} \hat{f}(\mathbf{x}^{(i)}) \tag{3.27}$$

(5) Estimated Global Optimum (EGOp)

$$\text{EGOp} = \hat{f}(\hat{\mathbf{x}}^*) \tag{3.28}$$

where $\mathbf{x}^* = \arg\min f(\mathbf{x})$, $\hat{\mathbf{x}}^* = \arg\min \hat{f}(\mathbf{x})$.

(6) Pearson correlation coefficient (PCC)

$$r^2 = \left(\frac{N \sum f\hat{f} - \sum f \sum \hat{f}}{\sqrt{[N \sum f^2 - (\sum f)^2][N \sum \hat{f}^2 - (\sum \hat{f})^2]}} \right)^2 \tag{3.29}$$

where N is the number of independent set of objective function data to be compared, f denotes objective function values from the independent test set and $\hat{f}$ are the corresponding surrogate model estimations. If $r^2 = 1$, the surrogate is exactly predicting the test data, while $r^2 = 0$ indicates no correlation between the surrogate and the objective function.
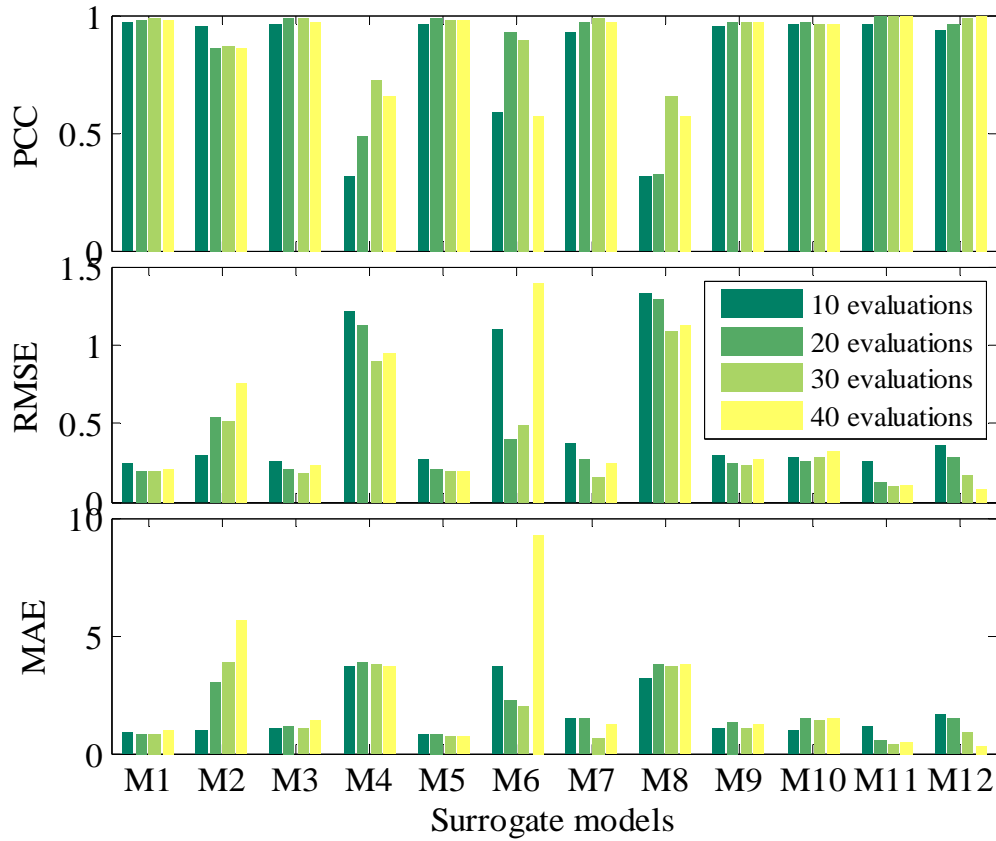
Figure 3-5: Validation of Surrogate Models.

Figure 3-5 shows the estimation errors by comparing the prediction values with true objective function for the uniform 20×20 grid. The Pearson correlation PCC, RMSE, MAE values are plotted for each surrogate model and different sample size. For all surrogate models except M2 and M6, the estimation accuracy would be higher when the sample size increases.

Table 3-3 shows the results of the 12 models in terms of five measures of effectiveness (MoEs) under the largest test sample size, i.e. 40 evaluations of the objective function. The minimum values of each column are highlighted in bold. Results show that the best model with the smallest errors of RMSE, MAX, NRMSE and NMAX is M11, and the second best model is M12. The smallest

EGOp = $\hat{F}(\hat{\mathbf{z}}^{*}, \mathbf{q}^{*})$ is 44.33 obtained by M6; however, its true response is 48.68 that

is larger than the estimated minimum. From other MoEs of M6, we can see its overall

prediction accuracy is poor. In the last column, seven models finally converge to the

global optima.

Method	RMSE	MAX	NRMSE	NMAX	EGOp	$f(\hat{\mathbf{x}}^{*})$
M1	0.21	0.95	0.44%	0.68	46.15	**46.22**
M2	0.76	5.62	1.57%	4.03	46.20	46.24
M3	0.24	1.44	0.50%	1.03	46.12	**46.22**
M4	0.94	3.67	1.95%	2.63	46.45	46.25
M5	0.19	0.71	0.40%	0.51	46.22	46.23
M6	1.30	7.21	2.69%	5.17	**44.33**	48.68
M7	0.24	1.20	0.50%	0.86	46.22	**46.22**
M8	1.13	3.76	2.34%	2.69	47.20	46.23
M9	0.36	1.41	0.75%	1.01	46.22	**46.22**
M10	0.23	1.09	0.48%	0.78	46.22	**46.22**
M11	**0.10**	**0.45**	**0.21%**	**0.32**	46.09	**46.22**
M12	0.16	0.57	0.33%	0.41	46.21	**46.22**

Table 3-3: The Estimation Accuracy Comparison of Surrogate Models, under 40
Times of Objective Function Evaluations.

The novelty in this work is the computational time savings. To demonstrate

how surrogate models can intelligently mimic simulation-based objective function

evaluation and reduce computational times, we compare the convergence and

efficiency of surrogate models with GA (the population size is 10, generations are 10,

other parameters are default value given by MATLAB R2010a GA Toolbox). Figure

3-6 quantifies the computational savings obtained from this method using the number

of objective function evaluations. We can see that the best surrogate model (M12) can find the global optima only using 10 evaluations.
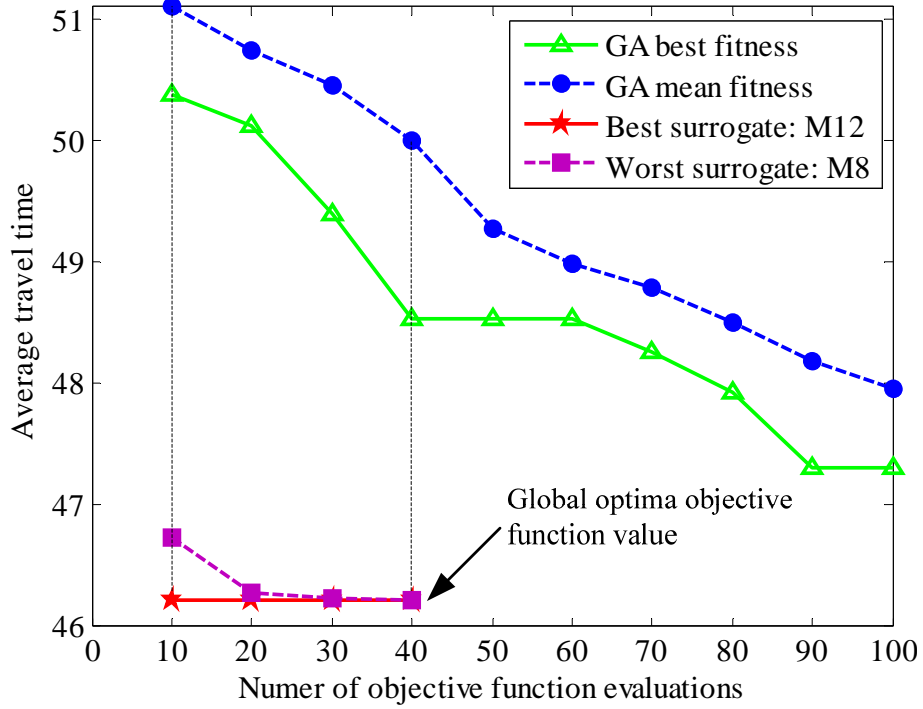
Figure 3-6: An Illustration of the Higher Computational Efficiency of the Surrogate Models compared to GA.

In summary, the three best one-state surrogate models are M1, M3 and M11 (M1 performs better mainly because the true response surface is not very complex as shown in Figure 3-4), and the best two-stage surrogate models is M12. In the following of this chapter, we will apply these four models in a real world transportation network.

3.4 Test Bed for Real World Application

In terms of application, this chapter would investigate the optimal dynamic pricing of toll facilities in a regional transportation network using the selected surrogate models in the previous section. In order to apply SBO for real world transportation planning and operations decisions, a road network as well as a well-developed simulator is required. This section will introduce these two components of our test bed for the real world application.

3.4.1 The ICC Road Network in Maryland

The Inter-County Connector (ICC) is probably the most significant and high-profile highway project in Maryland since the completion of the existing interstate freeway system several decades ago. It links existing and proposed development areas between the I-270/I-370 and I-95/US-1 corridors within central and eastern Montgomery County and northwestern Prince George's County.

The simulation model covers the central and eastern Montgomery County and the northwestern Prince George's County of the State of Maryland. Before the construction of ICC, there was no freeway connecting the areas lying northwest and northeast of the capital beltway. The traffic between these two areas usually travels through I-495, which contributes to the severe congestion on I-495 during peak hours. The ICC was constructed aiming at promoting development of the surrounding areas as well as alleviating congestion on I-495. The ICC is a toll facility with different toll rates for its five segments, and the toll rate for each segment is variable along time. Vehicles with E-ZPass, an electronic toll transponder, are charged directly when they

travel through ICC. If a vehicle without E-ZPass uses ICC, a $1 video processing fee

is added to the total price and a bill sent to the vehicle registration address.
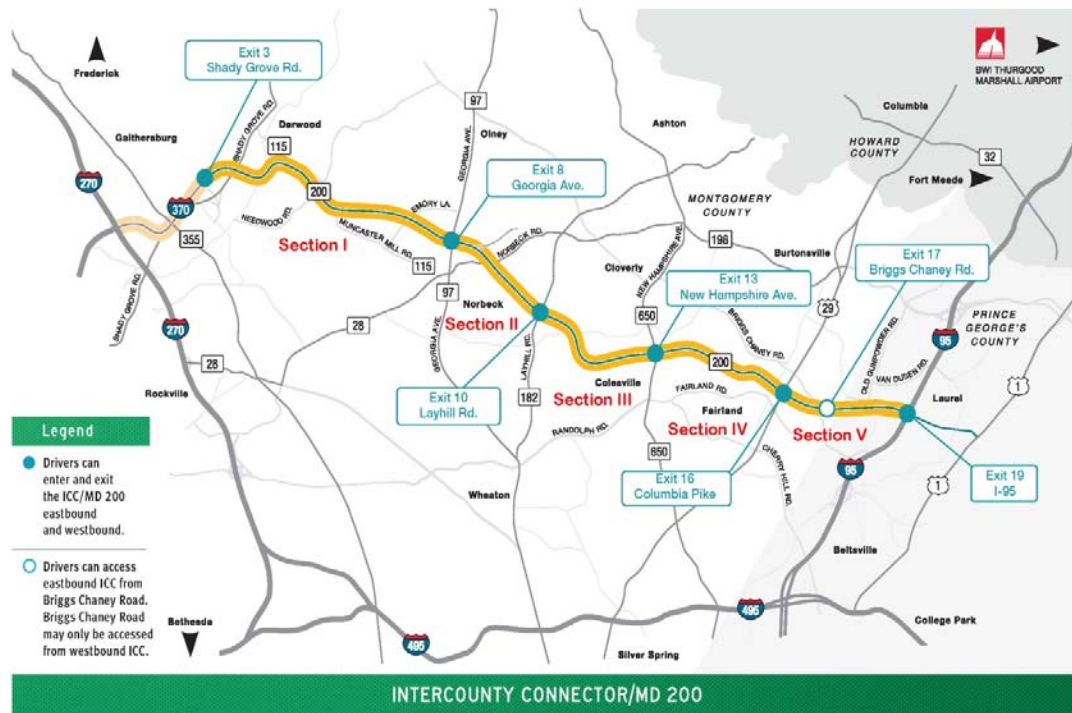
Figure 3-7: The Inter-County Connector (in thick line) and Regional Network.
(Source: http://www.mdta.maryland.gov/ICC/Toll_Rates.html)

To test the effectiveness of applying a simulation-based optimization method

to improve the transportation system performance, a case study on optimizing the toll

scheme of the ICC in Maryland has been conducted. A simulation model for the

regional network is developed to evaluate the system performance. All freeways and

arterial roads within the region in Figure 3-7 are included in the transportation

network, which is relatively large with 201 TAZs, 1077 nodes and 2158 links. In our

case study, actuated signal timings were coded into DynusT for all intersections that

have signals applied in real world in the network. In a previous research on the

before-and-after study of the ICC road (Zhang et al., 2013) for the same regional

network, the dynamic travel demand has already been calibrated and validated with field data from detectors and floating cars. The simplified current pricing scheme for two-axle vehicles with E-ZPass during different time periods as well as the proposed limit of the toll rates is summarized in Table 3-4.

Parameter (unit)	Symbol	Baseline	Lower bound	Upper bound
Toll charge in peak period, Segment I ($) Between I-370 and MD 97	z_1	1.45	0	3
Toll charge in peak period, Segment II ($) Between MD 97 and MD 182	z_2	0.60	0	1.5
Toll charge in peak period, Segment III ($) Between MD 182 and MD 650	z_3	0.75	0	1.5
Toll charge in peak period, Segment IV ($) Between MD 650 and US 29	z_4	0.65	0	1.5
Toll charge in peak period, Segment V ($) Between US 29 and I-95	z_5	0.70	0	1.5
Off-peak / Peak toll charge ratio	τ	80%	0	100%

Table 3-4: Selected Design Parameters and Baseline Value.

3.4.2 Open Source DTA Simulator: DynusT

DTA models fill the gap between static travel forecasting models and microscopic traffic simulation models, and enable modeling traffic dynamics at a relatively large scale within a reasonable amount of time. In the DTA framework, UE condition is only applied to travelers departing at the same time between the same OD pairs. Time-dependent shortest paths for travelers are computed based on time-varying link travel times when they arrive at the various links along a route.

DynusT (Dynamic Urban Systems in Transportation) is a simulation-based DTA model, which adopts the dynamic interactions between the network supply and user demand. DynusT performs well regarding its computational efficiency. However, it is essentially a route choice model. Some important aspects of travel demand analysis such as trip generation, mode choice and departure time choice are not enabled in DynusT. Time-varying link travel time needed for DTA in DynusT is retrieved from the Anisotropic Mesoscopic Simulation (AMS) model (Chiu, et al., 2010b), which is a vehicle-based mesoscopic traffic simulation approach that explicitly considers the anisotropic property of traffic flow in the vehicle state update at each simulation step. DynusT applies a gap function vehicle-based (GFV) solution algorithm to solve the DTA problem (Chiu and Bustillos, 2009). For each iteration and each OD-departure time combination, the number of vehicles to be updated with a new path is dependent on the relative gap function value, and vehicles with longer travel time are prioritized for path updating. Compared with the widely used successive average method, GFV can avoid over adjustments of flow and thus lead to more consistent and robust assignment results. Meanwhile, DynusT adopts a method of isochronal vehicle assignment which divides analysis periods into epochs and sequentially performs vehicle assignment in each epoch (Nava and Chiu, 2012). This significantly improves the model scalability regardless of the total analysis period. In the newly released 2012 version, DynusT has been fully parallelized in simulation, time-dependent shortest path and assignment algorithms, and therefore boosts the computational speed dramatically. However, the current simulator doesn't address capacity drop due to congestion.

Although other models of both microscopic and mesoscopic traffic simulation are widely available, (e.g. DynaMIT, DYNASMART, and Dynameq for mesoscopic models; TransModeler, VISSIM and AIMSUN for microscopic models) and some of them may possess some desirable features, DynusT is selected in this study due to its advantage in computation time.

3.5 Application: Optimal Dynamic Pricing for Toll Roads

For the real world application, we focus on the decision making of a public highway operator (such as the government or public agency) on the dynamic pricing a toll road. We assume that most commuting demands do not change departure time or cancel trips due to toll charges. Therefore, fixed commuting demands are assumed for the regional road network for this study. DynusT is selected as the DTA and mesoscopic vehicle simulation tool to evaluate network performance given various link-additive highway pricing rates. The computation time needed to obtain a solution from this black-box function evaluated by DynusT can be considerably reduced by surrogate-based optimization models, which make noisy data processing and computation intensive global optimization feasible.

3.5.1 Background

The idea of charging the use of road originates from the economic theory that users should be aware of the costs they impose upon one another and pay for the

additional congestion they create. The well-known marginal cost pricing principle was first discussed in Beckmann (1967). Yang and Huang (1998) then further examined this classical principle in connection with the general deterministic network equilibrium problem. They found that the optimal tolls for the variable-demand network equilibrium problems took the same form as the marginal cost toll in terms of maximizing net economic benefits. However, the optimal toll would be different in the presence of queues due to limited road capacity. Optimal toll charge under the stochastic network equilibrium was investigated later in Yang (1999). For the problem of charging toll on only a subset of links in the network, the assumptions of fixed demand (Yang and Lam, 1996) and elastic demand (Yang and Bell, 1997) have been studied. More complex optimization problems with equity constraints and multiple user classes were also investigated (Yang and Zhang, 2002; Yang and Huang, 2004).

Although the optimization of the highway toll has been studied extensively in previous literatures, the network model used for traffic routing was always macro-level static user equilibrium models. This type of model suffers from several limitations, e.g. the dynamics of travel behavior along time cannot be captured, and the influence of micro or meso level operational improvements such as traffic signals and dynamic message signs can't be considered. On the other hand, dynamic network supply models can overcome these shortcomings, and provide more detailed information regarding the system performance. All these features make dynamic network supply models a relatively better approach to evaluate the performance of transportation system under different pricing strategies. Yet the drawbacks with

utilizing dynamic network supply model in optimizing road tolls should also be noticed. The DTA is usually coupled with simulation, which makes it impossible to retrieve a closed form of the function mapping toll strategies to the output performance of network. Meanwhile, the simulation of a medium or large scale network may take hours to days to converge. The computational burden incurred by simulations would be a big challenge for the optimization. Moreover, the random noise in simulation as well as DTA outputs cannot be neglected during the process of optimization.

Formulating the problem of toll road pricing optimization in the simulation-based optimization framework is advantageous over mathematical formulation in that it relaxes various unreasonable assumptions in mathematical functions and can better represent the traffic realism as well as the travelers' responses to policy stimuli. On the other hand, surrogate-based method clearly serves as a feasible and promising way to tackle all the challenges associated with simulation-based optimization problems. In the following sections, the application of surrogate-based optimization of toll road pricing with different objective functions would be presented.

3.5.2 Problem Setting and Formulation

As introduced in section 3.4.1, this real world application problem is formulated with 6 decision variables, which are the independent peak period toll rate for the five segments of the ICC and the uniform off-peak to peak toll charge ratio. To apply the simulation-based optimization, the first step is to generate an initial sample plan. Utilizing the LHS method, we obtain the optimized sample plan, i.e. $\mathbf{X}$,

including 64 initial LHS points plus three chosen inputs: the minimal toll plan $\mathbf{x} = \mathbf{x}_{\min} = \mathbf{0}$, the maximal toll plan $\mathbf{x} = \mathbf{x}_{\max}$ and the baseline inputs $\mathbf{x} = \mathbf{x}_{\text{baseline}}$. The initial sample points are listed in Table 3-5.

Sample	z_1 $	z_2 $	z_3 $	z_4 $	z_5 $	τ %	Sample	z_1 $	z_2 $	z_3 $	z_4 $	z_5 $	τ %
1	1.24	1.29	0.07	1.31	0.95	0%	33	2.24	0.36	0.86	0.93	0.24	52%
2	2.57	0.76	1.10	0.83	0.81	92%	34	0.95	1.40	0.81	0.14	0.71	56%
3	0.90	0.98	0.95	1.43	0.10	71%	35	1.71	0.57	0.10	0.12	0.29	86%
4	2.52	0.38	1.29	1.05	1.26	70%	36	0.38	1.45	1.02	0.43	1.21	75%
5	2.67	0.95	1.48	0.55	1.43	60%	37	1.14	1.21	0.93	0.74	1.00	46%
6	1.81	1.24	0.24	0.71	0.93	33%	38	0.19	0.17	0.05	1.12	0.98	38%
7	0.43	1.07	0.43	0.76	0.31	29%	39	0.57	0.64	1.50	1.07	0.88	41%
8	1.48	0.00	1.26	0.05	0.90	35%	40	1.52	1.00	0.29	1.14	0.40	21%
9	1.86	0.40	1.31	0.38	0.43	59%	41	2.33	0.12	1.43	0.33	1.38	89%
10	1.76	0.86	0.83	0.50	0.69	30%	42	0.10	0.88	0.17	0.95	1.14	17%
11	2.48	1.17	0.57	0.36	0.07	63%	43	1.38	0.48	0.71	0.21	1.29	90%
12	0.86	0.24	0.19	0.48	0.05	32%	44	2.10	0.93	0.40	0.90	0.74	95%
13	0.52	0.55	0.88	0.45	1.24	54%	45	0.24	0.05	1.33	1.10	0.17	65%
14	0.62	1.38	1.38	0.79	1.12	16%	46	1.33	0.60	0.74	1.45	0.67	62%
15	0.71	1.48	1.21	0.67	0.50	67%	47	2.38	0.14	1.45	1.19	0.45	81%
16	3.00	1.10	0.36	0.31	0.62	24%	48	1.57	0.90	0.76	1.33	1.45	19%
17	0.05	0.02	0.14	0.02	0.38	10%	49	0.00	0.50	0.67	1.40	0.26	57%
18	2.29	0.07	1.14	1.24	0.57	25%	50	0.76	0.45	0.79	0.81	0.55	40%
19	2.71	0.81	0.60	0.62	1.50	87%	51	1.43	0.43	0.38	0.98	1.02	68%
20	1.19	0.71	0.50	0.00	1.19	37%	52	1.95	1.05	1.07	1.00	1.48	84%
21	2.43	1.43	1.12	0.40	0.76	3%	53	1.00	1.12	0.00	0.07	0.21	14%
22	1.29	0.79	0.26	0.29	0.14	49%	54	0.33	0.19	1.24	0.86	0.86	94%
23	0.48	1.33	1.19	1.50	1.36	51%	55	2.05	1.50	0.45	1.38	0.83	44%
24	0.29	0.21	1.05	0.60	0.33	2%	56	2.00	1.26	0.64	0.10	1.31	48%
25	0.81	1.02	0.12	0.57	0.48	100%	57	0.14	1.36	1.36	0.19	0.00	97%
26	2.86	0.52	1.17	0.17	0.79	22%	58	1.05	1.31	0.69	0.26	1.05	8%
27	2.81	0.74	0.52	0.88	0.12	78%	59	0.67	0.83	0.33	1.48	1.07	76%
28	2.62	0.26	0.55	1.17	1.40	43%	60	2.90	0.67	0.48	1.29	0.02	13%
29	2.76	0.33	0.98	1.02	1.10	11%	61	2.19	0.31	0.21	1.21	0.36	79%
30	1.67	0.62	1.00	0.52	1.33	6%	62	1.90	1.14	0.90	0.64	0.52	83%
31	1.10	0.10	0.02	1.26	0.60	5%	63	2.95	0.69	1.40	1.36	0.19	27%
32	1.62	1.19	0.31	0.24	0.64	73%	64	2.14	0.29	0.62	0.69	1.17	98%
Lower bound	0.00	0.00	0.00	0.00	0.00	0%	Upper bound	3	1.5	1.5	1.5	1.5	1
Baseline	1.45	0.6	0.75	0.65	0.7	80%							

Table 3-5: Space-Filling Latin Hypercube of Parameters for DoE.

After the initial sample is generated, all the sample points should be evaluated through the simulator. The observation of these sample points would then be used for the constructing of surrogate models as well as the search for optimal solution.

From a public agency's perspective, the objective would be to minimize the total social costs of the whole network or maximize total social welfare, while if a road is privately operated, maximizing total toll revenue may be the main objective. In the current study, the objective is set to minimize the average travel time for all network users given fixed OD demands. The objective function can be formulated as

$$\min_{\mathbf{x} \in \mathbb{R}^k} \quad \mathrm{E}[f_{\mathrm{TT}}(\mathbf{z}, \tau)] = \mathrm{E}\left[\frac{\sum_{r \in R}\sum_{s \in S}(d^{rs}_{\mathrm{peak}} \cdot \mathrm{TT}^{rs}_{\mathrm{peak}}(\mathbf{z}) + d^{rs}_{\mathrm{off}} \cdot \mathrm{TT}^{rs}_{\mathrm{off}}(\mathbf{z}, \tau))}{\sum_{r \in R}\sum_{s \in S}(d^{rs}_{\mathrm{peak}} + d^{rs}_{\mathrm{off}})} \right] \tag{3.30a}$$

$$\text{s.t.} \quad \mathbf{z}_{\min} \leq \mathbf{z} \leq \mathbf{z}_{\max}, \quad \mathbf{z} \in \mathbb{R}^{k-1} \tag{3.30b}$$

$$0 \leq \tau \leq 1 \tag{3.30c}$$

$$\mathbf{x}^{\mathrm{T}} = [\mathbf{z}^{\mathrm{T}}, \tau], \quad \mathbf{x} \in \mathbb{R}^k \tag{3.30d}$$

where $f_{\mathrm{TT}}(\mathbf{z}, \tau)$ is the stochastic average travel time function of the network, we minimize its expectation given the same input $\mathbf{z}$ and τ to eliminate random simulation errors. The decision vector $\mathbf{x}$ includes toll rates $\mathbf{z}$ of each toll road segment and the ratio τ of off-peak-hour toll rates to the peak-hour values, so $\mathbf{x}$ is a k-dimensional decision variable vector; the origin and destination sets are $r \in R$ and $s \in S$, (r, s) is the OD pair; $\mathrm{TT}^{rs}_{\mathrm{peak}}$ is the average travel time for trips during peak hours corresponding to the OD pair (r, s); $\mathrm{TT}^{rs}_{\mathrm{off}}$ is the average travel time of the OD pair (r, s) in off-peak hours; d^{rs}_{peak} and d^{rs}_{off} are the peak and off-peak demands of the

OD pair (r, s), respectively; the toll rate ratio is $0 \leq \tau \leq 1$; the box constraints are considered in this model, i.e. $\mathbf{z}_{\min}$ and $\mathbf{z}_{\max}$, which are lower and upper boundaries for segment toll rates, respectively.

3.5.3 Surrogate Models Evaluation

All the sample points are evaluated by DynusT to provide observations for surrogate model construction and SBO. To achieve convergence, 10 iterations of DTA and simulations are run for each toll plan, and the relative gaps for the DTA are found to be below 3% for all experiments. However, to save computation effort as much as possible, every toll plan is evaluated by the simulation-based DTA only once but including 10 iterations, despite the existence of noise in the route choice results. For each simulation run, the simulator obtains valid results when the convergence is achieved after several times of assignments and vehicular platoon simulations.

To simplify the optimization problem, this case study only cares about the travel time for users departing during the extended morning peak period (5–10 am). This extended morning peak consists of two off-peak periods (5-6 am and 9-10 am) and the morning peak period (6-9 am). To make the simulation results for the research period more reliable, one hour before 5 am is added into the simulation for warming up, and an extra hour after 10 am is also introduced to let the vehicles in the network disseminate. Thus a total of 7 hours is simulated in DynusT. Using a desktop with 3.10 GHz-quad CPU and 4 GB-Ram, one evaluation of each point takes about 5 hours. The average travel time for all network users during the extended morning peak period that finished their trips is computed after each simulation. Besides the

single simulation for each sample points, to characterize the baseline, we run 10 repetitions that produce 17.96, 17.94, 17.71, 18.22, 18.30, 17.83, 18.04, 17.78, 19.31, 18.44 minutes, respectively. The average baseline output is 18.15 min, and the standard deviation is 0.47 min, which could be an estimate of the simulation noise standard deviation, i.e. $\hat{\sigma}_{noise} = 0.47$ min. To estimate the mean and standard deviation of stochastic simulation outputs of the optimal variables, we will also run 10 repetitions, and then compare the estimated mean objective function value and standard deviation with the baseline to see how much improvement can be achieved after optimization.

As the simulation of the ICC network costs about 5 hours for each sample shown in Table 3-5, the surrogate models help reduce tremendous computation time compared to a traditional scenario study, which needs to evaluate all possible solutions through the entire feasible domain.

Methods	RMSE	MAX	NRMSE	NMAX	EGO	$E[f_{TT}(\hat{\mathbf{x}}^*)]$
M1	0.64	1.63	3.63%	3.18	**15.69**	-
M3	0.52	1.16	2.91%	2.26	17.14	-
M11	**0.52**	**1.41**	**2.92%**	**2.74**	17.51	-
M12	0.53	1.48	2.99%	2.90	17.36	**17.70**

Table 3-6: Leave-One-Out Cross Validation Results.

Using the four models we chose in the numerical test section, Table 3-6 shows the evaluation of response surfaces using the leave-one- out CV based on the first five MoEs introduced previously. We can see that the ordinary Kriging (M3), regressing Kriging (M11) and the expected improvement infill re-interpolating Kriging (M12)

approaches produce smaller estimation errors than the quadratic polynomial model (M1), which doesn't regress the response surface very well (larger RMSE, MAE and NMAX) though its EGO is extremely small. However, the simulation outputs of the samples should be treated as random variables with simulation errors instead of an accurate number. In this case, the aforementioned 5 MoEs are not suitable for evaluating the performance of these response surface models any more. Thus a new method analyzing the effectiveness of estimated confidence interval is proposed to evaluate model performance for the cases with significant simulation errors. The best model is then identified using this approach, and is then applied to search for the optimal toll rate in minimizing network average travel time.

Since no *a priori* information of simulation random errors is available, such random deviations from the expected smooth response can be simplified as uniformly distributed across the feasible domain, i.e. $\text{Var}[f_{\text{TT}}(\mathbf{x})] = \sigma_{\text{noise}}^2, \forall (\mathbf{z}, \tau)$, which is independently distributed with the regression error variance $\hat{s}^2(\mathbf{x})$. Then the estimated response variance $\hat{s}_{\text{TT}}^2(\mathbf{x})$ at $\mathbf{x}$ is given by

$$\hat{s}_{\text{TT}}^2(\mathbf{x}) = \hat{s}^2(\mathbf{x}) + \hat{\sigma}_{\text{noise}}^2 \qquad (3.31)$$

where $\hat{\sigma}_{\text{noise}}^2$ is the estimate of the simulation noise variance.

Figure 3-8 applies the leave-one-out CV to predict the response of each sample point using other points. Four surrogate methods (M1, M3, M11 and M12) are compared. Figure 3-8(a) shows the estimated mean of average travel time values $\hat{\mathbf{f}}_{\text{TT}}(\mathbf{x})$ at the initial 67 points as shown in Table 3-5 using M1. The total estimation standard error given $\hat{\sigma}_{\text{noise}} = 0.47$ is formulated by

$$\hat{s}_{\mathrm{TT}}(\mathbf{x}) = \sqrt{\hat{s}^2(\mathbf{x}) + \hat{\sigma}_{\mathrm{noise}}^2} \qquad (3.32)$$

The estimated mean values with one standard error upper and lower bounds are given by $\hat{\mathbf{f}}_{\mathrm{TT}}(\mathbf{x}) \pm \hat{\mathbf{s}}_{\mathrm{TT}}(\mathbf{x})$ for the training set $\mathbf{X}$. As a comparison, we plot the random observations $\mathbf{y}$ as well. The sample points in $\mathbf{X}$ are sorted according to estimated mean values of the average travel time in a descending order. The $\hat{\mathbf{f}}_{\mathrm{TT}}(\mathbf{x})$ of M1 decreases quickly and the estimated response of the last point adds bias to the observations.
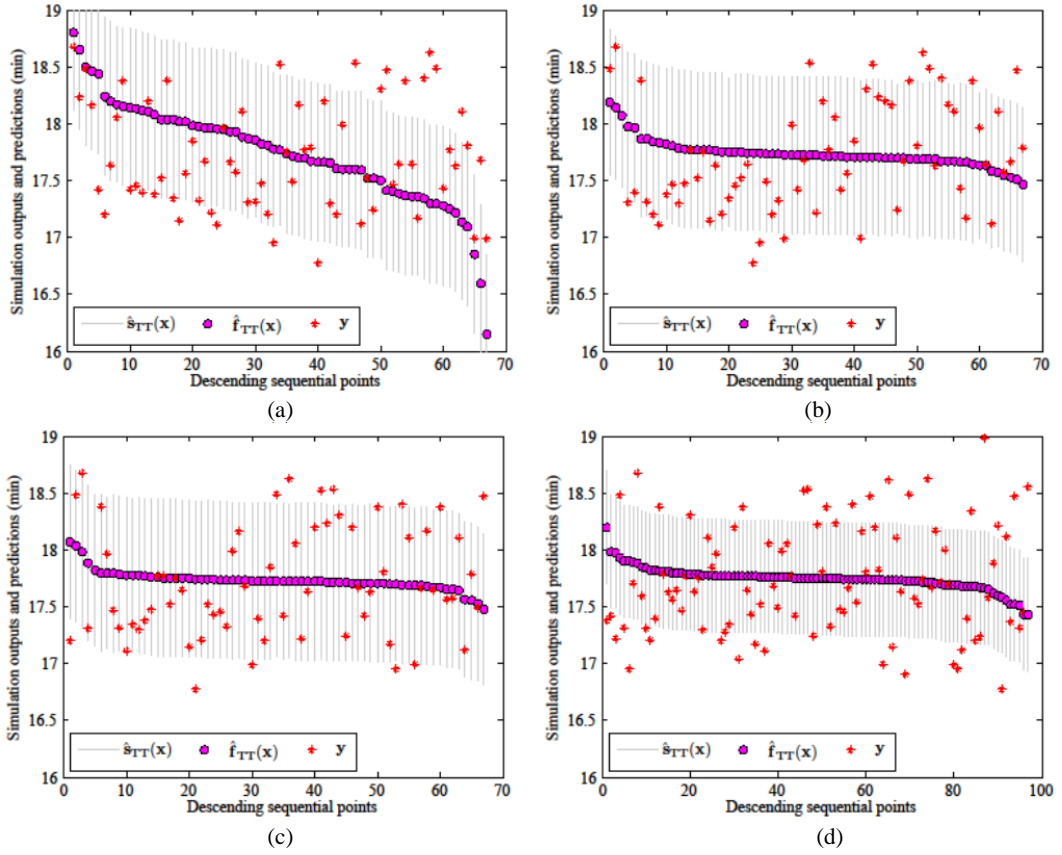
Figure 3-8: Prediction Accuracy of the Leave-One-Out Cross Validation: (a) M1; (b) M3; (c) M11; (d) M12.

Figure 3-8(b-c) show the CV results of M3 and M11 for the 67 initial sample points. The difference between them is very small, which is also shown in Table 3-6.

Figure 3-8(d) shows the results of the re-interpolating Kriging model (M12) for 97 points (including the initial samples and 30 infill points using the EI maximization criterion). The estimated standard deviation is smaller in this model, which means the optima after 30 infill points would have a smaller variance. Note that all random observations are within two standard deviations from the mean accounting for about 95% confidence level. The results show the advantage of M12: much narrower estimation bounds is shown for M12 than for M3, M11 or M1, which indicates a higher predicting concentration.

Therefore, based on the overall regression performance (indicated by RMSE, MAX, NRMSE, NMAX and EGO) and the prediction error bounds, we find the infill re-interpolating Kriging (M12) gives the best solution to the problem.

3.5.4 Optimization Results

At the end of the 30th update, the estimated best solution is $\hat{\mathbf{x}}^* = [\$\,2.28, \$\,0.15, \$\,1.29, \$\,1.31, \$\,0.24, 69\%]^{\mathrm{T}}$, the estimated global mean value of the minimized objective function is $\hat{f}_{\mathrm{TT}}^* = \hat{f}_{\mathrm{TT}}(\hat{\mathbf{x}}^*) = 17.36$ min. The ratio of the off-peak to peak pricing is reduced, not suggesting that tolls in the off-peak are increased, because the peak-hour toll charge rates of the optima increase for highway Segments I, III and IV, but the rates decrease for highway Segments II and V. Based on 10 runs of simulation under the same best input, we find the mean value of simulation outputs is 17.70 min that is 2.5% less than the mean value of the baseline average travel time. We compute the F statistic and p-value based on the analysis of variance (ANOVA) for the baseline and optimized toll charge plans. Results show the F statistic is 4.68

and the corresponding *p*-value is 0.04 that is close to zero indicating mean travel times are significantly different.
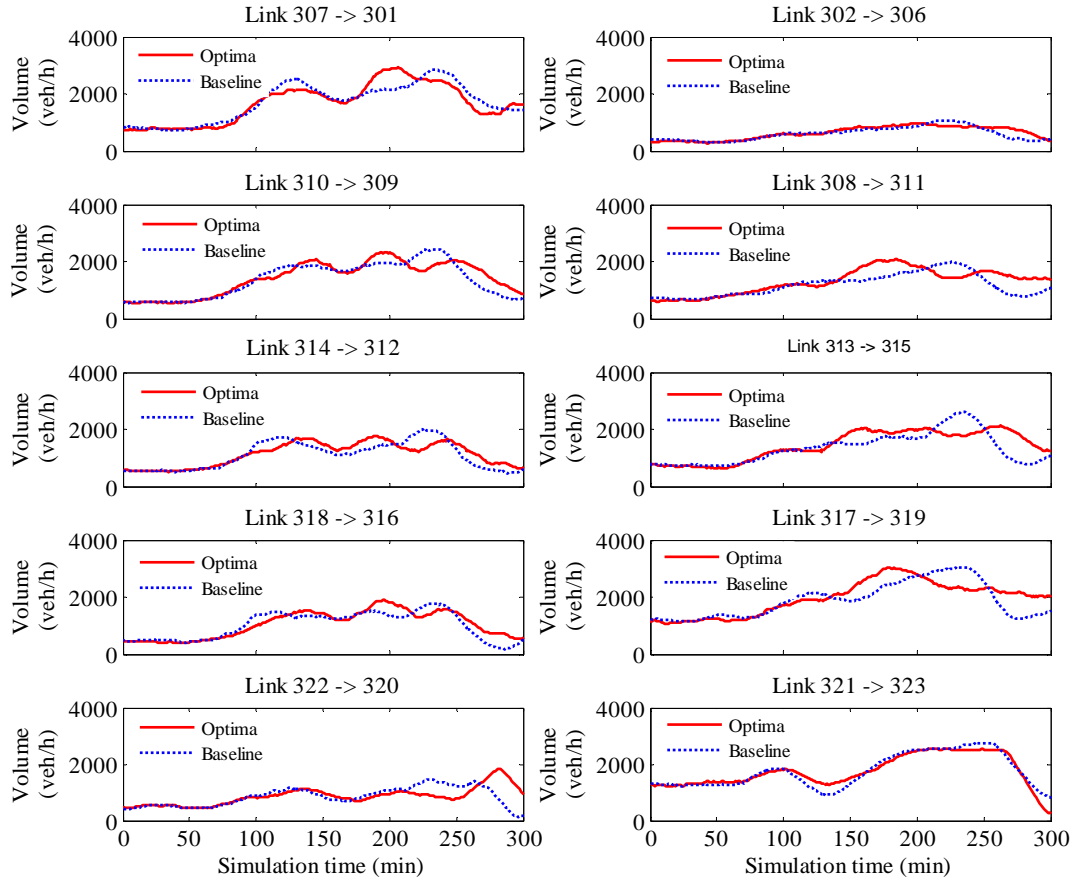
Figure 3-9: Comparisons of Traffic Flow Volumes of Additive Toll Links between the Baseline and Optimized Toll Rates**.**

The simulation generates statistics of the network performance every minute. As link volume fluctuates significantly at the one-minute interval, Figure 3-9 shows 10-minute moving average traffic flow volume of the 10 additive toll links in the network. The left column shows westbound ICC, and the right column shows eastbound ICC. Two subfigures in each row represent westbound and eastbound toll segment, i.e. Segments I, II, III, IV and V from top to bottom. Overall, the time series of toll link volumes is changed under the optimized toll compared to the baseline,

while the total flow passing through each toll link during the simulation period did not change a lot. In 8 out of the 10 additive toll links, traffic volume is significantly higher during the fifth hour, i.e. 9–10 am (240 min to 300 min) under the optimal toll.
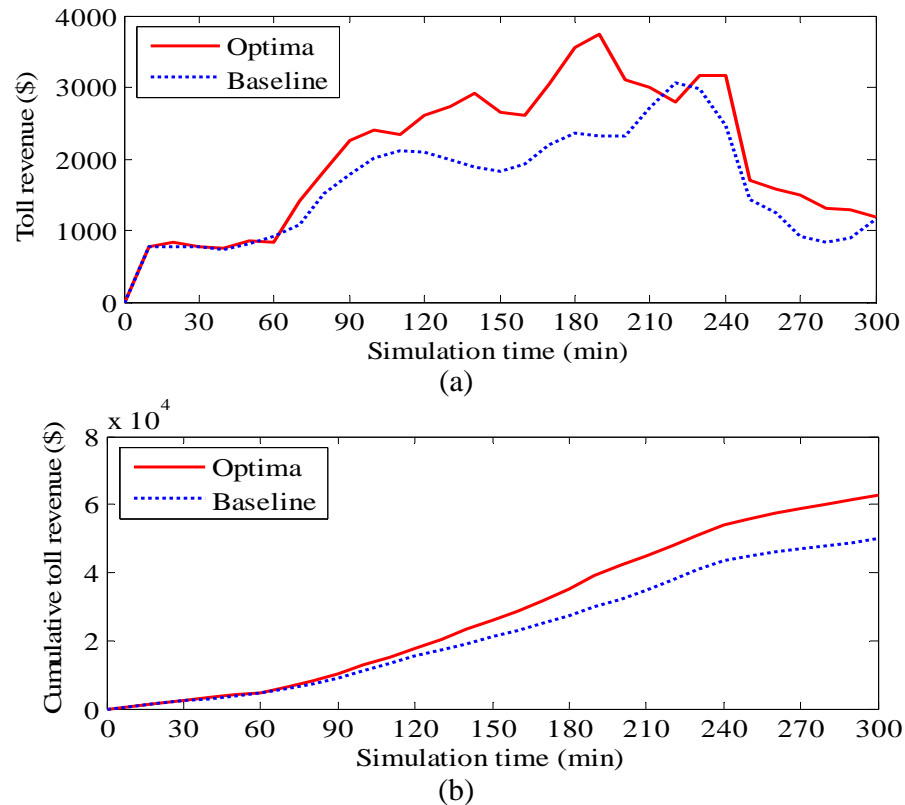
Figure 3-10: Comparison of Toll Revenue Collected on Additive Toll Links between the Baseline and Optimized Toll Rates, (a) Toll Revenue; (b) Cumulative Toll Revenue.

Figure 3-10(a) shows the comparison of toll revenue dynamics between the optimal toll and the baseline. The curve of cumulative toll revenue collected along time in Figure 3-10(b) shows that the optimized toll case generates a toll revenue of around 62 thousand dollars, which is a 20% increase compared to the current toll case. During the first hour of the simulation period, toll revenue collected under both cases is almost the same. During the three peak hours from 6 to 9 am, as traffic flow

of all toll links is very close, the increased toll revenue of the optimal toll case mainly comes from the increased peak toll rates. The peak/off-peak ratio of the optimal toll case (69%) is smaller than that of the current toll case (80%), and the average off-peak toll rates of the two cases are about on the same level. The increase of toll revenue during the last hour of the simulation period under the optimal toll mainly comes from the increase of link volumes. Figure 3-11 compares the total flow throughput at the network exit for the optimal and the baseline solutions. The optima increase the throughput capacity in peak hours (especially from 180 to 240 min).

(a)

(b)
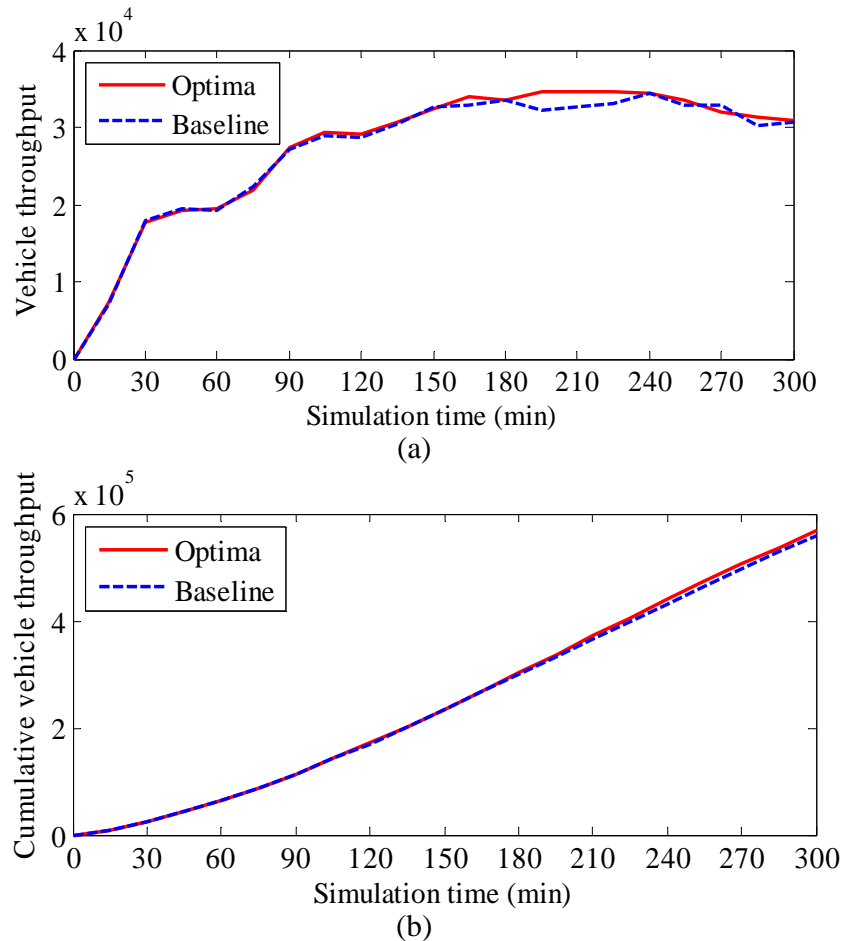
Figure 3-11: Comparison of the Vehicle Throughput between the Baseline and Optimized Toll Rates, (a) Vehicle Throughput; (b) Cumulative Vehicle Throughput.

Because we only suggest change the toll rate on one highway, so the influence of the small scale of toll rate change on the whole transportation network should be small. However, the small improvement in the mean travel time of all users of the transportation network (2.5% reduction) cannot be neglected. The average travel time under optimal toll was 17.70 minutes, which is 0.45 minutes shorter than the average travel time under the current tolls. The total vehicle throughput for the simulation period is around 570,000. The value of time is assumed to be \$15/hour for the network users. Thus the reduction in average travel time equals a total of \$65,000 saving in travel cost for the extended morning peak period (5 hours) for each day. If we consider the whole 24 hour of each day, and even consider a one-year effect, such a small improvement in mean travel time in the extended peak hours would mean a huge saving from an operational and policy standpoint.

Overall, the simulation results show that implementing the optimal toll predicted by the Kriging model considering simulation noise can benefit society in multiple ways. Travelers gain from the reduction of travel cost and the government benefits from the increase of toll revenue, while there is hardly any cost associating with the change of toll rates. Thus adjusting the current toll rate to the optimal toll rate should be an encouraging policy option to enhance transportation system performance in the study region.

3.5.5 Sensitivity Analysis

Moreover, to explore the sensitivity for the baseline $\mathbf{x}_{baseline}$ and the optima $\hat{\mathbf{x}}^*$ in the method M12, we provide a joint contour plot of the baseline in Figure 3-12(a).
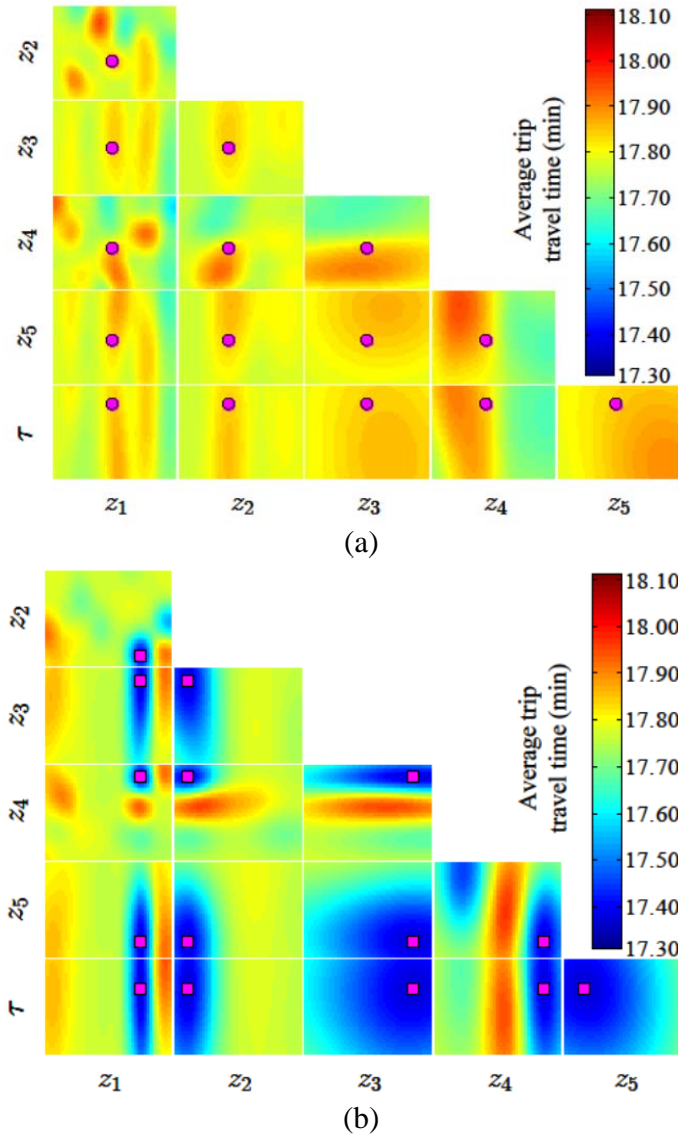
Figure 3-12: Sensitivity Analysis of the (a) Baseline and (b) Optima.

Each tile shows a contour of the estimated surrogate function $\hat{f}_{\mathrm{TT}}(\mathbf{x})$ (the average travel time) versus two of the six variables, with the remaining four variables held at the baseline value. This helps visualize how the surrogate values change around $\mathbf{x}_{\mathrm{baseline}}$. The baseline values and the ranges of each dimension can be found in Table 3-5. Take the upper-left contour plot of the average travel time surrogate

function as an example, it is a conditional function of $\hat{f}_{TT}(z_1, z_2 \mid z_3, z_4, z_5, \tau)$ given

$z_3 = \$\, 0.75$, $z_4 = \$\, 0.65$, $z_5 = \$\, 0.70$, $\tau = 80\%$. The warmer colors of the joint

contour plots indicate the longer average travel time, while the cooler colors show

shorter travel time values.

Analogically, the joint contour plots in Figure 3-12(b) show the sensitivity

analysis around the optima $\hat{\mathbf{x}}^{*}$, which is denoted by the squares. The main difference

from Figure 3-12(a) is that the values around $\hat{\mathbf{x}}^{*}$ are toward much cooler colors

(closer to cyan and blue) than the baseline sensitivity (closer to orange and red). It

also validates that the optimal solution performs better than the baseline inputs.

3.6 Conclusions

In this chapter, a systematic framework of transportation simulation-based

optimization is proposed to solve the highway toll optimization with expensive-to-

evaluate objective functions and obvious simulation random errors. The main novelty

of computational time savings can be achieved through applying the efficient

surrogate models which can intelligently mimic the simulation input-output mapping.

A family of surrogate-based optimization approaches to model response

surface of transportation simulation input-output is introduced. Both one-stage and

two-stage surrogate models are tested and compared using a small-scale numerical

study. Four out of the twelve surrogate models are identified to be promising

approximation of the true response surface of the numerical study, and are then applied to the real world application of dynamic pricing optimization.

Taking advantage of the simulation's power in reflecting the traffic realism, we evaluate the transportations system performance in response to different toll charges in a real transportation network with a simulation-based DTA model DynusT. Among the four selected promising models, the expected improvement infill re-interpolating Kriging performs the best. With only 97 samples, this model can produce highly reliable estimates of simulation outputs over the entire feasible domain, and thus successfully help find the optimal toll rate. The predicted optimal toll rate obtained from the Kriging model is then evaluated through simulation. The predicted output is relatively consistent with simulation outputs. Overall, a 2.5% improvement to the entire network in terms of average trip travel time can be achieved, through adjusting the toll rate of a single freeway from the baseline to the predicted optima. The travel time savings for the extended morning peak period of each day is equivalent to $65,000 assuming a $15 per hour value of time.

Chapter 4: Constrained and Multi-Objective Simulation-based Optimization

Single-objective optimization without any constraints is the simplest form for an optimization problem, and is usually the first stage in research efforts. Regarding the highway toll optimization problem we investigated in the previous chapter, there are usually multiple performance measures that highway operators would be concerned about. If the toll road is operated by a private company, the major objective is most likely to be maximizing toll revenue, while they would also pay attention to the traffic condition on the toll road, so as to guarantee a minimum level of service to the paying users. If a toll road is owned and operated by public agencies, maximizing social welfare for all users of the network is probably the primary objective. Meanwhile, they also care about the revenue that can be collected from the toll and some other measures like air pollution and safety. Under the situation that decision makers have specific expectation on several of the performance measures, the problem can be formulated as a constrained optimization problem. Otherwise, multi-objective optimization would be the proper form of the problem to be analyzed.

4.1 Constrained Simulation-based Optimization

In this section, we extend the previously investigated single-objective problem into a constrained optimization form, and try to deal with the toll road pricing optimization problem from the public agencies' point of view more realistically.

Specifically, the public agencies are assumed to be interested in two performance measures: average travel time for the network users and the toll revenue. The scenario to be investigated is that public agencies aim to minimize network wide average travel time while intending to keep the toll revenue higher or equal to a specific level.

Via our observation during the previous study, the simulation noise is very significant. Thus we decide to perform 3 repetitions of the evaluation for each sample point and use the average of outputs as the observations, so as to reduce the effect of the simulation random noises. The repetitions would significantly increase the computation time for the evaluation of initial sample points as well as the infill points. To reduce the overall computational burden, we only evaluate the transportation system performance during the 3-hour peak period (6-9 am). In this case, 5 hours (5-10 am, including one hour (5-6 am) for warming up and one hour (9-10 am) for dissemination) is simulated in DynusT. Using a desktop with 3.10 GHz-quad CPU and 4 GB-Ram, one evaluation of each point takes about 3 hours. Thus the 3-repetition evaluations for all the 67 initial points cost around 603 computer hours in all. The performance measures we are interested in are the average travel time for vehicles departing during the peak period and the total peak period toll revenue. As the network is relatively congested during the peak hours, there are still a few vehicles departing during the peak period cannot arrive at their destinations even one hour after the peak period. Thus we only compute the average travel time for those vehicles that finished their trips.

4.1.1 Problem Formulation

The constrained optimization problem can be formulated as follows:

$$\min_{\mathbf{x} \in \mathbb{R}^k} \mathrm{E}\left[f_{TT}(\mathbf{z}, \tau) \right] = \mathrm{E}\left[\frac{\sum_{r \in R} \sum_{s \in S} d_{\text{peak}}^{rs} \cdot \mathrm{TT}_{\text{peak}}^{rs}(\mathbf{z}, \tau)}{\sum_{r \in R} \sum_{s \in S} d_{\text{peak}}^{rs}} \right] \tag{4.1a}$$

$$\text{s.t.} \quad \mathrm{E}\left[f_{TR}(\mathbf{z}, \tau) \right] = \mathrm{E}\left[\sum_{a \in TA} v_{\text{peak}}^{a}(\mathbf{z}, \tau) \cdot \mathbf{z}(\sigma(a)) \right] \geq TollLimit \tag{4.1b}$$

$$\mathbf{z}_{\min} \leq \mathbf{z} \leq \mathbf{z}_{\max}, \quad \mathbf{z} \in \mathbb{R}^{k-1} \tag{4.1c}$$

$$0 \leq \tau \leq 1 \tag{4.1d}$$

$$\mathbf{x}^T = [\mathbf{z}^T, \tau], \quad \mathbf{x} \in \mathbb{R}^k \tag{4.1e}$$

where $f_{TR}(\mathbf{z}, \tau)$ is the function capturing the stochastic toll revenue for the peak period; TA is a subset of links that are tolled in the network; $v_{\text{peak}}^{a}(\mathbf{z}, \tau)$ represents the peak period volume on link a; $\sigma(a)$ is a function mapping the ID of link a to its corresponding index in the vector $\mathbf{z}$; $TollLimit$ is the predetermined lower bound of the peak-hour toll revenue. All other variables have the same meanings with those explained in the single-objective problem.

Although only the performance in peak period is of concern, the off-peak toll rate would still influence the objective value, because the route choice for vehicles departing at the beginning or end of the peak period would be indirectly affected by the off-peak toll due to its impact on the distribution of vehicles in the network both before and after the peak hours. After the simulation evaluation of a certain toll plan, the travel route and travel time for each vehicle trip can be generated and saved. At the link level, time dependent flow on each link can also be summarized. Both

network wide average travel time and toll revenue for the peak period can thus be recorded.

Expected improvement infill re-interpolating Kriging is identified as the best model in the single-objective optimization problem introduced in section 3.5.3, we thus will only investigate this model for the constrained optimization as well as the multi-objective optimization to be introduced in the next section. In fact, re-interpolating Kriging is just a slight revision of regressing Kriging (Forrester et al., 2006), to adapt for the deterministic experiments. As the simulator DynusT in the current study performs nondeterministic experiments, regressing Kriging model works perfectly fine with both approximating the response surface and the process of selecting infill points. Therefore, expected improvement infill regressing Kriging is finally applied for the constrained optimization and the multi-objective optimization problems.

4.1.2 Infill Strategy

For a constrained optimization problem with both the objective and constraint to be expensive-to-evaluate functions as illustrated in equation 4.1, considering only the expected improvement of the objective value at a point $\mathbf{x}^*$ is not sufficient. If the probability that the constraints would be violated at the point $\mathbf{x}^*$ is very high, the overall expectation of improvement should still be very low even if the EI of objective value is large at that point. The overall expectation of improvement at a certain point is large only if both the EI of objective value and the probability of satisfying the constraints are large.

Suppose the surrogate for the constraint function is $G(\mathbf{x})$ and the constraint limit is $g_{\min}$. The mean value and estimation variance at a point $\mathbf{x}^*$ can be computed in the same way as for the objective function, which are denoted by $\hat{f}_g(\mathbf{x}^*)$ and $\hat{s}_g(\mathbf{x}^*)$. The probability that the constraint is met can be given by

$$P[G(\mathbf{x}^*) \geq g_{\min}] = \frac{1}{\sqrt{2\pi}\hat{s}_g(\mathbf{x}^*)} \int_{g_{\min}}^{\infty} \exp\left(-\frac{(u - \hat{f}_g(\mathbf{x}^*))^2}{2\hat{s}_g^2(\mathbf{x}^*)}\right) du \qquad (4.2)$$

We define an indicator function to express the constrained improvement as follows.

$$CI(\mathbf{x}) = \begin{cases} y_{\min} - Y(\mathbf{x}) & \text{if } G(\mathbf{x}) \geq g_{\min} \\ 0 & \text{otherwise} \end{cases} \qquad (4.3)$$

As the surrogate for the objective function is dependent from that for the constraint function, the constrained expected improvement at the point $\mathbf{x}^*$ can thus be defined as

$$E[CI(\mathbf{x}^*)] = EI(\mathbf{x}^*) \bullet P[G(\mathbf{x}^*) \geq g_{\min}] \qquad (4.4)$$

The choice of infill points for the constrained optimization problem can be based on maximizing the constrained expected improvement.

4.1.3 Optimization Results

We then apply this infill strategy of constrained expected improvement in the ICC pricing optimization study. In the constrained optimization problem setting, the constraint limit of the toll revenue is set to be the revenue collected at the baseline case, which is \$7552.3. The target of this problem is to improve the network wide

traffic condition during the peak period as much as possible, while keeping collecting toll revenue at least at the same level as the baseline case.
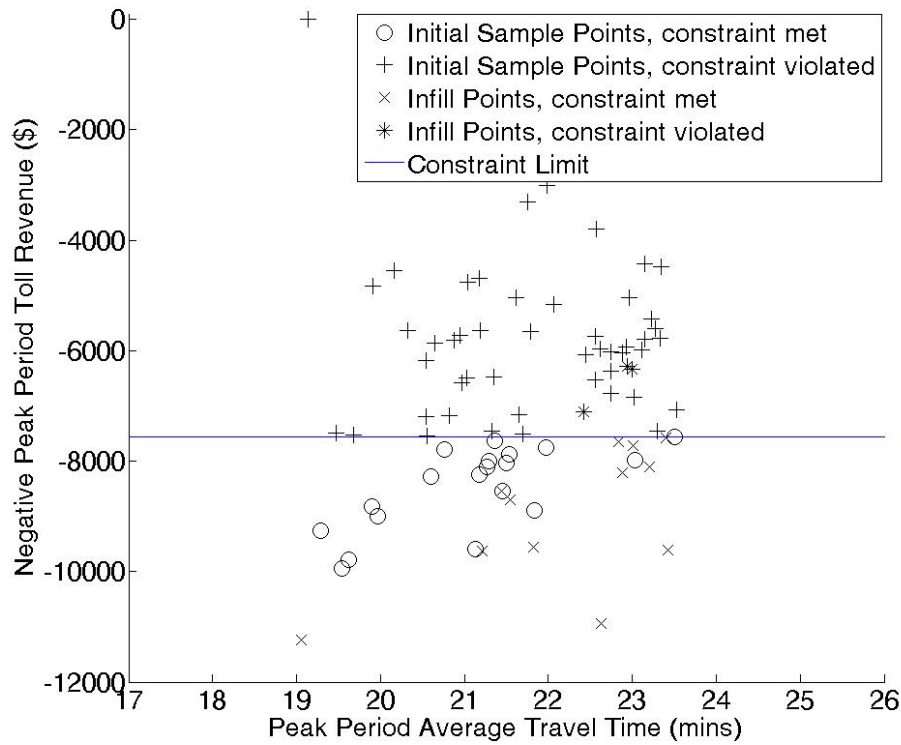
Figure 4-1: Initial and Infill Points for the Constrained Optimization Problem.

According to the criterion of expected improvement, 15 infill points (each infill point is simulated three times, and the total cost is about 135 computer hours) are added into the sample. Among the 15 infill toll plans, the toll revenue collected under 12 toll plans meets the constraint. All the initial points and infill points are plotted in Figure 4-1. The horizontal line represents the constraint limit, and the circle lies exactly on the line represents the output of the baseline case. The average travel time for the baseline case is about 23.51 minutes. Compared to other toll plans in both

the initial sample and infill sample, the baseline case performs very bad regarding average travel time.

Based on the initial sample, the best toll plan satisfying the constraint is

$$\mathbf{x}_{best}^{initial} = [\$\,0.67, \$\,0.83, \$\,0.33, \$\,1.48, \$\,1.07, 76\%]^{\mathrm{T}} ,$$

$$TT_{best}^{initial} = 19.29 \min s , \quad TR_{best}^{initial} = \$9,261.1 .$$

The infill points improved the average travel time a little bit. The best toll plan from the infill points is

$$\mathbf{x}_{best}^{infill} = [\$\,0.66, \$\,0.84, \$\,0.38, \$\,0.61, \$\,1.16, 39\%]^{\mathrm{T}} ,$$

$$TT_{best}^{infill} = 19.06 \min s , \quad TR_{best}^{infill} = \$11,243.4 .$$

The network average travel time in reduced by about 1.2%, and the added benefit is that toll revenue has increased by around $2,000, which is a 20% improvement.

Most of the infill points satisfy the constraint, which means the criterion of expected improvement is very effective in searching for feasible solutions. However, as the expected improvement method balances between local exploitation and global exploration, the regions with high average travel time would still be explored. The objective values (peak period average travel time) of those infill points vary a lot. Most of them perform better than the baseline case, while there is only one point found to outperform the best observation in the initial sample.

We search for the constrained optima after infilling 15 points, and find that the global optimum coincides with the infill point that produces the best performance. As DynusT provides detailed information regarding the traffic conditions, the transportation performance on ICC links and at the aggregate level are compared

between the estimated global optima and the baseline. Figure 4-2 shows the variations of traffic volume on the five tolled segments of the ICC during the peak period. Traffic flow for different directions is depicted separately. Even during the peak period, flow on the ICC is always lower than 3,000 vehicles per hour. As the ICC has three lanes in each direction, the capacity of ICC is not used sufficiently. Traffic volume under the optimal toll plan is significantly higher on three of the segments than the baseline case. This is an expected change since the toll rate for the three segments is lower than the baseline. The optimal toll for segment five is 70% higher than the baseline, and the volume for the east bound of segment five exhibits apparent drop compared to the baseline.
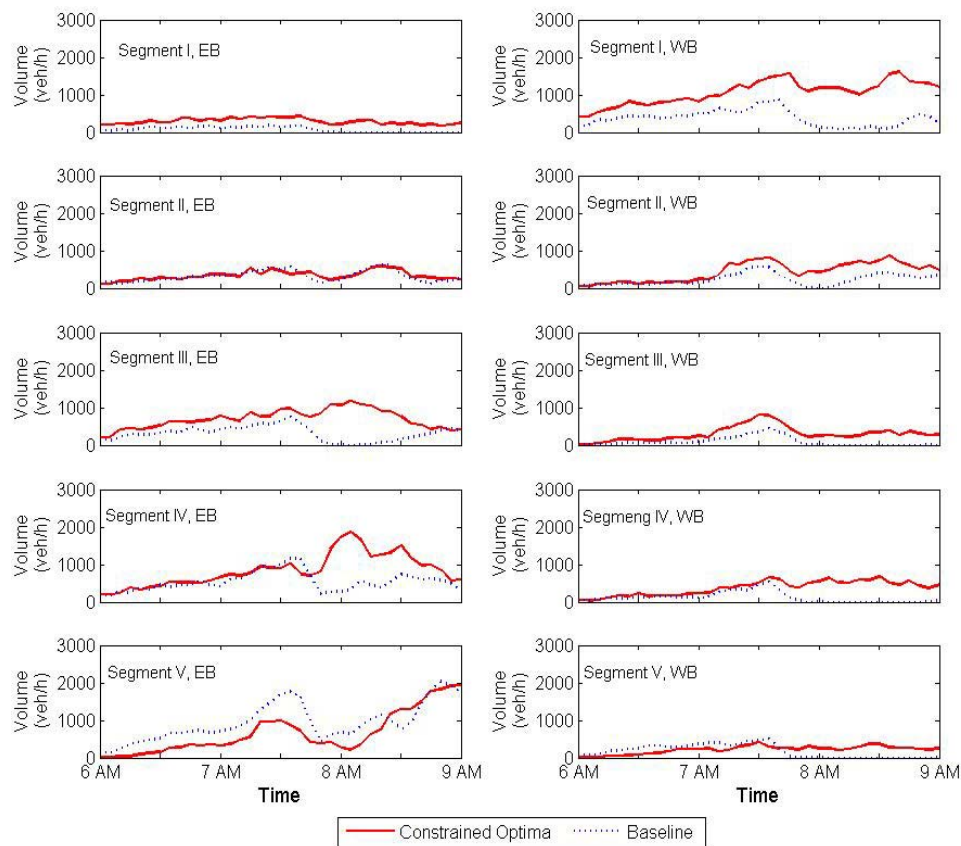
Figure 4-2: Traffic Volumes of ICC Segments under Optima and Baseline Toll.

Various network wide performance measures are also compared between the two toll plans. The network wide average trip travel time for the two cases is illustrated in Figure 4-3. The optimal toll plan reduces average travel time by almost 10 minutes during the most congested period (7:30-9:00 am). Looking back to Figure 4-2, we can find the most significant increase in the use of the ICC just occurs at that period, which means the optimal toll plan performs very well in diverting vehicles from very congested routes.
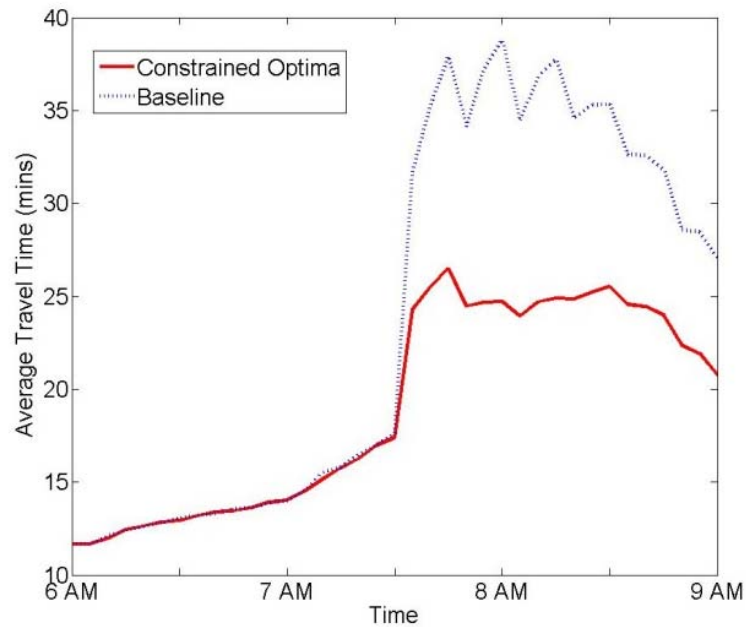
Figure 4-3: Comparison of Network Wide Average Travel Time between Optimal and Baseline Toll Plans.

In the same manner, due to the significant increase of users in the period of 7:30-9:00 am, toll revenue collected during this period is almost two times higher than the baseline case as shown in Figure 4-4. The overall increase of the toll revenue for the optimal toll case during the three-hour peak period is about $3,000. Moreover, the improvement in travel time brings in additional benefits. Figure 4-5 shows that the

throughput for the whole network increases significantly for the period 7:30-9:00 am, and the total throughput for the morning peak period increased by about 10% under the optimal toll. The adjustment of toll rates for one freeway in the regional network has successfully led to a more efficient usage of the road network capacity.

(a)   (b)
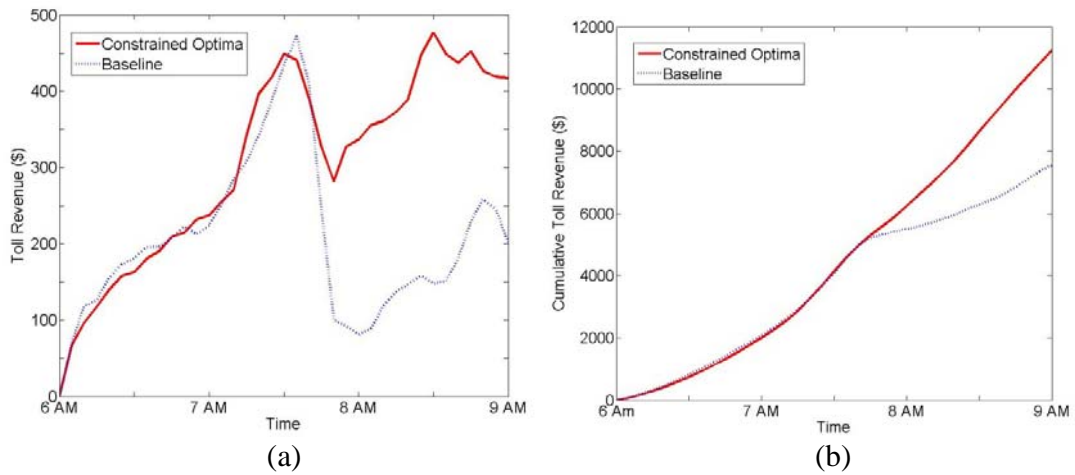
Figure 4-4: Comparison of Toll Revenue between Optimal and Baseline Toll Plans, (a) Toll Revenue; (b) Cumulative Toll Revenue.

(a)   (b)
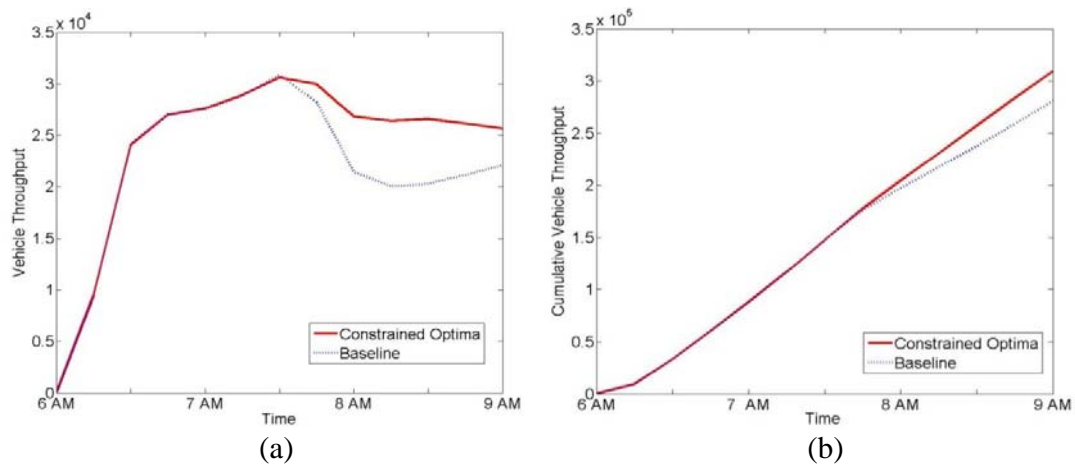
Figure 4-5: Comparison of Throughput between Optimal and Baseline Toll Plans, (a) Vehicle Throughput; (b) Cumulative Vehicle Throughput.

In summary, the optimal solution can reduce average travel time by around 20% (4.5 minutes in reduction). The overall throughput for the peak period is about

300,000, thus the total reduction in travel time for the morning peak hours is 22,250 hours (around $300,000 in travel time saving assuming $15 per hour value of time), which is very remarkable. Moreover, the added benefit is that the total toll revenue is increase by 50%.

4.2 Multi-Objective Simulation-based Optimization

As illustrated previously, there are usually multiple performance measures that highway operators would be concerned about. However, the constrained optimization problem can be formulated only when the decision makers are pretty sure about the limit they want to impose on several of the performance measures. If the highway operators have neither specific weights among the multiple goals nor expectations regarding any of the objectives, the dynamic pricing optimization for toll roads can be formulated as a multi-objective optimization problem. In this case, a set of optional toll plans which do not dominate each other can be generated. This set of compromise solutions is called the Pareto set, and it can serve as the foundation for decision making.

For the multi-objective optimization problem, the objectives are assumed to be minimizing peak-period average trip travel time and maximizing peak-period toll revenue. The same initial samples and their corresponding observations for the constrained optimization problem in the previous section are utilized to solve this multi-objective problem. However, the infill strategies and the infill points would be different.

4.2.1 Problem Formulation

The multi-objective optimization problem can be formulated as follows:

$$\min_{\mathbf{x} \in \mathbb{R}^k} \mathrm{E}\left[f_{TT}\left(\mathbf{z}, \tau\right) \right] = \mathrm{E}\left[\frac{\sum_{r \in R}\sum_{s \in S} d_{peak}^{rs} \cdot TT_{peak}^{rs}\left(\mathbf{z}, \tau\right)}{\sum_{r \in R}\sum_{s \in S} d_{peak}^{rs}} \right] \tag{4.5a}$$

$$\max_{\mathbf{x} \in \mathbb{R}^k} \mathrm{E}\left[f_{TR}\left(\mathbf{z}, \tau\right) \right] = \mathrm{E}\left[\sum_{a \in TA} v_{peak}^{a}\left(\mathbf{z}, \tau\right) \cdot \mathbf{z}\left(\sigma(a)\right) \right] \tag{4.5b}$$

$$\text{s.t.} \quad \mathbf{z}_{min} \le \mathbf{z} \le \mathbf{z}_{max}, \quad \mathbf{z} \in \mathbb{R}^{k-1} \tag{4.5c}$$

$$0 \le \tau \le 1 \tag{4.5d}$$

$$\mathbf{x}^T = [\mathbf{z}^T, \tau], \quad \mathbf{x} \in \mathbb{R}^k \tag{4.5e}$$

Overall, the specification of this problem is very similar with that for the constrained optimization problem introduced in the previous section. The only difference is that one of the constraints in the constrained optimization is re-formulated as an objective function.

The expected results from this model is a set of optional toll plans, in which no one can dominate any of the others in terms of both of the two objectives. This set of results can provide decision makers a clear understanding on how good a combination of objectives can be achieved. Further evaluation of these optional toll plans can be done later when the weighting between the two objectives is determined.

4.2.2 Infill Strategy

After the evaluation of all initial sample points, a Pareto set of solutions can be identified. All the solutions in the Pareto set are optimal in some sense, while they cannot dominate any other solutions in the Pareto set in terms of all objective values. If a solution in the Pareto set is better than another in one of the objective values, it should be worse in at least one of other objective values than the other. Figure 4-6 shows an example of a Pareto set with five non-dominated points (noted as stars in the figure) for a problem with two objectives.

New points found in the shaded area would augment the current Pareto front, while any points in the hatched area will at least dominate one of the points in the current Pareto set. Thus any new points found in the hatched area would lead to the update of the current Pareto front.
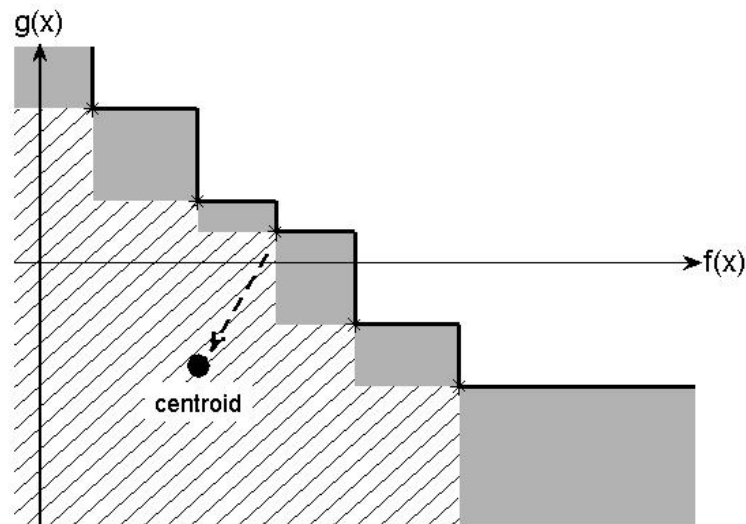
Figure 4-6: Example of a Pareto Set.

Similar to the case of the constrained optimization, the two surrogates for the two objective functions are assumed to be independent. The probability density function (PDF) of a combination of output regarding the two objectives $\phi_{P-Gaussian}(\mathrm{y}_f, \mathrm{y}_g)$ can be described as the product of two Gaussian PDFs. The probability that a new point $\mathbf{x}^*$ is an improvement to the current Pareto front can thus be computed by integrating the joint PDF $\phi_{P-Gaussian}(\mathrm{y}_f, \mathrm{y}_g)$ over the region below and to the left of the current Pareto front, which include both the shaded and hatched area shown in Figure 4-6. The probability of improvement at a new point $\mathbf{x}^*$ is expressed as

$$
\begin{aligned}
\mathrm{P}[\mathrm{I}(\mathbf{x}^*)] = \int_{-\infty}^{f^{*(1)}} \int_{-\infty}^{\infty} \phi_{P-Gaussian}(\mathrm{y}_f, \mathrm{y}_g) \mathrm{dy}_g \mathrm{dy}_f + \sum_{i=1}^{M-1} \int_{f^{*(i)}}^{f^{*(i+1)}} \int_{-\infty}^{g^{*(i)}} \phi_{P-Gaussian}(\mathrm{y}_f, \mathrm{y}_g) \mathrm{dy}_g \mathrm{dy}_f \\
+ \int_{f^{*(M)}}^{\infty} \int_{-\infty}^{g^{*(M)}} \phi_{P-Gaussian}(\mathrm{y}_f, \mathrm{y}_g) \mathrm{dy}_g \mathrm{dy}_f
\end{aligned}
$$

(4.6)

where $(f^{*(i)}, g^{*(i)})$ are the points in the current Pareto set, $i = 1, 2, \ldots, M$ and M is the total number of points in the Pareto set.

The probability of improvement represents integration of the PDF over the region below and to the left of the Pareto front. The expected improvement is then the first moment of the integral over the same region. Suppose we know the position of the centroid of the $\mathrm{EI}(\mathbf{x}^*)$ integral (see Figure 4-6), the Euclidean distance between the centroid and each member of the Pareto set can be computed. The expected improvement criterion is subsequently calculated using the Pareto set member closest to the centroid, $(\mathrm{y}_f^*(\mathbf{x}^*), \mathrm{y}_g^*(\mathbf{x}^*))$, by taking the product of the probability of improvement with the shortest Euclidean distance (Forrester and Keane, 2009).

Suppose the position of the centroid is $(\hat{\mathrm{y}}_f(\mathbf{x}^*), \hat{\mathrm{y}}_g(\mathbf{x}^*))$, the expected improvement at

the point $\mathbf{x}^*$ is given by

$$\mathrm{EI}(\mathbf{x}^*)=\mathrm{P}[\mathrm{I}(\mathbf{x}^*)]\bullet\sqrt{(\hat{\mathrm{y}}_f(\mathbf{x}^*)-\mathrm{y}_f^*(\mathbf{x}^*))^2+(\hat{\mathrm{y}}_g(\mathbf{x}^*)-\mathrm{y}_g^*(\mathbf{x}^*))^2} \tag{4.7}$$

where

$$\widehat{\mathrm{y}_f}(\mathbf{x}^*) = \frac{\left(\begin{array}{c}\displaystyle\int_{-\infty}^{f^{*(1)}}\int_{-\infty}^{\infty}\mathrm{y}_f\bullet\phi_{P-Gaussian}(\mathrm{y}_f,\mathrm{y}_g)d\mathrm{y}_g d\mathrm{y}_f + \sum_{i=1}^{M-1}\int_{f^{*(i)}}^{f^{*(i+1)}}\int_{-\infty}^{g^{*(i)}}\mathrm{y}_f\bullet\phi_{P-Gaussian}(\mathrm{y}_f,\mathrm{y}_g)d\mathrm{y}_g d\mathrm{y}_f\\[4mm] +\displaystyle\int_{f^{*(M)}}^{\infty}\int_{-\infty}^{g^{*(M)}}\mathrm{y}_f\bullet\phi_{P-Gaussian}(\mathrm{y}_f,\mathrm{y}_g)d\mathrm{y}_g d\mathrm{y}_f\end{array}\right)}{\mathrm{P}[\mathrm{I}(\mathbf{x}^*)]}$$

$$\tag{4.8}$$

and $\hat{\mathrm{y}}_g(\mathbf{x}^*)$ can be computed similarly.

4.2.3 Optimization Results

Through infill points to the sample, the Pareto front for the multi-objective problem is expected to be augmented or updated. The Pareto front for the initial sample and the augmented sample with infill points are presented in Figure 4-7.
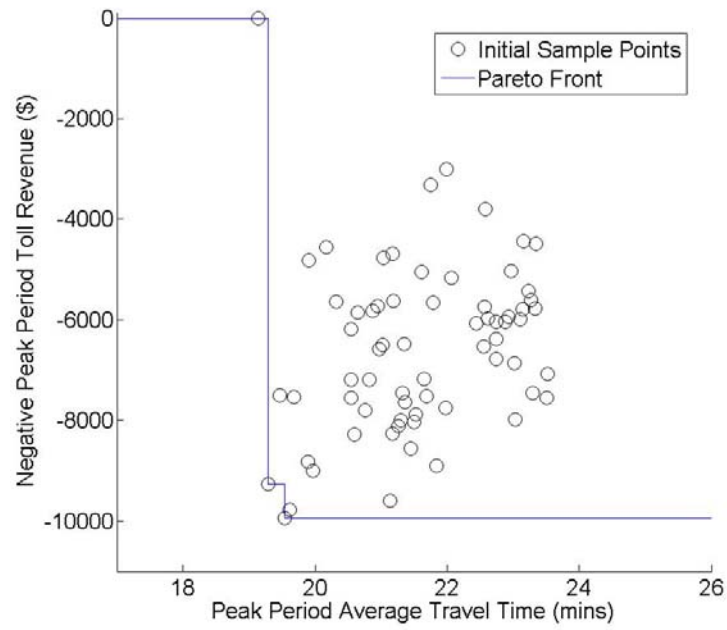
(a)
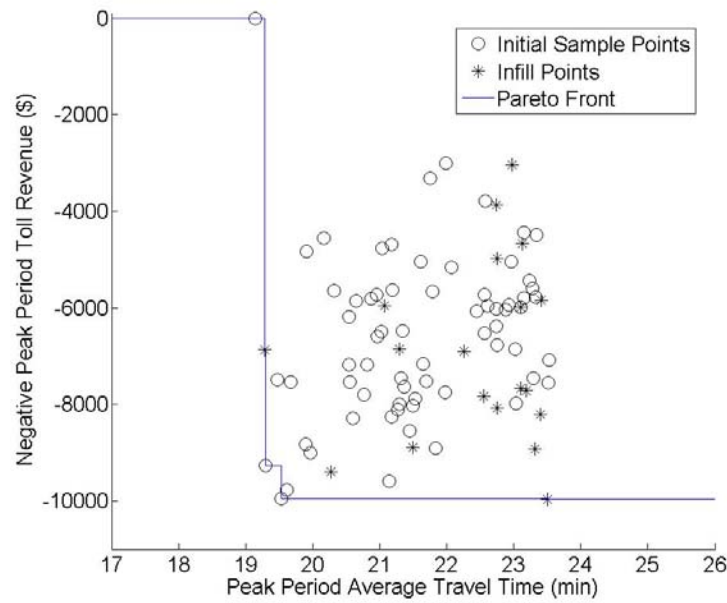
(b)

Figure 4-7: Pareto front for the multi-objective optimization problem, (a) Pareto Front Based on Initial Samples; (b) Pareto Front Based on Augmented Samples.

Similar to the method of evaluating initial sample points, the infill points are all evaluated three times through DynusT and the mean of the output is used as the observation.

Among the 67 initial sample points, only three of them are on the Pareto front. Any of the other points would be dominated by at least one of the points from the Pareto set in terms of both average travel time and toll revenue. The toll plan and the corresponding performance regarding the two objectives for the three Pareto set members are

$\mathbf{x}_1 = [\$\, 0, \$\, 0, \$\, 0, \$\, 0, \$\, 0, 0\%\,]^\mathrm{T}, TT_1 = 19.14 \min s, TR_1 = \$0$

$\mathbf{x}_2 = [\$\, 0.67, \$\, 0.83, \$\, 0.33, \$\, 1.48, \$\, 1.07, 76\%\,]^\mathrm{T}, TT_2 = 19.29 \min s, TR_1 = \$9,261.1$

$\mathbf{x}_3 = [\$\, 0.76, \$\, 0.45, \$\, 0.79, \$\, 0.81, \$\, 0.55, 40\%\,]^\mathrm{T}, TT_3 = 19.54 \min s, TR_3 = \$9,941.7$

Compared to the 0 toll case, more than 8,000 dollars can be collected in the 3-hour peak period through charging ICC users. At the same time, the average travel time for the whole network is just slightly increased. However, the maximal toll revenue can be collected is around 10,000 dollars. Keeping the toll revenue at this level, the network wide average trip travel time may increase dramatically by adjusting the toll plans. Just looking at the Pareto front based on the 67 initial points, we can say that the toll plans $\mathbf{x}_2$ and $\mathbf{x}_3$ are the two optimal solutions that the operator of the ICC may be interested in. The objective values under these two toll plans are significantly better than any other toll plans that are evaluated.

By infilling 20 points (costs about 180 computer hours) based on the expected improvement criterion, the Pareto front is augmented slightly with two points generated during the infill process. The two new Pareto set members are

$\mathbf{x}_4 = [\$\,0.36, \$\,0.54, \$\,0.76, \$\,0.33, \$\,0, 40\%\,]^T$, $TT_4 = 19.28$ min s, $TR_4 = \$6,860.7$

$\mathbf{x}_5 = [\$\,1.00, \$\,0.57, \$\,0.01, \$\,0.41, \$\,0.85, 50\%\,]^T$, $TT_5 = 23.50$ min s, $TR_5 = \$9,959.8$

However, no points that dominate the initial Pareto set members are found through the infilling. After the 20 infill iterations, the maximal expected improvement over the entire domain is about 1000, which means that there is still a very large possibility that better toll plans that dominate or augment current Pareto set members can be found through further infilling. As the purpose of this study is just to illustrate the effectiveness of the expected improvement-based infill strategy in finding new Pareto set members, further infilling is not performed here.

In summary, the expected improvement-based infill strategy is pretty effective in that two new Pareto set members are found by evaluating just 20 points. The surrogate-based optimization framework and this specific infill strategy work very well in providing optimal solutions to decision makers for the multi-objective optimization problems.

Comparing the result of multi-objective problem to that of the constrained optimization problem, we can find that the surrogate-based method with infill by the expected improvement criterion performs more efficiently in the constrained optimization problem. With less iterations of infill, a much better optimal solution is found for the constrained problem. A possible reason is that the constrained problem imposed a limit on one of the goals, which restricts the search for expected improvement in a smaller region than the case of multi-objective optimization problem. Further exploration of this issue can be done in the future. One research question we would like to investigate is whether the performance of the method will

be improved if restrictions are set for both objectives in the multi-objective optimization problem.

4.3 Conclusions

The decision making of transportation planning and operations policies always concerns about multiple system performance measures, e.g. efficiency, safety, pollution, reliability, economy, etc. To make the study more application ready, simulation-based optimization problems with multiple objectives (network average trip travel time and total toll revenue collected) considerations are formulated in this chapter for the ICC pricing optimization study introduced previously.

When there are predetermined goals regarding several of the objectives, a constrained optimization problem is formulated. Otherwise, if there is no prior knowledge on which objective is more important and no constraint on any of the objectives, a multi-objective optimization problem can be formulated.

Expected improvement based infill strategies designed separately for the constrained optimization and multi-objective optimization structures are utilized to solve the two types of problems. With a constraint that toll revenue should be higher than or equal to that collected at the baseline, an optimal solution is found for the constrained optimization problem after 15 iterations of infill, which improve the network wide average trip travel time for the peak period by 20% while satisfying the constraint. For the multi-objective optimization problem, a Pareto set with non-dominant optimal solutions based on initial sample evaluation is first generated for

decision makers' consideration. After 20 iterations of infill, 2 new points are found to augment the initial Pareto front. The effectiveness of the proposed multi-objective optimization procedures is thus proved.

Chapter 5: An Enhanced SVR Method: Adapting for Heteroscedastic Simulation Noise

This chapter further investigates the ICC dynamic pricing study mentioned in Chapter 3 and 4. Heteroscedasticity in simulation noise is observed. As this feature violates the assumptions of the surrogate models analyzed previously, a new method that can adapt for the heteroscedastic simulation noise is proposed and explored in this chapter.

5.1 Heteroscedasticity in Transportation Simulation Outputs

Using the observations of the 3-repetition evaluation of the 67 initial sample points for the ICC network introduced in Chapter 4, we can analyze the features of the simulation noise. The mean output as well as the standard deviation of the mean output for each sample point can be computed based on the observed value. Figure 5-1 plots the mean output against its corresponding standard deviation/coefficient of variation for all the 67 initial samples.
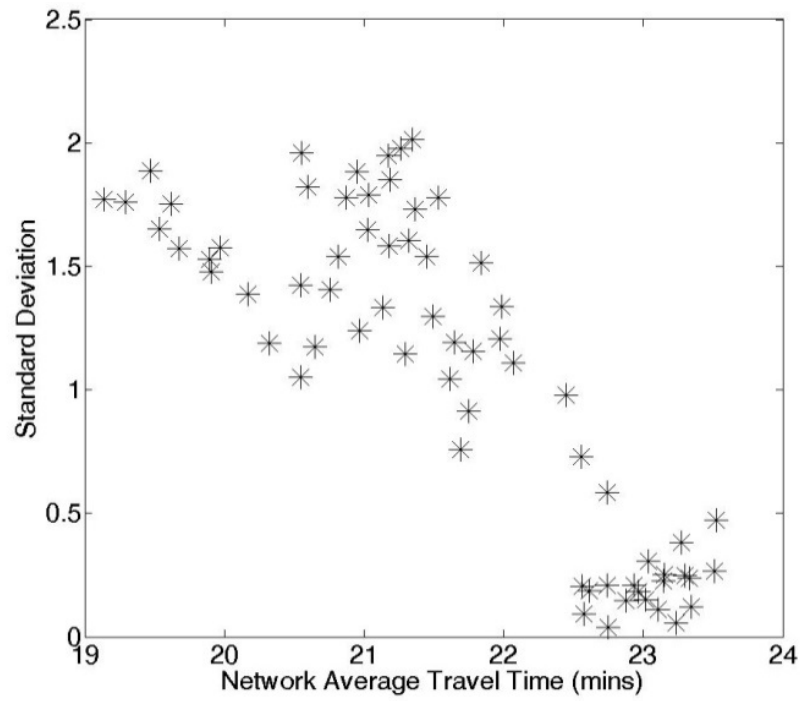
(a)
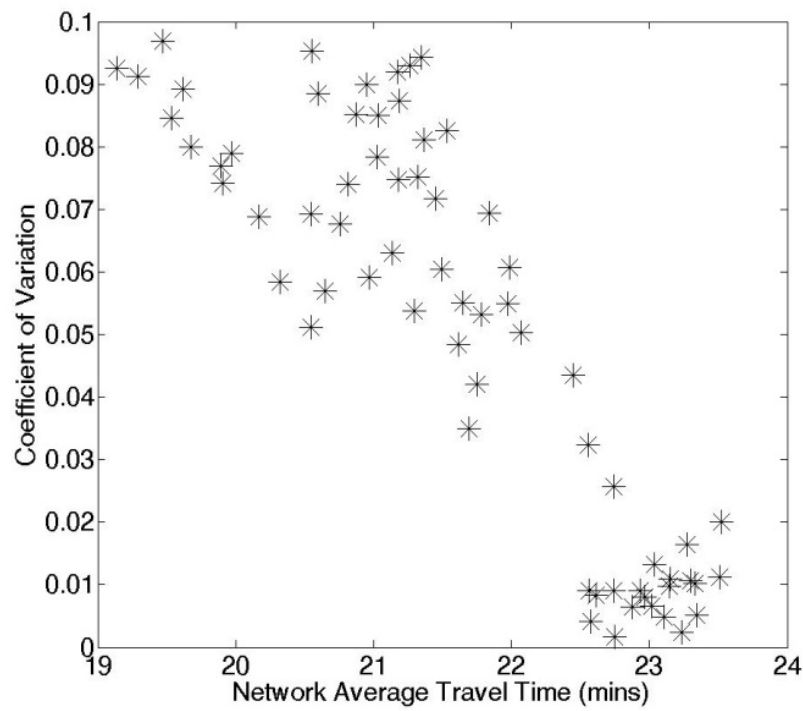
(b)

Figure 5-1: Heteroscedasticity Exhibited in the Simulation Outputs, (a) Standard Deviation vs. Mean; (b) Coefficient of Variation vs. Mean.

The standard deviation of the mean output varies significantly among those initial samples as shown in Figure 5-1(a), which means the objective function is heteroscedastic. The coefficient of variation (standard deviation divided by the mean) is a relative measure to reflect the variance level of the noise. Figure 5-1(b) plots the mean output against its corresponding coefficient of variation for all the initial samples. The level of variance differs from 0 to 0.1. Overall, the variance level is relatively low. In general, the magnitude of the simulation noise is highly dependent on the location of the design points, and the assumption of identical distribution of noise over the entire research domain in most surrogate models is not satisfied.

5.2 Methodology

SVR is a method that balances between model smoothness and the loss due to prediction error. When the variance of simulation noise is assumed to be homogenous, the error tolerance can be set to be uniform over the research domain. Otherwise, different error tolerance limit can be configured for different locations.

5.2.1 Traditional Support Vector regression (SVR)

SVR originates from the theory of support vector machine (SVM). Being different from SVM's objective of classifying sample data into different groups, SVR predicts the value at the sample points. This model provides very good compromise

between prediction accuracy and robustness of the approximations. The basic form of SVR prediction is the weighted sum of basis functions φ added to a base term b.

$$\hat{f}(\mathbf{x}) = b + \mathbf{w}^T \boldsymbol{\varphi} = b + \sum_{i=1}^{n} w^{(i)} \varphi(\mathbf{x}, \mathbf{x}^{(i)}) \tag{5.1}$$

where $\mathbf{w}$ is the vector of weights of the basis functions. The key attribute of SVR is that it allows us to specify a margin ε, and any errors in the sample data within the ε -margin are fully accepted without affecting the surrogate prediction. Under this assumption, the objective of SVR is to minimize the model complexity with the constraint that predictions at all sample points are within the ε -margin of the corresponding observation.

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \tag{5.2a}$$

$$\text{s.t.} \ -\varepsilon \le y^{(i)} - \mathbf{w}^T \boldsymbol{\varphi}^{(i)} - b \le \varepsilon, \ i = 1, 2, \ldots, n \tag{5.2b}$$

where $\boldsymbol{\varphi}^{(i)}$ is the abbreviation of $\varphi(\mathbf{x}, \mathbf{x}^{(i)})$. In the objective function, the model complexity is measured by the 2nd norm of the coefficient vector $\mathbf{w}$. $y^{(i)}$ in the constraint represents the observation at the point i. However, a solution that meets the constraints at all sample points may not exist. The constraint can be relaxed a little bit by allowing the difference between the prediction and observation to be larger than ε. As only errors smaller than ε are assumed not to affect the surrogate prediction, errors that are larger than ε should be penalized in the objective function.

A penalizing function called ε -insensitive loss function can be given by

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \le \varepsilon \\ |\xi| - \varepsilon & \text{otherwise} \end{cases} \tag{5.3}$$

where ξ is the actual error between prediction and observation. By minimizing the sum of $\frac{1}{2}\|\mathbf{w}\|^2$ and $|\xi|_\varepsilon$, a tradeoff between the model complexity and the loss due to prediction error is obtained. The updated SVR model can be expressed as

$$\min \ \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}(\xi^{+(i)}+\xi^{-(i)}) \tag{5.4a}$$

$$\text{s.t.} \ \ y^{(i)} - \mathbf{w}^\mathrm{T}\boldsymbol{\varphi}^{(i)} - b \le \varepsilon + \xi^{+(i)}, \ i=1,2,\ldots,n \tag{5.4b}$$

$$\mathbf{w}^\mathrm{T}\boldsymbol{\varphi}^{(i)} + b - y^{(i)} \le \varepsilon + \xi^{-(i)}, \ i=1,2,\ldots,n \tag{5.4c}$$

$$\xi^{+(i)}, \xi^{-(i)} \ge 0, \ i=1,2,\ldots,n \tag{5.4d}$$

The coefficient C ($C >= 0$) in the objective function governs the tradeoff between the model complexity and prediction accuracy. If C equals 0, then a flat function through μ would be generated. As C approaches infinity, the derived function would pass through all observed points. By introducing Lagrange multipliers, the above constrained optimization problem can be solved through the Lagrangian:

$$
\begin{aligned}
L = {} & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{C}{n}\sum_{i=1}^{n}(\xi^{+(i)}+\xi^{-(i)}) - \sum_{i=1}^{n}(\eta^{+(i)}\xi^{+(i)}+\eta^{-(i)}\xi^{-(i)}) \\
& - \sum_{i=1}^{n}\alpha^{+(i)}(\varepsilon + \xi^{+(i)} - y^{(i)} + \mathbf{w}^\mathrm{T}\boldsymbol{\varphi}^{(i)} + b) \\
& - \sum_{i=1}^{n}\alpha^{-(i)}(\varepsilon + \xi^{-(i)} + y^{(i)} - \mathbf{w}^\mathrm{T}\boldsymbol{\varphi}^{(i)} - b)
\end{aligned} \tag{5.5}
$$

The Lagrangian should be minimized with respect to primal variables $\mathbf{w}$, b, $\xi^{+(i)}$ and $\xi^{-(i)}$. Meanwhile, it should be maximized with respect to dual variables $\eta^{\pm(i)}$ and $\alpha^{\pm(i)}$. Meanwhile, all of the dual variables should be larger than or equal to

0. At the optimal solution, the derivatives of L with respect to the primal variables would vanish.

$$\partial_\mathbf{w} L = \mathbf{w} - \sum_{i=1}^{n} (\alpha^{+(i)} - \alpha^{-(i)})\mathbf{x}^{(i)} = 0 \tag{5.6}$$

$$\partial_b L = \sum_{i=1}^{n} (\alpha^{+(i)} - \alpha^{-(i)}) = 0 \tag{5.7}$$

$$\partial_{\xi^{+(i)}} L = \frac{C}{n} - \alpha^{+(i)} - \eta^{+(i)}, \ \ i = 1, 2, \ldots, n \tag{5.8}$$

$$\partial_{\xi^{-(i)}} L = \frac{C}{n} - \alpha^{-(i)} - \eta^{-(i)}, \ \ i = 1, 2, \ldots, n \tag{5.9}$$

By substituting Equation 5.6 into Equation 5.1, the SVR prediction can be expressed as

$$\hat{f}(\mathbf{x}) = b + \mathbf{w}^\mathrm{T}\boldsymbol{\varphi} = b + \sum_{i=1}^{n} (\alpha^{+(i)} - \alpha^{-(i)})\varphi(\mathbf{x}, \mathbf{x}^{(i)}) \tag{5.10}$$

The parameters $\alpha^{\pm(i)}$ and μ are still unknown at this stage, where $\alpha^{\pm(i)}$ can be estimated through the dual variable optimization problem.

$$\begin{aligned}
\max \quad & -\frac{1}{2}\sum_{i,j=1}^{n} (\alpha^{+(i)} - \alpha^{-(i)})(\alpha^{+(j)} - \alpha^{-(j)})\varphi(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) - \varepsilon\sum_{i=1}^{n} (\alpha^{+(i)} + \alpha^{-(i)}) \\
& + \sum_{i=1}^{n} y^{(i)}(\alpha^{+(i)} - \alpha^{-(i)})
\end{aligned} \tag{5.11a}$$

$$\text{s.t.} \ \ \sum_{i=1}^{n} (\alpha^{+(i)} - \alpha^{-(i)}) = 0 \tag{5.11b}$$

$$\alpha^{\pm(i)} \in [0, C/n], \ \ i = 1, 2, \ldots, n \tag{5.11c}$$

The maximization problem in equation 5.11 can be formulated as a quadratic programing problem in the following form.

$$\min \quad \frac{1}{2}\begin{pmatrix}\boldsymbol{\alpha}^+\\\boldsymbol{\alpha}^-\end{pmatrix}^T\begin{pmatrix}\boldsymbol{\Psi} & -\boldsymbol{\Psi}\\-\boldsymbol{\Psi} & \boldsymbol{\Psi}\end{pmatrix}\begin{pmatrix}\boldsymbol{\alpha}^+\\\boldsymbol{\alpha}^-\end{pmatrix}+\begin{pmatrix}\mathbf{1}^T\varepsilon-\mathbf{y}\\\mathbf{1}^T\varepsilon+\mathbf{y}\end{pmatrix}^T\begin{pmatrix}\boldsymbol{\alpha}^+\\\boldsymbol{\alpha}^-\end{pmatrix} \tag{5.12a}$$

$$\text{s.t.} \quad \mathbf{1}^T\begin{pmatrix}\boldsymbol{\alpha}^+\\-\boldsymbol{\alpha}^-\end{pmatrix}=0 \tag{5.12b}$$

$$\boldsymbol{\alpha}^+,\boldsymbol{\alpha}^- \in[\mathbf{0},\mathbf{C}/\mathrm{n}] \tag{5.12c}$$

where $\boldsymbol{\Psi}$ is the $n\times n$ matrix for the basis functions, with elements to be $\varphi(\mathbf{x}^{(i)},\mathbf{x}^{(j)})$.

According to the KKT conditions, the base term b can be computed as

$$\hat{b}=\begin{cases}y^{(i)}-\mathbf{w}^T\mathbf{x}^{(i)}+\varepsilon & \text{for } \alpha^{+(i)}=0,\alpha^{-(i)}\in(0,C/n)\\y^{(i)}-\mathbf{w}^T\mathbf{x}^{(i)}-\varepsilon & \text{for } \alpha^{+(i)}\in(0,C/n),\alpha^{-(i)}=0\end{cases} \tag{5.13}$$

In the current study, we use Gaussian function form for the basis function φ. The

choices of the parameter σ^2 (the Gaussian kernel variance) and the weight coefficient

C are determined by cross validation.

5.2.2 An Enhanced SVR

For the traditional SVR model, the margin ε is set to be a constant for all

sample points, which means that the tolerance of the error is the same throughout the

entire domain. However, this is definitely not the case for heteroscedastic samples.

Because observations at points with larger variance would have larger confidence

intervals, at the same confidence level, larger errors should be tolerated at those

corresponding points. If there is no prior knowledge of the variance of noise in the

data, the margin ε can be calculated by using $\nu-\text{SVR}$ (Schölkopf and Smola, 2002).

When prior knowledge regarding the noise level is available, ε is usually chosen as the standard deviation of the noise (Forrester and Keane, 2009). Thus for heteroscedastic samples, the margin ε at each point can be chosen as the corresponding standard deviation of the noise at that point. The SVR model can then be revised as follows to adapt for heteroscedasticity in sample data. This enhanced SVR model will be named as heteroscedastic SVR model (H-SVR) hereafter. In equation 5.4, the constant margin ε in the first two constraints will be replaced by the unique margin $\varepsilon^{(i)}$ corresponding to each sample point i. The parameters $\alpha^{\pm(i)}$ can then be computed through solving the quadratic program 5.14.

$$\min \quad \frac{1}{2}\begin{pmatrix}\boldsymbol{\alpha}^+\\\boldsymbol{\alpha}^-\end{pmatrix}^T\begin{pmatrix}\boldsymbol{\Psi} & -\boldsymbol{\Psi}\\-\boldsymbol{\Psi} & \boldsymbol{\Psi}\end{pmatrix}\begin{pmatrix}\boldsymbol{\alpha}^+\\\boldsymbol{\alpha}^-\end{pmatrix}+\begin{pmatrix}\boldsymbol{\varepsilon}-\mathbf{y}\\\boldsymbol{\varepsilon}+\mathbf{y}\end{pmatrix}^T\begin{pmatrix}\boldsymbol{\alpha}^+\\\boldsymbol{\alpha}^-\end{pmatrix} \tag{5.14a}$$

$$\text{s.t.} \quad \mathbf{1}^T\begin{pmatrix}\boldsymbol{\alpha}^+\\-\boldsymbol{\alpha}^-\end{pmatrix}=0 \tag{5.14b}$$

$$\boldsymbol{\alpha}^+,\boldsymbol{\alpha}^- \in [\mathbf{0},\mathbf{C}/\mathrm{n}] \tag{5.14c}$$

where $\boldsymbol{\varepsilon}$ is a $n \times 1$ vector with the elements to be $\varepsilon^{(i)}$, which is the standard deviation of noise at point i. The base term μ can be calculated through equation 5.15, which is revised from equation 5.13 accordingly.

$$\hat{b}=\begin{cases}y^{(i)}-\mathbf{w}^T\mathbf{x}^{(i)}+\varepsilon^{(i)} & \text{for } \alpha^{+(i)}=0, \alpha^{-(i)}\in(0,C/n)\\ y^{(i)}-\mathbf{w}^T\mathbf{x}^{(i)}-\varepsilon^{(i)} & \text{for } \alpha^{+(i)}\in(0,C/n), \alpha^{-(i)}=0\end{cases} \tag{5.15}$$

Figure 5-2 shows the difference between the H-SVR model and the traditional SVR model through a one-variable function, and this figure is created mainly for illustrative purpose. The variance of the function noise is assumed to be heteroscedastic.
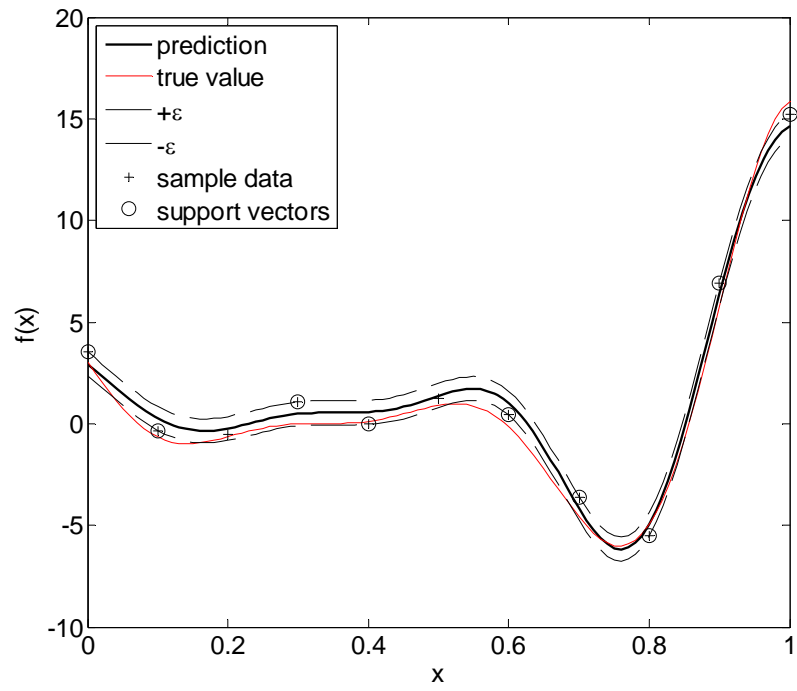
(a) traditional SVR model
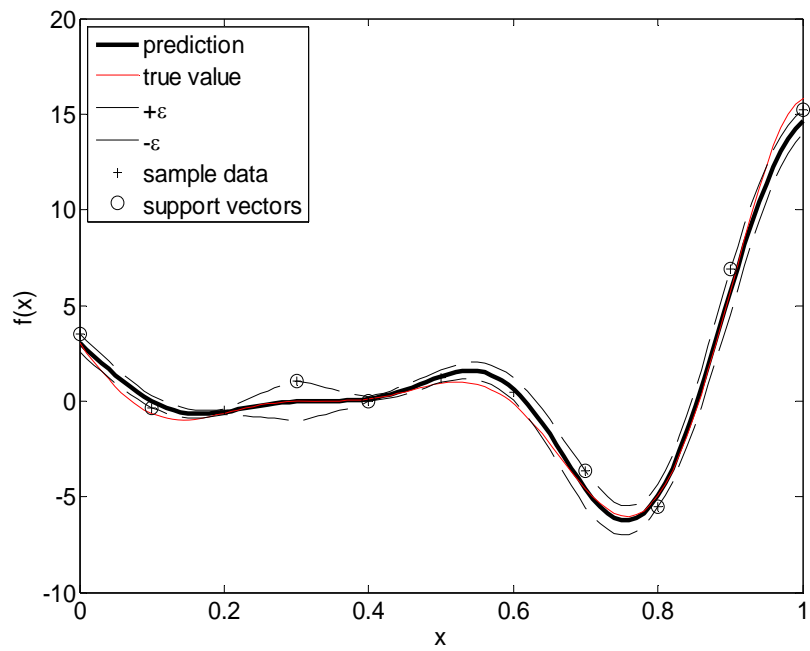
(b) heteroscedastic SVR model

Figure 5-2: Comparison of the Traditional and the Heteroscedastic SVR Model, (a) Traditional SVR Model; (b) Heteroscedastic SVR Model.

It clearly shows that the errors that can be tolerated at different locations vary significantly, and the ε-tube is no longer parallel to the predicted function. The advantage of the H-SVR model in terms of the overall prediction accuracy will be investigated through numerical test functions in the following section.

5.3 Numerical Tests

Different forms of surrogate models including quadratic polynomial function, radial basis function, Kriging model and SVR have been investigated in the numerical study in Chapter 3. The objective function of the numerical study is computed by solving a static user equilibrium model. In that case, the output is in fact deterministic instead of stochastic. It was found that the regressing Kriging model performs best in terms of prediction accuracy.

As the focus of this study is to investigate the capability of the H-SVR model in dealing with heteroscedastic sample data, the performance of the H-SVR model is compared to that of the regressing Kriging model in approximating several mathematical functions with heteroscedastic noise.

The first test problem is a two-dimension function

$$f_1(x_1, x_2) = x_1 \sin(x_1) + x_2 \sin(x_2), \quad x_1, x_2 \in [-2\pi, 2\pi] \tag{5.16}$$

with global optimal solutions of $\mathbf{x}^* = \pm(4.9132, 4.9132)^{\mathrm{T}}$, $f_1(\mathbf{x}^*) = -9.6289$.

In order to introduce heteroscedastic noise into the function output, the observation at a certain point is set to be

$$\tilde{f}_1(x_1, x_2) = f_1(x_1, x_2) + 0.1 f_1(x_1, x_2) \bullet N(0,1), \quad x_1, x_2 \in [-2\pi, 2\pi] \qquad (5.17)$$

$$\widehat{f}_1(x_1, x_2) = f_1(x_1, x_2) + 0.5 f_1(x_1, x_2) \bullet N(0,1), \quad x_1, x_2 \in [-2\pi, 2\pi] \qquad (5.18)$$

where $N(0,1)$ represents a normal distribution with mean 0 and standard deviation 1. Equation 5.17 creates samples with a relatively low level of noise while Equation 5.18 creates samples with a relatively high level of noise. As the variance of noise is dependent on the magnitude of true function value at each point, the samples created from these two equations exhibit heteroscedasticity. For Equation 5.17, the standard deviation of the output at point (x_1, x_2) is $0.1 f_1(x_1, x_2)$, while the standard deviation would be $0.5 f_1(x_1, x_2)$ for Equation 5.18.

To investigate the impact of sample size on the prediction accuracy, two different samples are generated using LHS separately, with a small sample including 25 points and a large sample including 60 points. Prior knowledge regarding the variance of noise at each point is needed by the H-SVR model. As in real world applications, evaluation through simulation usually costs significant amount of time, we mimic the situation with high computational cost and constrain the number of evaluation for these simple test functions. Each point in the samples by either Equation 5.17 or Equation 5.18 is allowed to be evaluated three times to calculate the mean and variance of the corresponding output. The mean output, the sample variance of the mean output and the location of the corresponding points will then be used to construct the surrogate models. As each point is only evaluated three times, the computed mean and variance may not be reliable estimates of the real distribution attributes. To achieve a robust comparison between the performance of the H-SVR model and the regressing Kriging model, 10 random seeds are used for the three-time

evaluation of each sample point. Thus for both the small sample and the large sample, the whole process including creation of 3-replication dataset, surrogate model estimation and validation is repeated 10 times.

After the surrogate models are constructed, an independent denser validation data set with 576 points is generated, and the true function value at each of the validation point is evaluated through Equation 5.16. Predictions at the validation points will be obtained using the surrogate models. Four MoEs are chosen to evaluate the performance of the two types of surrogate models, including RMSE, MAE and EGOp that are introduced in section 3.3, and a new MoE named absolute error of optimal value (AEOV).

$$\text{AEOV} = \left| f(\hat{\mathbf{x}}^*) - f^* \right|, \ \hat{\mathbf{x}}^* = \arg\min \hat{f}(\mathbf{x}), \ f^* = \min f(\mathbf{x}) \quad (5.19)$$

where f is the true function, $\hat{f}$ is the surrogate function, $\hat{\mathbf{x}}^*$ is the predicted global optima, and f^* is the true optimal value. After the MoEs for each surrogate model is computed, the average performance for the 10-time model estimation and validation process is then compared between the H-SVR model and the regressing Kriging model.

To investigate the impact of problem dimension on the performance of two types of models, we test a second problem, which is a six-dimension function, belonging to the Hartman's families.

$$f_2(\mathbf{x}) = -\sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{6} a_{ij}(x_j - p_{ij})^2 \right), \ \mathbf{x} \in \mathbb{R}^6, 0 \le x_j \le 1 \quad (5.20)$$

Similarly, the outputs with different level of noise variance are generated by equation 5.21 and 5.22.

$$\tilde{f}_2(\mathbf{x}) = f_2(\mathbf{x}) + 0.1 f_2(\mathbf{x}) \cdot N(0,1), \quad \mathbf{x} \in \mathbb{R}^6, 0 \le x_j \le 1, 1 \le j \le 6 \qquad (5.21)$$

$$\hat{f}_2(\mathbf{x}) = f_2(\mathbf{x}) + 0.5 f_2(\mathbf{x}) \cdot N(0,1), \quad \mathbf{x} \in \mathbb{R}^6, 0 \le x_j \le 1, 1 \le j \le 6 \qquad (5.22)$$

The parameters for this six-dimension test function are as follows:

$$a_{ij} = \begin{bmatrix} 10.0 & 3.0 & 17.0 & 3.50 & 1.7 & 8.0 \\ 0.05 & 10.0 & 17.0 & 0.10 & 8.0 & 14.0 \\ 3.0 & 3.50 & 1.70 & 10.0 & 17.0 & 8.0 \\ 17.0 & 8.0 & 0.05 & 10.0 & 0.1 & 14.0 \end{bmatrix} \qquad (5.23)$$

$$c_i = \begin{bmatrix} 1.0 \\ 1.2 \\ 3.0 \\ 3.2 \end{bmatrix} \qquad (5.24)$$

$$p_{ij} = \begin{bmatrix} 0.1312 & 0.1696 & 0.5569 & 0.0124 & 0.8283 & 0.5886 \\ 0.2329 & 0.4135 & 0.8307 & 0.3736 & 0.1004 & 0.9991 \\ 0.2348 & 0.1451 & 0.3522 & 0.2883 & 0.3047 & 0.6650 \\ 0.4047 & 0.8828 & 0.8732 & 0.5743 & 0.1091 & 0.0381 \end{bmatrix} \qquad (5.25)$$

The global optima of this function is

$$\mathbf{x}^* = (0.20169, 0.150011, 0.47687, 0.275332, 0.311652, 0.6573), \ f(\mathbf{x}^*) = -3.32237.$$

For the second test function, the small sample includes 65 points and the large sample includes 380 points. All the procedures of constructing the surrogate models and evaluating the performance of the surrogate models are the same as the case of the first test function, except that the independent validation sample for this function includes 729 points.

Overall, three aspects of the problems are considered for the numerical test, which are problem dimension, sample size and variance level. Thus there are in total 8 scenarios to be investigated. The average performance of the two types of surrogate

models under different test functions, different sample sizes and different variance levels is listed in Table 5-1.

The better performance regarding the four MoEs under each scenario is highlighted in bold in Table 5-1. In general, the MoE RMSE evaluates the overall prediction accuracy of the two surrogate models over the entire research domain, while MAE, EGOp and AEOV measures their capability of estimating the function value accurately at some specific small regions. For the current numerical study, as the validation sample is selected using the space filling algorithm LHS and no evolutionary process (e.g. global optimal infill) is performed, the MoE RMSE evaluating the performance over the entire domain are of more concern.

Test Function	Sample Size	Variance Level	Heteroscedastic SVR Model				Regressing Kriging Model			
			RMSE	MAE	EGOp	AEOV	RMSE	MAE	EGOp	AEOV
2-variable function	small	low	**2.33**	**7.10**	**-7.89**	1.81	2.42	7.26	-7.89	**0.88**
		high	2.68	7.62	-8.03	3.38	**2.55**	**7.40**	**-9.40**	**2.41**
	large	low	**0.80**	**7.00**	**-9.09**	**0.07**	0.98	7.10	-7.43	3.99
		high	**1.52**	**5.17**	-6.22	**1.77**	1.56	5.82	**-7.58**	3.31
6-variable function	small	low	**0.24**	**1.90**	**-2.01**	0.85	0.29	1.94	-1.83	**0.62**
		high	0.49	2.49	-1.25	**1.34**	**0.30**	**2.08**	**-1.82**	1.87
	large	low	**0.14**	**1.31**	-3.02	**0.48**	0.18	1.65	**-3.14**	0.49
		high	0.44	2.14	-1.73	1.01	**0.26**	**1.70**	**-2.38**	**0.98**

Table 5-1: Performance of the Heteroscedastic SVR Model and Regressing Kriging Model.

It's very clear from Table 5-1 that the H-SVR model always outperforms the regressing Kriging model in terms of overall prediction accuracy when the variance level of the data is relatively low. For the four low-variance scenarios, ANOVA is

conducted to test the statistical significance of the difference between the RMSE predicted by the H-SVR model and the regressing Kriging model. The p value is computed to be 0.000, 0.015, 0.025 and 0.000 for the 2-dimension small-sample, 2-dimension large-sample, 6-dimension small-sample and 6-dimension large-sample cases, respectively. The differences in RMSE for the two models in all of the four low-variance cases are all significant at the 95% confidence level. In general, the increase of sample size would enhance the performance of both surrogate models under the low variance cases.

Worse performance should be expected when the variance level is increased. However, the comparison between the two surrogate models becomes more complicated if the variance level is relatively high. Each of the two surrogate models does not show consistent better performance than the other. An interesting finding is that the performance of the H-SVR model varies significantly among the 10 sets of input-output data for the same scenario, while the regressing Kriging model provides quite consistent performance using the same 10 sets of data. The RMSEs for both of the two surrogates using 10 sets of data under the case of 6-variable function with a high variance level are plotted in Figure 5-3. The H-SVR model is shown to be rather sensitive to the generated random outputs of the samples. Thus it's not a promising option to deal with problems with a high variance level.

(a)

(b)
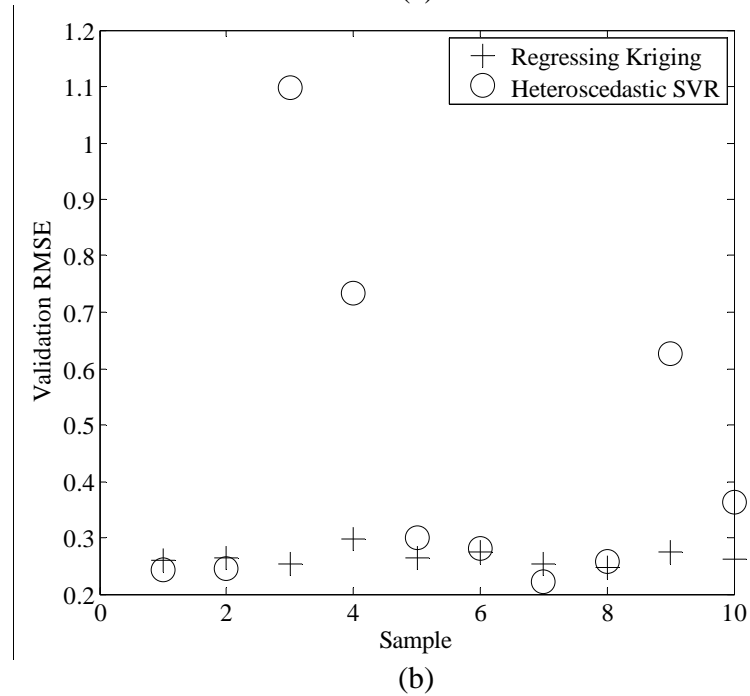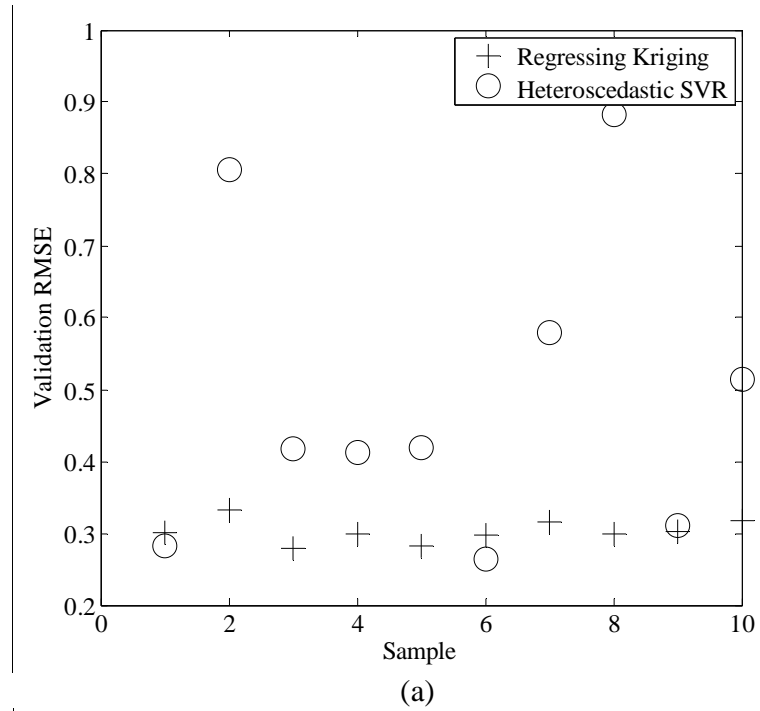
Figure 5-3: Validations RMSE for the 6-Variable Function with a High Variance
Level, (a) Small Sample; (b) Large Sample.

112

5.4 Application

The real world application to be investigated for this chapter is still the ICC dynamic pricing optimization problem. And the objective of the problem is to minimize network wide average trip travel time for vehicles departing during the morning peak period (6-9 am). The variance level of the ICC simulation output has been analyzed in section 5.1. Figure 5-1 shows that the coefficient of variation of the outputs for all the initial samples differs from 0 to 0.1. According to the conclusion from the numerical test, the H-SVR model outperforms the regressing Kriging model for heteroscedastic data when the variance level is low. Thus the H-SVR model should be suitable for solving this real world application problem.

Based on the 3-time evaluation of the 67 initial samples, both the H-SVR model and the regressing Kriging model are constructed, and their capability of predicting optimal solutions is compared.

The optimal solution found by the H-SVR model is $\hat{\mathbf{x}}^*_{SVR} = [\$\,0.30, \$\,0.37, \$\,0.00, \$\,0.00, \$\,0.00, 1.5\%]^{\mathrm{T}}$, and the estimated network average travel time under this toll plan is $\widehat{TT}^*_{SVR} = 19.35 \min s$. This toll plan is then evaluated through 10-time DynusT simulation runs, and the average of the network mean travel time is $TT^*_{SVR} = 18.37 \min s$, and the standard deviation is $\sigma^*_{SVR} = 0.35 \min s$

The regressing Kriging model is also constructed to approximate the response surface. However, after an extensive search for global optima based on the Kriging model using the genetic algorithm, the global optima is found to lie at the boundary

domain, with the toll rate of all the ICC sections to be 0. The estimated network mean

travel time is $\widehat{TT}^*_{Kriging} = 20.19 \min s$. The actual network average travel time under 0

tolls is $TT^*_{Kriging} = 19.16 \min s$ according to the 10-time simulation evaluation at this

point, and the standard deviation is $\sigma^*_{Kriging} = 0.29 \min s$.

We perform the ANOVA for the 10-time simulation results of the optima predicted by the H-SVR and the regressing Kriging models, respectively. The F statistic is 3.40 and the corresponding p value is 0.08, which means that the network wide average trip travel times in response to the two optimal toll plans predicted by different surrogate models are statistically different at the 90% confidence level.

Figure 5-4 shows the variations of traffic volume on the five tolled segments of the ICC during the peak period under the two optimal toll plans predicted by the H-SVR and the regressing Kriging model as well as the baseline toll. Traffic flow for different directions is depicted separately. The traffic volume on the ICC sections is very low under the baseline toll plan. It's almost always lower than 1,000 vehicles per hour. As the ICC has 3 lanes per direction, the capacity is definitely not used sufficiently. Traffic volume under the two optimal toll plans is much higher on all of the 5 segments than the baseline case. This change is expected since the toll rate on all of the segments is reduced. It's not surprising that the traffic volume is the highest when the ICC is totally free.
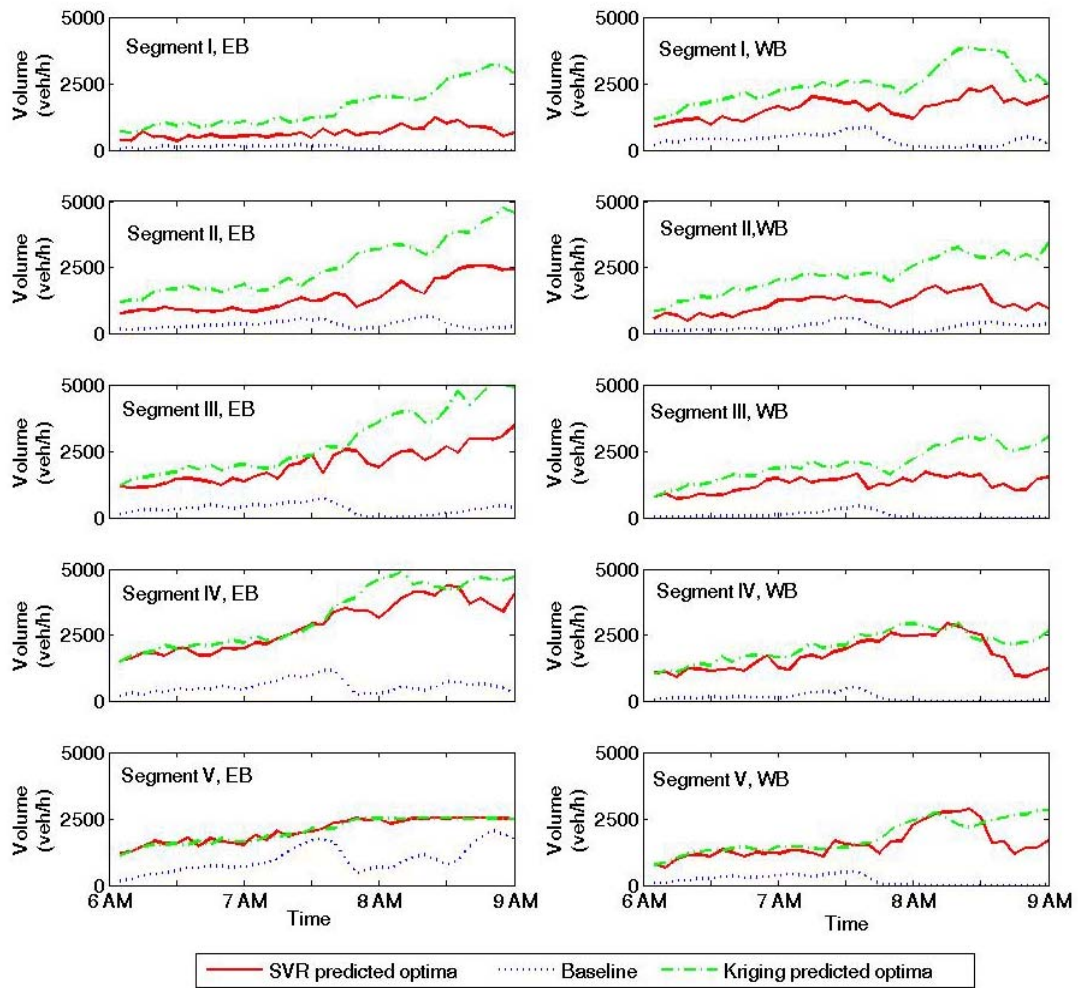
Figure 5-4: Traffic Volumes of ICC Segments under Optima and Baseline Tolls.

In addition to the comparison of traffic volume on the tolled links, the network level performance under the 3 toll plans is also analyzed. The average travel time for vehicles departing within every 5-minute interval is depicted in Figure 5-5. The average travel time is significantly reduced under both the optimal toll plan predicted by the regressing Kriging model and the H-SVR model, especially during the most congested period (7:30-9:00 am). The reduction of average travel time is around 30% during this period. Although the optimal solution predicted by the regressing Kriging

115

model makes the ICC road free, and introduces a lot of traffic onto the ICC, the network wide traffic condition is not as efficiently improved as the solution given by the H-SVR model. A possible reason is that most of the traffic diverted onto the ICC is short trips, and these trips do not contribute a lot to the network wide performance. Those long trips are not attracted to the ICC and stay jammed on their usual routes.
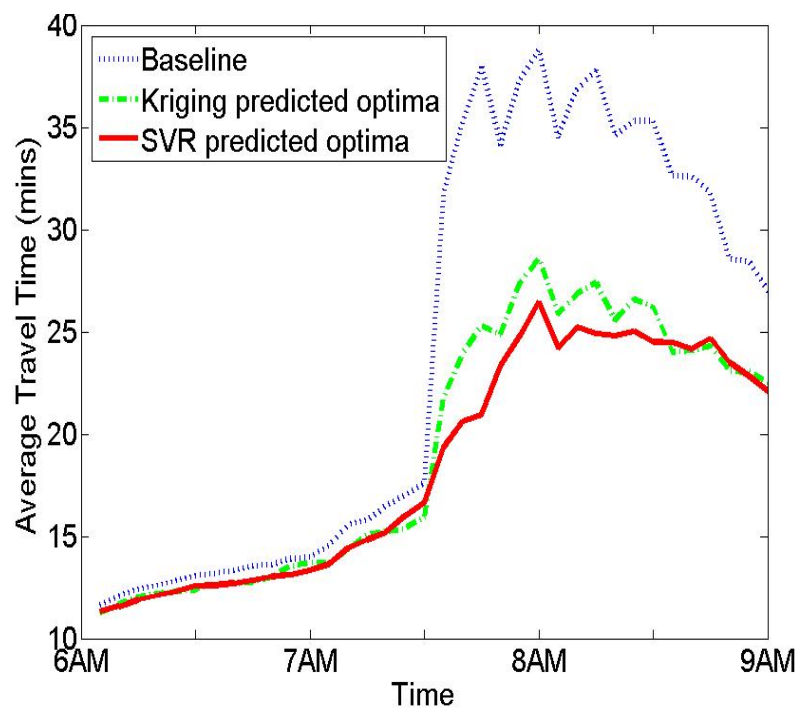
Figure 5-5: Average Travel Time under Optima and Baseline Tolls.

As the traffic condition is improved during the most congested period (7:30-9:00 am) under the two optimal toll plans, more vehicles are able to arrive at their destination during the peak period. The throughput for the whole network increases significantly for the period 7:30-9:00 am, and the total throughput for the morning peak period increases by about 11% under the optimal toll predicted by the regressing Kriging model. The increase in throughput is around 14% for the optimal

116

toll predicted by the H-SVR model. The adjustment of toll rates for one freeway in the regional network has successfully led to a more efficient usage of the road network capacity. As the average travel time is reduced and the overall throughput is increased at the same time, the total amount of travel time saving for all the network users during the morning peak hours is very huge.

(b) Vehicle Throughput   (c) Cumulative Vehicle Throughput
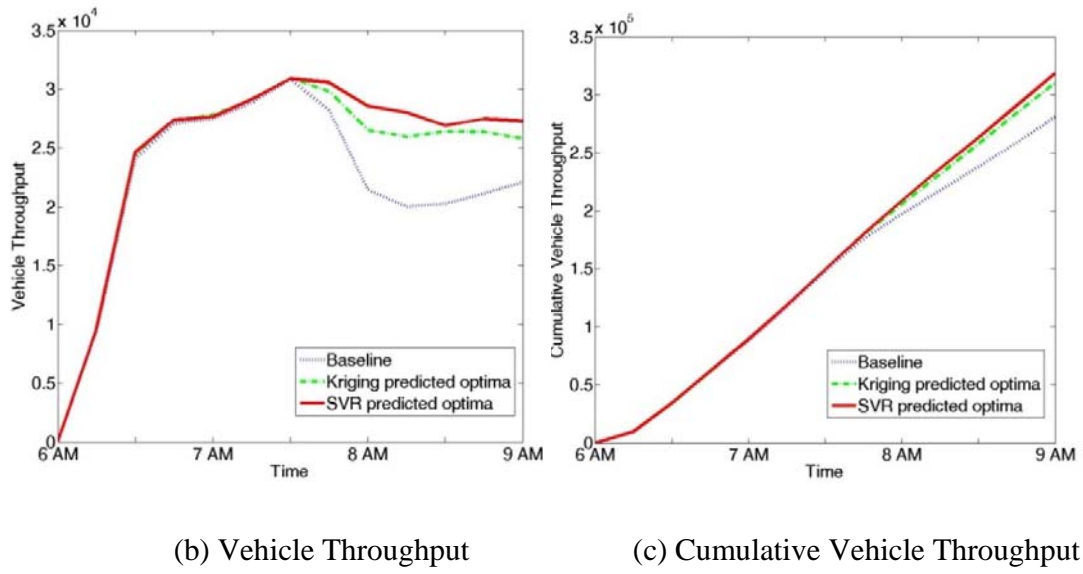
Figure 5-6: Throughput under Optima and Baseline Tolls, (a) Vehicle Throughput; (b) Cumulative Vehicle Throughput.

5.5 Conclusions

Most existing surrogate models that can approximate expensive-to-evaluate simulation response surfaces implicitly assume identical distribution of simulation noise over the entire feasible domain. However, via the observation of the simulation output of the regional network for our case study, we find that the noise of the

transportation simulation exhibits very significant heteroscedasticity. To take into account heteroscedastic noise in the sample data, an enhanced SVR method (H-SVR) is proposed in this chapter, which allows different levels of error tolerance throughout the feasible domain according to the level of variance at the specific locations.

To test the effectiveness of using the H-SVR model to approximate heteroscedastic simulation output, a complete set of numerical test is designed, which includes 8 scenarios characterized by problem dimension, variance level and sample size. The performance of the H-SVR model is then compared to that of the regressing Kriging model. The test results show that the H-SVR model outperforms the regressing Kriging model in terms of overall prediction accuracy when the variance level of the noisy data is relatively low. However, when the level of variance in the data is relatively high, both of the two surrogates cannot provide satisfying performance.

The H-SVR model is then applied to the application of optimizing the dynamic pricing of the ICC road as introduced in Chapter 4. The objective of this problem is set to be minimizing network wide average trip travel time for vehicles departing during the morning peak period. The optimal toll plan predicted by the H-SVR model is found to outperform that predicted by the regressing Kriging model (best surrogate model identified in the study introduced in Chapter 3). Overall, under the optimal toll plan predicted by the H-SVR model, the network wide average trip travel time for the peak period is reduced by around 20% compared to the baseline case, and the total vehicle throughput is increased by 14%.

However, no updating process is introduced when constructing the H-SVR models in both the numerical study and the real world application. The reason is that SVR models can only provide predicted value at unevaluated design points. Information such as the variance of the prediction at a particular point is not available. Therefore, the global optimal infill strategies used in Kriging models cannot be applied. In this case, the capability of the H-SVR surrogate model in finding the true optimal solution (or near-optimal solution) is heavily dependent on the choice of initial samples. This is the main limitation of the proposed H-SVR method.

Chapter 6: Bayesian Stochastic Kriging for Simulations with Heteroscedastic Errors

To deal with the heteroscedastic noise in transportation simulation outputs, we propose another surrogate model in this chapter, which is named Bayesian stochastic Kriging model. Similar to the ordinary Kriging, this model can provide prediction of the distribution of output at any unevaluated design point. Thus the global optimal infill strategies introduced in section 3.2.3 can be easily applied. Moreover, to further illustrate the power of the simulation-based optimization approach in transportation research, a new problem of integrated optimization of transportation planning and operations decisions is developed as the case study in this chapter, which is very difficult to be formulated as an analytical model and thus is very hard to be solved using pure mathematical methods.

6.1 Bayesian Stochastic Kriging Model

To account for both sampling and response-surface uncertainty that is inherent to a stochastic simulation, stochastic Kriging method can be incorporated with heteroscedastic noisy data. Ankenman et al. (2010) provides a mathematical foundation for stochastic Kriging that extends the power of Kriging metamodeling for deterministic computer experiments to modeling responses from heteroscedastic, stochastic simulations. Although the ordinary Kriging treats the unknown response surface as a Gaussian random field that exhibits spatial correlation, the method

assumes measurement errors to be independent and identically distributed, while stochastic Kriging accounts for unequal variances and correlation of the random errors across the design space in stochastic simulation (Chen et al., 2012). In this chapter, we enhance the stochastic Kriging model by incorporating parameter uncertainties. The proposed Bayesian stochastic Kriging model is derived in a Bayesian analysis framework which endeavors to estimate parameters of an underlying distribution-based on the observed distribution.

6.1.1 Model Formulation

Suppose the unknown true response at point $\mathbf{x}$ is stochastic and following a probabilistic distribution $\mathbf{g}(\mathbf{x})$, and the mathematical expectation of the true response satisfies $\mathrm{E}[\mathbf{g}(\mathbf{x})] = \mathbf{f}(\mathbf{x})$, where $\mathbf{f}$ is the true mean response surface. The expected square prediction error at $\mathbf{x}$ is $\mathrm{E}[(\mathbf{y} - \mathbf{g}(\mathbf{x}))^2]$.

The stochastic Kriging method predicts a response by summarizing a linear model and a high frequency variation component that represents fluctuations around the trend. In this study, we consider the following stochastic Kriging model

$$y_r(\mathbf{x}) = \mathbf{q}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}) + \varepsilon_r(\mathbf{x}), \quad r = 1, 2 \ldots \tag{6.1}$$

where $y_r(\mathbf{x})$ is the observed response obtained from the r th simulation replication at point $\mathbf{x}$, $\mathbf{q}(\mathbf{x}) = [q_1(\mathbf{x}), \ldots, q_m(\mathbf{x})]^{\mathrm{T}}$ is an $m \times 1$ vector of known regression basis functions (e.g. polynomial functions) of $\mathbf{x}$, and $\boldsymbol{\beta}$ is an $m \times 1$ vector of unknown weight parameters of each regression basis function. The stochastic nature of $Z(\mathbf{x})$ is referred as extrinsic uncertainty (Ankenman et al., 2010) because it is imposed to

reconstruct the metamodel. $\varepsilon_r(\mathbf{x})$ is the intrinsic uncertainty that is only associated with simulation random noises. A series of simulation replications $\{\varepsilon_1(\mathbf{x}), \varepsilon_2(\mathbf{x}), \ldots\}$ at the same point $\mathbf{x}$ are assumed to be independent and identically distributed.

The Kriging surrogate approach assumes that the joint distribution of $f_0 = f(\mathbf{x}^{(0)})$ and $\mathbf{f} = [f(\mathbf{x}^{(1)}), f(\mathbf{x}^{(2)}), \ldots, f(\mathbf{x}^{(n)})]^{\mathrm{T}}$ is a $(n+1)$-multivariate Gaussian distribution given by

$$\begin{pmatrix} f_0 \\ \mathbf{f} \end{pmatrix} \sim \mathcal{N}_{1+n}\left[\begin{pmatrix} \mathbf{q}_0^{\mathrm{T}} \\ \mathbf{Q} \end{pmatrix}\boldsymbol{\beta}, \sigma_z^2 \begin{pmatrix} 1 & \boldsymbol{\psi}^{\mathrm{T}} \\ \boldsymbol{\psi} & \boldsymbol{\Psi} \end{pmatrix} \right] \tag{6.2}$$

where $\mathbf{q}_0^{\mathrm{T}} = \mathbf{q}^{\mathrm{T}}(\mathbf{x}^{(0)})$ is an $1 \times m$ vector of regressors at $\mathbf{x}^{(0)} \in \mathbb{R}^k$. Take the $n \times m$ matrix of regression basis functions as $\mathbf{Q} = [\mathbf{q}(\mathbf{x}^{(1)}), \mathbf{q}(\mathbf{x}^{(2)}), \ldots, \mathbf{q}(\mathbf{x}^{(n)})]^{\mathrm{T}}$, the (i,j)th element of which is $q_j(\mathbf{x}^{(i)})$ for $1 \le i \le n$, $1 \le j \le m$.

To reduce the influence of simulation random noises, multiple replications of simulations are necessary for each design point $\mathbf{x}^{(i)}$, because the response mean and variance are given by

$$\overline{y}(\mathbf{x}^{(i)}) = \frac{1}{R_i} \sum_{r=1}^{R_i} y_r(\mathbf{x}^{(i)}), \quad i = 1, \ldots, n \tag{6.3}$$

$$s_\varepsilon^2(\mathbf{x}^{(i)}) = \frac{1}{R_i - 1} \sum_{r=1}^{R_i} (y_r(\mathbf{x}^{(i)}) - \overline{y}(\mathbf{x}^{(i)}))^2, \quad i = 1, \ldots, n \tag{6.4}$$

where $\overline{y}(\mathbf{x}^{(i)})$ is the sample mean at $\mathbf{x}^{(i)}$, R_i is the number of simulation replications at $\mathbf{x}^{(i)}$, the mean response vector is $\overline{\mathbf{y}} = [\overline{y}(\mathbf{x}^{(1)}), \overline{y}(\mathbf{x}^{(2)}), \ldots, \overline{y}(\mathbf{x}^{(n)})]^{\mathrm{T}}$, $s_\varepsilon^2(\mathbf{x}^{(i)})$ is the unbiased sample variance of R_i simulation replications. So the variance estimation of the mean response is $s^2(\overline{y}(\mathbf{x}^{(i)})) = s_\varepsilon^2(\mathbf{x}^{(i)}) / R_i$.

Similar to the $n \times n$ extrinsic covariance matrix of $\sigma_z^2 \boldsymbol{\Psi}$, let $\boldsymbol{\Sigma}_\varepsilon$ be the $n \times n$ intrinsic covariance matrix implied by the simulation noise. To simplify the covariance structure of simulation noises and keep its heteroscedasticity feature, the estimation of $\boldsymbol{\Sigma}_\varepsilon$ can be given by

$$\hat{\boldsymbol{\Sigma}}_\varepsilon = \begin{pmatrix} s_\varepsilon^2(\mathbf{x}^{(1)})/R_1 & 0 & \cdots & 0 \\ 0 & s_\varepsilon^2(\mathbf{x}^{(2)})/R_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_\varepsilon^2(\mathbf{x}^{(n)})/R_n \end{pmatrix} \tag{6.5}$$

Lemma 1 (Conditional distribution of the multivariate Gaussian distribution). *Suppose that $[f_0, \mathbf{f}]^{\mathrm{T}}$ follow the $(n+1)$ dimensional multivariate Gaussian distribution $\mathcal{N}_{1+n}$, i.e.*

$$\begin{pmatrix} f_0 \\ \mathbf{f} \end{pmatrix} \sim \mathcal{N}_{n+1}\left[\begin{pmatrix} \mathbf{q}_0^{\mathrm{T}} \\ \mathbf{Q} \end{pmatrix} \boldsymbol{\beta}, \sigma_z^2 \begin{pmatrix} 1 & \boldsymbol{\psi}^{\mathrm{T}} \\ \boldsymbol{\psi} & \boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2 \end{pmatrix} \right] \tag{6.6}$$

then the conditional distribution of f_0 given $\mathbf{f}$ is an univariate Gaussian distribution $\mathcal{N}_1$ given by

$$f_0 \mid \mathbf{f} \sim \mathcal{N}_1\left[\mathbf{q}_0^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}(\mathbf{f} - \mathbf{Q}\boldsymbol{\beta}), \sigma_z^2[1 - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\boldsymbol{\psi}] \right] \tag{6.7}$$

Proof. Consider the $(n+1) \times (n+1)$ matrix $\boldsymbol{\Sigma} = \sigma_z^2 \begin{pmatrix} 1 & \boldsymbol{\psi}^{\mathrm{T}} \\ \boldsymbol{\psi} & \boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2 \end{pmatrix}$, according to the Sherman-Morrison-Woodbury matrix inverse formula (Woodbury, 1950), we have

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_z^2}\begin{pmatrix} (1 - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\boldsymbol{\psi})^{-1} & -\boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2 - \boldsymbol{\psi}\boldsymbol{\psi}^{\mathrm{T}})^{-1} \\ -(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\boldsymbol{\psi}(1 - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\boldsymbol{\psi})^{-1} & (\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2 - \boldsymbol{\psi}\boldsymbol{\psi}^{\mathrm{T}})^{-1} \end{pmatrix} \tag{6.8}$$

To diagonalize the covariance matrix $\boldsymbol{\Sigma}$ of the multivariate Gaussian distribution, we conduct the LDU decomposition (Eaton, 1983) as follows

$$\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Lambda} = \sigma_z^2 \begin{pmatrix} 1 - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\boldsymbol{\psi} & 0 \\ 0 & \boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2 \end{pmatrix} \tag{6.9}$$

where $\boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 \\ -(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\boldsymbol{\psi} & 1 \end{pmatrix}$.

Since that $[f_0, \mathbf{f}]^{\mathrm{T}}$ follow the multivariate Gaussian distribution, we premultiply $\boldsymbol{\Lambda}^{\mathrm{T}}$ in Equation 6.6, then

$$\boldsymbol{\Lambda}^{\mathrm{T}} \begin{pmatrix} f_0 \\ \mathbf{f} \end{pmatrix} \sim \mathcal{N}_{n+1}\left[\boldsymbol{\Lambda}^{\mathrm{T}} \begin{pmatrix} \mathbf{q}_0^{\mathrm{T}} \\ \mathbf{Q} \end{pmatrix}\boldsymbol{\beta}, \boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Lambda} \right] \tag{6.10}$$

i.e.

$$\begin{pmatrix} f_0 - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\mathbf{f} \\ \mathbf{f} \end{pmatrix} \sim \mathcal{N}_{n+1}\left[\begin{pmatrix} \mathbf{q}_0^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\mathbf{Q}\boldsymbol{\beta} \\ \mathbf{Q}\boldsymbol{\beta} \end{pmatrix}, \boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Lambda} \right] \tag{6.11}$$

The covariance matrix has now been diagonalized. This is useful because zero covariance implies independence for normally distributed random variables and so it follows that

$$f_0 - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\mathbf{f} \sim \mathcal{N}_1\left[\mathbf{q}_0^{\mathrm{T}}\boldsymbol{\beta} - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\mathbf{Q}\boldsymbol{\beta}, \sigma_z^2[1 - \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\boldsymbol{\psi}] \right]$$
$$\tag{6.12}$$

When we move the second term on the left hand to right in Equation 6.12, it follows the Gaussian distribution conditional on $\mathbf{f}$ given by Equation 6.7.

Lemma 2. *For any $m \times m$ symmetric and positive definite matrix $\boldsymbol{\Sigma}_\beta$, and $m \times 1$ vector $\mathbf{v}$, if the probability density function and an $m \times 1$ multivariate random variable $\boldsymbol{\beta}$ satisfies $p(\boldsymbol{\beta}) \propto \exp\{-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta} + \mathbf{v}^{\mathrm{T}}\boldsymbol{\beta}\}$, then $\boldsymbol{\beta} \sim \mathcal{N}_m\left(\boldsymbol{\Sigma}_\beta \mathbf{v}, \boldsymbol{\Sigma}_\beta\right)$.*

Proof. Consider an $m \times 1$ multivariate random variable $\mathbf{w}$ that has a multivariate Gaussian probability density function $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, i.e.

$$p(\mathbf{w}) = \frac{1}{(2\pi)^{m/2} \det(\boldsymbol{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\mu})\right\} \tag{6.13}$$

Calculate integrals in terms of $\mathbf{w}$ on both sides of Equation 6.13, we have

$$
\begin{aligned}
(2\pi)^{m/2} \mid \boldsymbol{\Sigma} \mid^{1/2} &= \int_{\mathbb{R}^m} \exp\left\{-\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\mu})\right\} d\mathbf{w} \\
&= \int_{\mathbb{R}^m} \exp\left\{-\frac{1}{2}\mathbf{w}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{w} + \boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{w} - \frac{1}{2}\boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})\right\} d\mathbf{w}
\end{aligned}
\tag{6.14}
$$

so that

$$\int_{\mathbb{R}^m} \exp\left\{-\frac{1}{2}\mathbf{w}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{w} + \boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{w})\right\} d\mathbf{w} = (2\pi)^{m/2} \det(\boldsymbol{\Sigma})^{1/2} \exp\left\{-\frac{1}{2}\boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu})\right\} \tag{6.15}$$

Substitute $\boldsymbol{\beta}$ for $\mathbf{w}$, $\boldsymbol{\Sigma}_\beta$ for $\boldsymbol{\Sigma}$, and $\mathbf{v}^{\mathrm{T}}$ for $\boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}$, so we have $\boldsymbol{\mu} = \boldsymbol{\Sigma}_\beta \mathbf{v}$.

Since $\mathbf{w} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})^{\mathrm{T}}$, then $\boldsymbol{\beta} \sim \mathcal{N}_m(\boldsymbol{\Sigma}_\beta \mathbf{v}, \boldsymbol{\Sigma}_\beta)$.

Lemma 3. *Suppose that σ_z^2, $\boldsymbol{\Psi}$, $\boldsymbol{\Sigma}_\varepsilon$ are known, then for an arbitrary $\boldsymbol{\beta}$ priori, the best linear unbiased predictor, which minimizes the mean squared prediction error between the linear predictor of the response at $\mathbf{x}^{(0)}$, is*

$$\hat{f}_0 = \mathbf{q}_0^{\mathrm{T}} \mathrm{E}[\boldsymbol{\beta} \mid \mathbf{f}] + \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}(\bar{\mathbf{y}} - \mathbf{Q}\mathrm{E}[\boldsymbol{\beta} \mid \mathbf{f}]) \tag{6.16}$$

Specifically, when $\boldsymbol{\beta}$ has a non-informative priori, i.e. $p(\boldsymbol{\beta}) \propto 1$, the best linear unbiased predictor of the response at $p(\boldsymbol{\beta}) \propto 1$ is

$$\hat{f}_0 = \mathbf{q}_0^{\mathrm{T}} \hat{\boldsymbol{\beta}} + \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \hat{\boldsymbol{\Sigma}}_\varepsilon / \sigma_z^2)^{-1}(\bar{\mathbf{y}} - \mathbf{Q}\hat{\boldsymbol{\beta}}) \tag{6.17}$$

where $\hat{\boldsymbol{\beta}} = [\mathbf{Q}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\mathbf{Q}]^{-1}\mathbf{Q}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\bar{\mathbf{y}}$.

Proof. First, we prove that Equation 6.16 is an unbiased predictor of f_0.

Using Lemma 1, we have

$$
\begin{aligned}
\mathrm{E}[f_0] &= \mathrm{E}[f_0 \mid \mathbf{f}] \\
&= \mathrm{E}[\mathrm{E}[f_0 \mid \mathbf{f}, \boldsymbol{\beta}] \mid \mathbf{f}] \\
&= \mathrm{E}[\mathbf{q}_0^{\mathrm{T}}\boldsymbol{\beta} + \sigma_z^2 \boldsymbol{\psi}^{\mathrm{T}} (\sigma_z^2 \boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon)^{-1}(\overline{\mathbf{y}} - \mathbf{Q}\boldsymbol{\beta}) \mid \mathbf{f}] \\
&= \mathbf{q}_0^{\mathrm{T}}\mathrm{E}[\boldsymbol{\beta} \mid \mathbf{f}] + \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}(\overline{\mathbf{y}} - \mathbf{Q}\mathrm{E}[\boldsymbol{\beta} \mid \mathbf{f}])
\end{aligned}
\tag{6.18}
$$

So Equation 6.16 is the unbiased predictor of f_0, i.e. $\mathrm{E}[f_0 - \hat{f}_0] = 0$.

Second, fix an arbitrary unbiased predictor f_0^*, its mean squared prediction error is given by

$$
\begin{aligned}
s^2(f_0^*) &= \mathrm{E}[(f_0^* - f_0)^2] \\
&= \mathrm{E}[(f_0^* - \hat{f}_0 + \hat{f}_0 - f_0)^2] \\
&= \mathrm{E}[(f_0^* - \hat{f}_0)^2] + s^2(\hat{f}_0) + 2\mathrm{E}[(f_0^* - \hat{f}_0)(\hat{f}_0 - f_0)] \\
&\geq s^2(\hat{f}_0) + 2\mathrm{E}[(f_0^* - \hat{f}_0)(\hat{f}_0 - f_0)] \\
&= s^2(\hat{f}_0)
\end{aligned}
\tag{6.19}
$$

So $\hat{f}_0$ is the best linear unbiased predictor that corresponds to the minimum mean squared prediction error.

In particular, when $\boldsymbol{\beta}$ has a non-informative priori, i.e. $p(\boldsymbol{\beta}) \propto 1$, its conditional distribution given $\mathbf{f}$ is

$$
\begin{aligned}
p(\boldsymbol{\beta} \mid \mathbf{f}) &= p(\overline{\mathbf{y}} \mid \boldsymbol{\beta}) p(\boldsymbol{\beta}) / p(\mathbf{f}) \\
&\propto \exp\left\{ -\frac{1}{2\sigma_z^2}(\overline{\mathbf{y}} - \mathbf{Q}\boldsymbol{\beta})^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}(\overline{\mathbf{y}} - \mathbf{Q}\boldsymbol{\beta}) \right\} \\
&\propto \exp\left\{ -\frac{1}{2\sigma_z^2}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\mathbf{Q}\boldsymbol{\beta} - 2\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\overline{\mathbf{y}}) \right\}
\end{aligned}
\tag{6.20}
$$

So Equation 6.20 satisfies the condition in Lemma 2, we have

$$
p(\boldsymbol{\beta} \mid \mathbf{f}) = \mathcal{N}_m\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}\right)
\tag{6.21}
$$

where $\hat{\boldsymbol{\beta}} = [\mathbf{Q}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\mathbf{Q}]^{-1}\mathbf{Q}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}\overline{\mathbf{y}}$, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} = [\mathbf{Q}^{\mathrm{T}}(\sigma_z^2\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon)^{-1}\mathbf{Q}]^{-1}$.

Substitute Equation 6.21 into Equation 6.16, i.e. $\mathrm{E}[\boldsymbol{\beta} \,|\, \mathbf{f}] = \hat{\boldsymbol{\beta}}$, we have the best

linear unbiased predictor of the response at $\mathbf{x}^{(0)}$ is Equation 6.17.

Theorem 1. *When σ_z^2, $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}_\varepsilon$ are known, (I) if $p(\boldsymbol{\beta}) \propto 1$ on $\mathbb{R}^m$, then the*

predictive distribution of the response at $\mathbf{x}^{(0)} \in \mathbb{R}^k$ belongs to a Gaussian

distribution, i.e. $p(f_0 \,|\, \mathbf{f}) = \mathcal{N}_1(\mu_{f|\mathbf{f}}, \sigma_{f|\mathbf{f}}^2)$, where

$$\mu_{f|\mathbf{f}} = \mathbf{q}_0^{\mathrm{T}}\hat{\boldsymbol{\beta}} + \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}(\overline{\mathbf{y}} - \mathbf{Q}\hat{\boldsymbol{\beta}}) \tag{6.22}$$

$$\sigma_{f|\mathbf{f}}^2 = \sigma_z^2 \left[1 - \left(\mathbf{q}_0^{\mathrm{T}}, \boldsymbol{\psi}^{\mathrm{T}}\right) \begin{pmatrix} \mathbf{0}_{m\times m} & \mathbf{Q}^{\mathrm{T}} \\ \mathbf{Q} & \boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{q}_0 \\ \boldsymbol{\psi} \end{pmatrix} \right] \tag{6.23}$$

(II) if $p(\boldsymbol{\beta}) = \mathcal{N}_m\left(\boldsymbol{\beta}_0, \mathbf{B}\right)$ on $\mathbb{R}^m$, then the predictive distribution of the

response at $\mathbf{x}^{(0)} \in \mathbb{R}^k$ belongs to Gaussian distribution, i.e. $p(f_0 \,|\, \mathbf{f}) = \mathcal{N}_1(\mu_{f|\mathbf{f}}, \sigma_{f|\mathbf{f}}^2)$,

where

$$\mu_{f|\mathbf{f}} = \mathbf{q}_0^{\mathrm{T}}\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}} + \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2)^{-1}(\overline{\mathbf{y}} - \mathbf{Q}\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}}) \tag{6.24}$$

where $\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}} = [\mathbf{Q}^{\mathrm{T}}(\sigma_z^2\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon)^{-1}\mathbf{Q} + \mathbf{B}^{-1}]^{-1}[\mathbf{Q}^{\mathrm{T}}(\sigma_z^2\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon)^{-1}\overline{\mathbf{y}} + \mathbf{B}^{-1}\boldsymbol{\beta}_0]$, and

$$\sigma_{f|\mathbf{f}}^2 = \sigma_z^2 \left[1 - \left(\mathbf{q}_0^{\mathrm{T}}, \boldsymbol{\psi}^{\mathrm{T}}\right) \begin{pmatrix} -\sigma_z^2\mathbf{B}^{-1} & \mathbf{Q}^{\mathrm{T}} \\ \mathbf{Q} & \boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{q}_0 \\ \boldsymbol{\psi} \end{pmatrix} \right] \tag{6.25}$$

Proof. We first prove the informative $\boldsymbol{\beta}$ priori, i.e. the Gaussian distribution

priori in (II), when $|\mathbf{B}| \to \infty$, the informative $\boldsymbol{\beta}$ priori reduces to the non-informative

priori shows in (I).

Given a known σ_z^2, the conditional density of $f_0 \,|\, \mathbf{f}$ can be written as

$$p(f_0 \mid \mathbf{f}) = \int_{\mathbb{R}^m} p(f_0 \mid \mathbf{f}, \boldsymbol{\beta}) p(\boldsymbol{\beta} \mid \mathbf{f}) \mathrm{d}\boldsymbol{\beta} \tag{6.26}$$

Consider the following conditional probability density function by applying Lemma 1, we have the first term in the right-side integrand of Equation 6.26

$$p(f_0 \mid \mathbf{f}, \boldsymbol{\beta}) = \mathcal{N}_1\left(\mu_{f|\mathbf{f}, \boldsymbol{\beta}}, \sigma^2_{f|\mathbf{f}, \boldsymbol{\beta}} \right) \tag{6.27}$$

where $|\mu_{f|\mathbf{f}, \boldsymbol{\beta}} = \mathbf{q}_0^\mathrm{T}\boldsymbol{\beta} + \boldsymbol{\psi}^\mathrm{T}(\boldsymbol{\Psi} + \hat{\boldsymbol{\Sigma}}_\varepsilon / \sigma_z^2)^{-1}(\overline{\mathbf{y}} - \mathbf{Q}\boldsymbol{\beta})$, $\sigma^2_{f|\mathbf{f}, \boldsymbol{\beta}} = \sigma_z^2[1 - \boldsymbol{\psi}^\mathrm{T}(\boldsymbol{\Psi} + \hat{\boldsymbol{\Sigma}}_\varepsilon / \sigma_z^2)^{-1}\boldsymbol{\psi}]$.

According to the Bayes' theorem, we have the second term in the right-side integrand of Equation 6.26 as the following

$$p(\boldsymbol{\beta} \mid \mathbf{f}) = \frac{p(\overline{\mathbf{y}} \mid \boldsymbol{\beta}) p(\boldsymbol{\beta})}{p(\mathbf{f})}$$

$$\propto \exp\left\{ -\frac{1}{2}(\overline{\mathbf{y}} - \mathbf{Q}\boldsymbol{\beta})^\mathrm{T}(\sigma_z^2\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon)^{-1}(\overline{\mathbf{y}} - \mathbf{Q}\boldsymbol{\beta}) \right\} \times \exp\left\{ -\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\mathrm{T}\mathbf{B}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right\}$$

$$\propto \exp\left\{ -\frac{1}{2}\boldsymbol{\beta}^\mathrm{T}\left(\mathbf{Q}^\mathrm{T}(\sigma_z^2\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon)^{-1}\mathbf{Q} + \mathbf{B}^{-1} \right)^{-1}\boldsymbol{\beta} + \boldsymbol{\beta}^\mathrm{T}\left(\mathbf{Q}^\mathrm{T}(\sigma_z^2\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon)^{-1}\overline{\mathbf{y}} + \mathbf{B}^{-1}\boldsymbol{\beta}_0 \right) \right\}$$

$$\tag{6.28}$$

Since Equation 6.28 satisfies the condition in Lemma 2, we have

$$p(\boldsymbol{\beta} \mid \mathbf{f}) = \mathcal{N}_1\left(\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}} \right) \tag{6.29}$$

where

$$\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}} = [\mathbf{Q}^\mathrm{T}(\sigma_z^2\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon)^{-1}\mathbf{Q} + \mathbf{B}^{-1}]^{-1}[\mathbf{Q}^\mathrm{T}(\sigma_z^2\boldsymbol{\Psi} + \hat{\boldsymbol{\Sigma}}_\varepsilon)^{-1}\overline{\mathbf{y}} + \mathbf{B}^{-1}\boldsymbol{\beta}_0] \tag{6.30}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}} = [\mathbf{Q}^\mathrm{T}(\sigma_z^2\boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon)^{-1}\mathbf{Q} + \mathbf{B}^{-1}]^{-1} \tag{6.31}$$

To simplify the notations in the followings, let $|\tilde{\boldsymbol{\Psi}} = \boldsymbol{\Psi} + \boldsymbol{\Sigma}_\varepsilon / \sigma_z^2$. Substitute Equation 6.27 and Equation 6.29 into Equation 6.26, and ignore constants of proportionality, we obtain

$$p(f_0 \mid \mathbf{f}) \propto \int_{\mathbb{R}^m} \exp\left\{-\frac{(f_0 - \mu_{f|\mathbf{f},\boldsymbol{\beta}})^2}{2\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2} - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}})^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}})\right\}\mathrm{d}\boldsymbol{\beta}$$

$$\propto \exp\left\{-\frac{(f_0 - \boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\bar{\mathbf{y}})^2}{2\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2}\right\} \times \int_{\mathbb{R}^m}\exp\left\{\begin{array}{l} -\dfrac{\boldsymbol{\beta}^{\mathrm{T}}(\boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\mathbf{Q} - \mathbf{q}_0^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\mathbf{Q} - \mathbf{q}_0^{\mathrm{T}})\boldsymbol{\beta}}{2\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2} \\[2mm] -\dfrac{(f_0 - \boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\bar{\mathbf{y}})(\boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\mathbf{Q} - \mathbf{q}_0^{\mathrm{T}})\boldsymbol{\beta}}{\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2} \\[2mm] -\dfrac{1}{2}(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}\boldsymbol{\beta} + \boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}} - 2\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}\boldsymbol{\beta}) \end{array}\right\}\mathrm{d}\boldsymbol{\beta}$$

$$\propto \exp\left\{-\frac{(f_0 - \boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\bar{\mathbf{y}})^2}{2\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2}\right\} \times \int_{\mathbb{R}^m}\exp\left\{-\frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}\left(\frac{\mathbf{h}\mathbf{h}^{\mathrm{T}}}{\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2} + \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}\right)^{-1}\boldsymbol{\beta} + \mathbf{v}^{\mathrm{T}}\boldsymbol{\beta}\right\}\mathrm{d}\boldsymbol{\beta}$$

$$(6.32)$$

where $\mathbf{h}^{\mathrm{T}} = \boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\mathbf{Q} - \mathbf{q}_0^{\mathrm{T}}$, $\mathbf{v} = -\dfrac{\mathbf{h}(f_0 - \boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\bar{\mathbf{y}})}{\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2} + \boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}}$.

By Lemma 2, we can further derive Equation 6.32 as follows

$$p(f_0 \mid \mathbf{f}) \propto (2\pi)^{m/2}\det(\boldsymbol{\Sigma})^{1/2}\exp\left\{-\frac{(f_0 - \boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\bar{\mathbf{y}})^2}{2\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2} + \frac{1}{2}\mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{v}\right\}$$

$$\propto \exp\left\{\begin{array}{l} -\dfrac{f_0^2}{2\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2} + \dfrac{\boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\bar{\mathbf{y}}}{\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2}f_0 \\[3mm] + \dfrac{1}{2\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2}(f_0 - \boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\bar{\mathbf{y}})\mathbf{h}^{\mathrm{T}}\left(\mathbf{h}\mathbf{h}^{\mathrm{T}} + \sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}\right)^{-1}\mathbf{h}(f_0 - \boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\bar{\mathbf{y}}) \\[3mm] - \dfrac{1}{\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2}(f_0 - \boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\bar{\mathbf{y}})\mathbf{h}^{\mathrm{T}}\left(\mathbf{h}\mathbf{h}^{\mathrm{T}} + \sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}\right)^{-1}\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}} \end{array}\right\}$$

$$\propto \exp\left\{\begin{array}{l} -\dfrac{f_0^2}{2}\left(\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2 + \mathbf{h}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}\mathbf{h}\right)^{-1} \\[3mm] + \boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\bar{\mathbf{y}}\left(\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2 + \mathbf{h}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}\mathbf{h}\right)^{-1}f_0 - \mathbf{h}^{\mathrm{T}}\left(\mathbf{h}\mathbf{h}^{\mathrm{T}} + \sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}\right)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}}f_0 \end{array}\right\}$$

$$\propto \exp\left\{-\frac{f_0^2}{2}\left(\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2 + \mathbf{h}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}\mathbf{h}\right)^{-1} + \left(\boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\bar{\mathbf{y}} - \mathbf{h}^{\mathrm{T}}\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}}\right)\left(\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2 + \mathbf{h}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}\mathbf{h}\right)^{-1}f_0\right\}$$

$$(6.33)$$

where we apply the Sherman-Morrison-Woodbury matrix inverse formula (Woodbury, 1950) to derive the inverse matrices in Equation 6.33 as follows

$$\left(\mathbf{h}\mathbf{h}^{\mathrm{T}}+\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^{2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}\right)^{-1}=\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^{-2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}-\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^{-2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}\mathbf{h}\left(\mathbf{h}^{\mathrm{T}}\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^{-2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}\mathbf{h}+1\right)^{-1}\mathbf{h}^{\mathrm{T}}\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^{-2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}} \tag{6.34}$$

$$\mathbf{h}^{\mathrm{T}}\left(\mathbf{h}\mathbf{h}^{\mathrm{T}}+\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^{2}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}\right)^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}}=\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^{-2}\mathbf{h}^{\mathrm{T}}\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}}[1-\mathbf{h}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}\mathbf{h}\left(\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^{2}+\mathbf{h}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}\mathbf{h}\right)^{-1}] \tag{6.35}$$

We track the "decision variable" f_0 in Equation 6.33, again, by Lemma 2, we conclude that $p(f_0 | \mathbf{f}) = \mathcal{N}_1(\mu_{f|\mathbf{f}}, \sigma_{f|\mathbf{f}}^2)$ by substituting $\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^2$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}}$, thus

$$\mu_{f|\mathbf{f}}=\left(\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^{2}+\mathbf{h}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}\mathbf{h}\right)\left(\boldsymbol{\psi}^{\mathrm{T}}\tilde{\boldsymbol{\Psi}}^{-1}\overline{\mathbf{y}}-\mathbf{h}^{\mathrm{T}}\boldsymbol{\mu}_{\boldsymbol{\beta}|\mathbf{f}}\right)\left(\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^{2}+\mathbf{h}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}\mathbf{h}\right)^{-1} \tag{6.36}$$

$$\sigma_{f|\mathbf{f}}^{2}=\sigma_{f|\mathbf{f},\boldsymbol{\beta}}^{2}+\mathbf{h}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}|\mathbf{f}}\mathbf{h} \tag{6.37}$$

Then substitute Equations 6.27, 6.30 and 6.31 into Equations 6.36 and 6.37, we obtain the results of Equations 6.24 and 6.25. So we the case of informative $\boldsymbol{\beta}$ priori is proven. If $\mathbf{B} | \rightarrow \infty$ or $\mathbf{B}^{-1}=\mathbf{0}_{m\times m}$, the non-informative $\boldsymbol{\beta}$ priori leads to the results shown in Equations 6.22 and 6.23.

Note that the priori of $\boldsymbol{\beta}$ can be estimated by the first-round simulations across all design points. The least squares error estimations of $\left(\boldsymbol{\beta}_0, \mathbf{B}\right)$ are

$$\hat{\boldsymbol{\beta}}_0 = (\mathbf{Q}^{\mathrm{T}}\mathbf{Q})^{-1}\mathbf{Q}^{\mathrm{T}}\mathbf{y}_1 \qquad \text{and} \qquad \hat{\mathbf{B}}=\frac{(\mathbf{y}_1-\mathbf{Q}\hat{\boldsymbol{\beta}}_0)^{\mathrm{T}}(\mathbf{y}_1-\mathbf{Q}\hat{\boldsymbol{\beta}}_0)(\mathbf{Q}^{\mathrm{T}}\mathbf{Q})^{-1}}{n-m} \qquad , \qquad \text{where}$$

$\mathbf{y}_1 = [y_1^{(1)}, y_1^{(2)}, \dots, y_1^{(n)}]^{\mathrm{T}}$ is the first-round simulation output vector of all n points.

6.1.2 Model Evaluation

As Kriging surrogate models provide predictive distributions for the outputs of any design points, we use the Kullback-Leibler divergence (Kullback and Leibler,

1951) as the performance measure. It is also called relative entropy, which is a non-symmetric measure of the difference between two probability distributions. Specifically, the Kullback-Leibler divergence of an unknown multivariate Gaussian distribution $\mathcal{N}_{\text{True}}$ from an estimated multivariate Gaussian distribution $\mathcal{N}_{\text{Est}}$, is a measure of the information lost when using $\mathcal{N}_{\text{Est}}$ to approximate $\mathcal{N}_{\text{True}}$. The analytical derivation of the Kullback-Leibler divergence of Gaussian distributions is given by

$$D_{\text{KL}}(\mathcal{N}_{\text{Est}} \parallel \mathcal{N}_{\text{True}}) = \frac{1}{2}\left[\text{tr}(\mathbf{\Sigma}_{\text{True}}^{-1}\mathbf{\Sigma}_{\text{Est}}) - (\mathbf{\mu}_{\text{True}} - \mathbf{\mu}_{\text{Est}})^{\text{T}}\mathbf{\Sigma}_{\text{True}}^{-1}(\mathbf{\mu}_{\text{True}} - \mathbf{\mu}_{\text{Est}}) - n - \ln\left(\frac{\det(\mathbf{\Sigma}_{\text{Est}})}{\det(\mathbf{\Sigma}_{\text{True}})} \right) \right]$$

$$(6.38)$$

where D_{KL} is the Kullback-Leibler divergence, $\mathbf{\mu}_{\text{True}}$ and $\mathbf{\mu}_{\text{Est}}$ are the mean values of $\mathcal{N}_{\text{True}}$ and $\mathcal{N}_{\text{Est}}$, respectively, $\mathbf{\Sigma}_{\text{True}}$ is the nonsingular covariance matrix of $\mathcal{N}_{\text{True}}$, and the estimated covariance matrix of $\mathcal{N}_{\text{Est}}$ is given by

$$\mathbf{\Sigma}_{\text{Est}} = \hat{\sigma}_z^2\hat{\mathbf{\Psi}} + \hat{\mathbf{\Sigma}}_\varepsilon \qquad (6.39)$$

6.2 Numerical Example

Is it beneficial to include both parameter uncertainties and heteroscedastic simulation noise into the ordinary Kriging surrogate model? In this section, we work with a simplified transportation network to gain some insights of the Bayesian stochastic Kriging model using the DynusT simulator.

To illustrate the methodology developed in this chapter and compare it with existing models, we set up a toy network to simulate how travelers' value of time (VOT) influence the route choice behaviors. Our purpose in this section is three-fold: To provide some intuition about what the Bayesian stochastic Kriging method does on approximating the black-box function of a simulation-based dynamic traffic assignment problem; to assess the parameter uncertainty; and to evaluate the estimation robustness by comparing the heteroscedastic errors in our model.

6.2.1 Toy Network

The toy network is depicted in Figure 6-1. Overall, the network consists of 10 links and 6 nodes. There is only one origin-destination (OD) pair from node 1 to node 3 in this network. Two major parallel corridors serve for the travel demand between the OD pair. The route 1-2-3 is a freeway corridor with two general-purpose lanes and one High-occupancy vehicle (HOV) lane, while the route 1-4-5-6-3 is an arterial with 2 lanes. Vehicles may change their route choices between the freeway and the arterial under different VOT via links 8 and 9. The configuration of the links is illustrated in Table 6-1.

Figure 6-1: Numerical Illustration Network.

Link	1	2	3	4	5	6	7	8	9	10
Length (mile)	45	45	45	45	30	30	0.2	0.2	0.2	0.2
Lanes	2	2	1	1	2	2	2	2	2	2
Speed limit (mph)	60	60	60	60	40	40	40	40	40	40
Free flow travel time (min)	45	45	45	45	45	45	0.3	0.3	0.3	0.3
Capacity (veh/lane/hour)	1500	1500	1500	1500	1000	1000	1500	1500	1500	1500

Table 6-1: Input Data for a Small Road Network.

This network is coded into DynusT, and simulation-based dynamic traffic assignment is applied to capture the route choice behavior. Five-hour dynamic travel demand between the OD pair is set up as in Table 6-2. To allow all vehicles to dissipate, we simulate the network for 7 hours, with no travel demand during the last two hours.

Time (hour)	SOV[a]	HOV
0 - 1	3000	300
1 - 2	5000	500
2 - 3	5800	850
3 - 4	4700	950
4 - 5	2800	350

[a] SOV: single occupancy vehicles.

Table 6-2: Dynamic Travel Demand.

In this numerical example, the average travel time for all vehicles is selected as the dependent variable, and the surface to be approximated is the response of average travel time to VOT. Due to the effect of random seeds and the stochastic

route choice behavior, the output of the DynusT model is noisy. To create the real response surface, we first generate a uniformly distributed sample of VOT, and then run 100 replications of simulations for each specific input of VOT. The mean of the output from the 100 replications is assumed to be the true response. This true response would be used for validating the surrogate models. In addition, we also consider the variation of the output. Thus the distribution of the output from the 100 replications of simulations for each VOT value is utilized to validate the estimated distribution of response value from the surrogate models.

For the estimation of the Bayesian stochastic Kriging model, results from the first 10 replications of simulation for each corresponding VOT value are utilized. Moreover, to verify the advantageous of incorporating parameter uncertainties and heteroscedastic simulation noise into the surrogate model, we estimate a regressing Kriging model with the same input as that for the Bayesian stochastic Kriging model, and compare their performance on approximating the real response surface.

6.2.2 Numerical Results

To compare the developed Bayesian stochastic Kriging model with existing surrogate-based optimization approaches, e.g. quadratic polynomial response surface method, ordinary Kriging for deterministic input-output relationship, and regressing Kriging, the first 10 replications are used to estimate the heteroscedastic simulation errors at different design points and the entire 100 replications are used to approximate the true response distributions, respectively.

134

| Goodness-of-fit	Quadratic Polynomial	Ordinary Kriging	Regressing Kriging	Bayesian stochastic Kriging
D_{KL}	N/A[a]	N/A	65.58	38.33
RMSE	1.90	0.19	0.87	0.20
MAE	5.67	0.59	2.70	0.56
NRMSE	1.89%	0.19%	0.86%	0.20%
NMAE	3.10	0.33	1.52	0.31
PCC	0.00	0.99	0.98	0.99

[a] N/A: not applicable.

Table 6-3: The Goodness-of-Fit of Surrogate Models.

Table 6-3 compares results of the four models with six MoEs. It should be pointed out that the Kullback-Leibler divergence measures the difference of two probability density functions. It is an important performance measure to evaluate the surrogate function of a stochastic simulation optimization problem. At 30 design points (VOT increases from US$ 1/hour to US$ 30/hour with a step of US$ 1/hour), the ordinary Kriging method predicts zero errors to their observed mean values, while both regressing Kriging and Bayesian stochastic Kriging predict non-zero standard errors. For this stochastic simulation numerical example in the synthetic network, we can see that the Kullback-Leibler divergence of the Bayesian stochastic Kriging is smaller than regressing Kriging, indicating that the proposed model generates better predictions given heteroscedastic data. Table 6-3 also shows estimation errors by comparing the prediction values with true objective function (estimated by 100 replications of simulations) over these 30 design points. The RMSE, MAE, NRMSE, NMAE, PCC values are used for the mean output evaluation. We can see that the

Bayesian stochastic Kriging results in similar accuracy as the ordinary Kriging in mean value prediction.
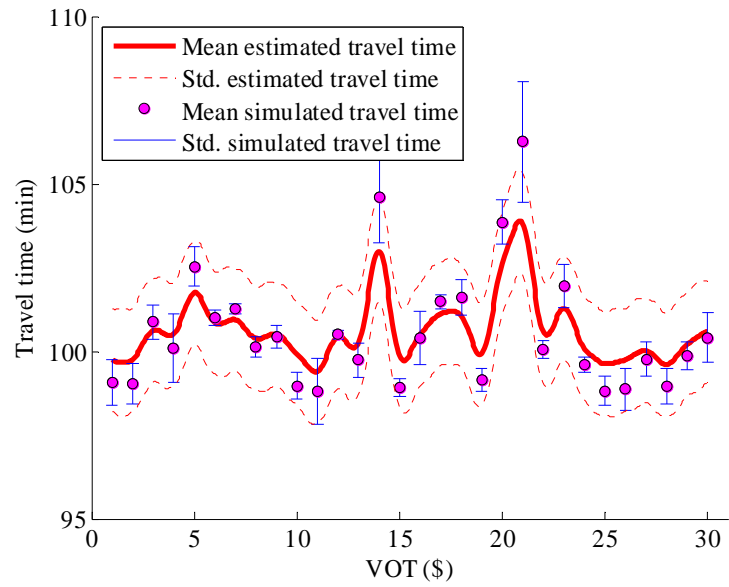
(a)
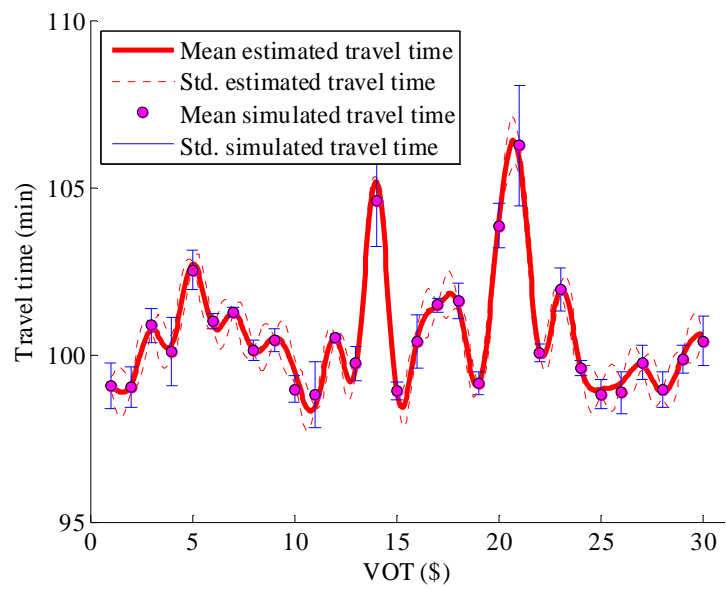
(b)

Figure 6-2: Comparison of Surrogate Models, (a) Regressing Kriging; (b) Bayesian Stochastic Kriging.

Figure 6-2 shows the surrogates of regressing Kriging and Bayesian stochastic Kriging corresponding to Table 6-3. These results verify the capability of the proposed Bayesian stochastic Kriging surrogate model in estimations of both mean values and standard errors given heteroscedastic noisy simulation input-output relationships.

6.3 Application of Integrated Corridor Planning and Operational Optimization

6.3.1 Research Problem

This application aims to jointly optimize a transportation planning policy (i.e. HOT toll) and an operations strategy (i.e. DMS) for a freeway-arterial corridor network. Both of the two strategies have been separately investigated extensively in existing literature.

HOT lanes reserve a set of freeway lanes HOV, while allowing low-occupancy vehicles to enter for a toll. A good pricing scheme of HOT facilities would increase the HOT lane usage and relieve congestion on the parallel general-purpose lanes. Benefits of HOT lanes in travel time reduction, freeway efficiency improvement, and bottleneck congestion mitigation have been widely studied in multiple region-specific case studies (Burris et al., 2009; Goel and Burris, 2012). Yin and Lou (2009) delivered a reactive self-learning approach for determining time-

varying tolls in response to the detected traffic arrivals. A Logit lane choice model was further adopted by Lou et al. (2011) to specify toll rates that maximized the freeway's throughput. Gardner et al., (2013) modeled the choice process of individual drivers to pay and take the HOT lane, which could be helpful for the determining the toll setting.

DMS is the most common way to provide real time travel information to drivers and encourage diversion before they approach the congested areas. The theory of random utility maximization has been applied in Ben-Akiva and Lerman (1985) and Khattack et al. (1995), while other research adopted if-then rules based on fuzzy logics to model the diversion in response to information (Paz and Peeta, 2009; Xiong and Zhang, 2013). In previous research, the optimization of diversion rate has been analyzed through heuristic methods. Jacob et al. (2006) proposed a reinforcement learning approach to investigate the optimal diversion control for an express/collector corridor affected by work zones. Chen et al. (2005) developed a simulated annealing-based algorithm to search for the optimal alternative route and diversion rate for a two-lane highway resurfacing project. In addition to the theoretical models, the diversion rate in response to DMS was investigated with field data in Horowitz et al. (2003). However, the focus of this study is not to analyze the actual diversion rate in response to DMS. Instead, we assume that the relationship between diversion rate and the information provided through DMS is already clear, and the objective of this study is to search for the optimal diversion rate that creates the best system performance.

The major challenge of jointly optimizing these two corridor management strategies is that it is very hard to formulate the objective function in a mathematical equation, as the two strategies concern about the analysis of travel behavior at different level. While on the other hand, simulation can conveniently evaluate the impact of the combined strategies to the system performance. Therefore, we take advantage of simulation, and apply SBO method to solve this problem.

6.3.2 Study Area

The study freeway/arterial corridor is along a 15.50-mile freeway segment of Interstate 270 (I-270 is known as the Washington National Pike) in the Montgomery County, State of Maryland. It lies between I-495 (the Capital Beltway) north of Bethesda, and Maryland Route (MD 124) in the city of Gaithersburg, Montgomery County. The I-270 consists of a 12.40-mile mainline as well as a 2.10-mile spur that provides access to and from southbound I-495. The freeway corridor heads northwest from an interchange with I-495 and MD 355 (Rockville Pike) in suburban Bethesda as an up to twelve-lane freeway with a 55 mph speed limit. The left lane on each side is used as a HOV lane in the northbound direction between 15:30 and 18:30 weekdays and in the southbound direction between 6:00 AM and 9:00 AM weekdays. I-270 takes on a local-express lane configuration with the outer two lanes serving as local lanes and the inner three lanes and the HOV lane serving as express lanes.
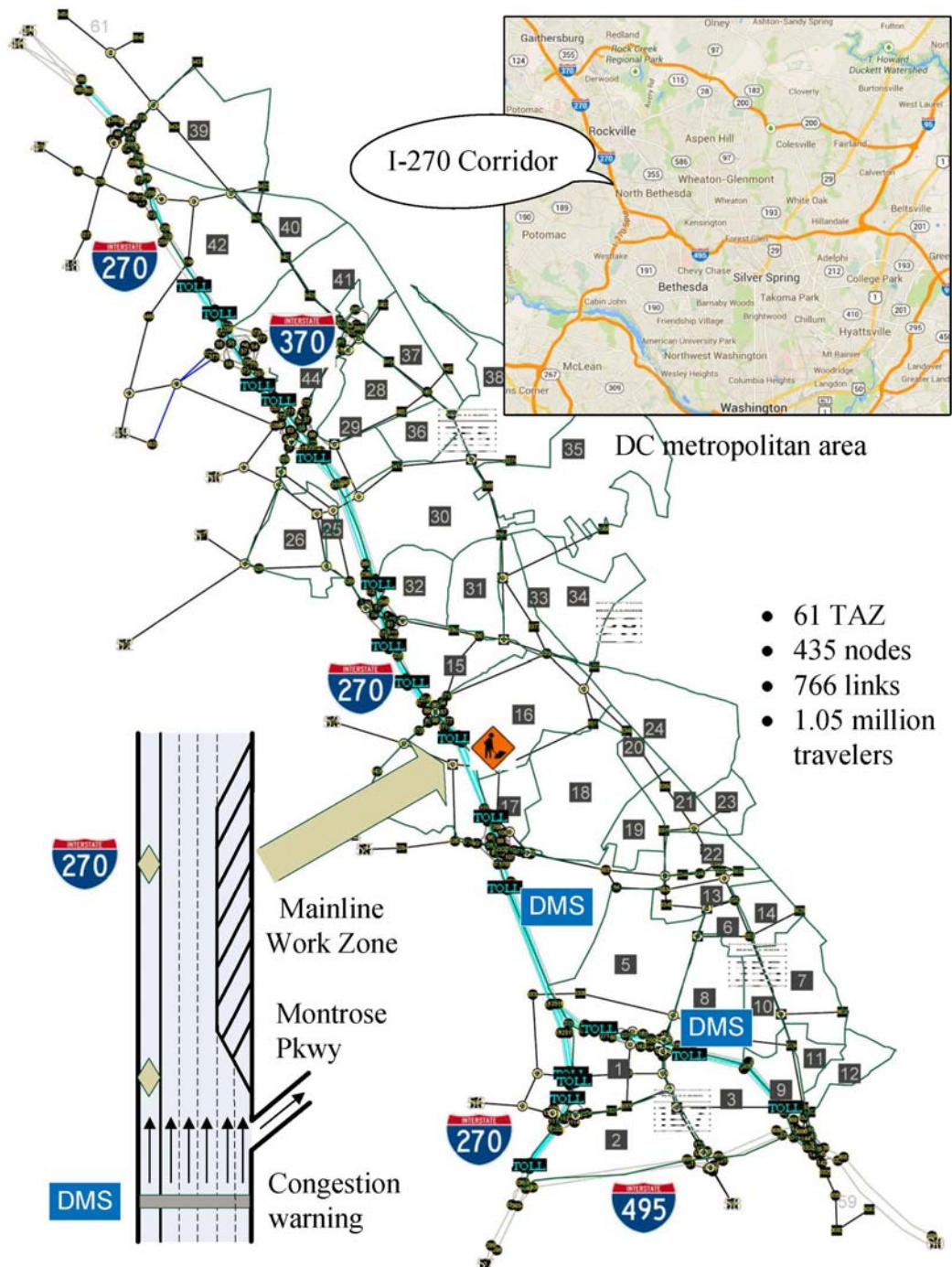
139

Figure 6-3: Simulation Network of I-270 Freeway/Arterial Corridor.

The urban network of arterials and freeways is coded into DynusT. All freeways and major arterials illustrated in Figure 6-3 are included in the network,

which has 61 traffic analysis zones, 435 nodes and 766 links. Three modes of dynamic OD matrices, i.e. SOV, HOV and trucks, are estimated based on demand data from the regional planning model (Zhang et al. 2012; Xiong et al., 2014).

To investigate the impact of the two traffic management strategies (i.e. HOT toll and DMS) on travel behavior, we develop a work zone scenario in the aforementioned simulation network, which would create recurrent congestion in the corridor area. The effectiveness of jointly utilizing those strategies to divert traffic from congested areas can then be tested. The work zone scenario is simulated for a period of time from 14:00 to 19:00.

As shown in Figure 6-3, the work zone is deployed on the north bound of I-270 mainline, at the segment between the intersections to Montrose Road and Falls Road. Two lanes out of the initial five general-purpose lanes of this segment are closed for the road work. The initial HOV lane during the afternoon peak period (15:30-18:30) is converted to an HOT lane, allowing SOVs travelling through this lane with an additional monetary cost. Meanwhile, two DMSs are deployed at two major off-ramps of I-270 at the upstream of the work zone. These DMS can provide congestion warning to help divert traffic to the parallel arterial MD 355 before they approach to the work zone. Hereafter, we define the diversion rate as the proportion of drivers who make route choices based on their own perception of the congestion warning among all travelers. However, the response of travelers to the information provided by DMS is out of the scope of this study. Please refer to Paz and Peeta (2009) and Xiong and Zhang (2013) for more details. Instead, the objective is to search for the optimal diversion rate, which then serves as the goal for DMS. In

summary, the objective of the optimization problem is minimizing the average travel time for all finished trips during the 5-hour period, and the decision variables are HOT toll rate for the afternoon peak period (i.e. 15:30-18:30) and the independent diversion rate for the two DMS.

The optimization problem is given by

$$\min_{\mathbf{x} \in \mathbb{R}^3} \quad E[f(\mathbf{x})] = E[f(x_1, x_2, x_3)] \tag{6.40a}$$

$$\text{s.t.} \quad \mathbf{x}_{min} \leq \mathbf{x} \leq \mathbf{x}_{max} \tag{6.40b}$$

where $f(\mathbf{x})$ represents the unknown true average trip travel time for all travelers of the corridor given the input $\mathbf{x}$, x_1 is the HOT toll rate, x_2 is the diversion rate of the DMS next to the work zone, guiding travelers to Montrose Pkwy, x_3 is the diversion rate of the DMS at the off-ramp to MD 187. So $\mathbf{x}$ is a three-dimensional decision variable vector. The box constraints are $\mathbf{x}_{min} = [0, 0, 0]^T$ and $\mathbf{x}_{max} = [US\$ 5.00, 100\%, 100\%]^T$, which are lower and upper boundaries for planning and operational strategies, respectively. The lower bound collects no toll on the HOV lane, which equals to open the HOV lane to SOV in peak hours, while the upper bounds collects US\$ 5.00 per SOV use of the HOT lane.

6.3.3 Simulation Demand and Supply Calibration

This I-270 corridor network is a sub cut of the ICC regional network we introduced previously. Before using the simulation to evaluate system performance, the demand and supply of the simulated network should be calibrated.

Statistics	SOV	HOV	Truck	Total vehicles
Number of vehicles	955,095	56,671	38,763	1,050,529
Average overall trip time (min)	8.26	8.48	6.94	8.23
Average trip times (min)	8.13	8.33	6.81	8.09
Average entry queuing time (min)	0.14	0.15	0.13	0.14
Average stop time (min)	0.69	0.62	0.60	0.69
Average trip distance (mile)	5.00	5.35	3.92	4.98
Average travel speed (mph)	–[a]	–	–	36.93

[a]–: not available.

Table 6-4: Summaries of the 24-Hour Simulation for the Corridor.

We first simulated the travels of the corridor for the whole 24-hour weekday. A total number of 1,053,052 vehicles are loaded in the network during 24 hours. Field collections of urban street signal timing are also included in the network. From the 24-hour simulation results, we see that the largest vehicle volumes in the corridor are on the freeway I-270 and its parallel major arterial MD 355, as well as several other arterials (e.g. MD 28, MD 124 and MD 187). At the end of 20 iterations of DTA and mesoscopic simulation, the CPU time of operation is 572 min (CPU time of simulation is 348 min, CPU time of assignment is 224 min) with a 2.66 GHz-quad CPU and 4 GB-Ram computer. For the scenario after both demand and supply calibration, at the end of 20 iterations, a percentage of 99.76% vehicles complete their trips, see detailed vehicle statistics in Table 6-4.

To calibrate the simulation model, we collect traffic flow data of 35 detector stations from January 1, 2013 to June 30, 2013 along the I-270 corridor, including 11 detectors on I-270 general-purpose lanes, 6 detectors on I-270 HOV lanes, 10

detectors on I-495, 6 detectors on MD 187, and 2 detectors on MD 355. We use 6-month (January 1 to June 30, 2013) empirical loop/microwave data of the freeway network, including lane-by-lane speed, occupancy and volume extracted from fixed detectors (CATT Lab, 2013). Each station detects 1 through 5 lanes at a time interval of 15 minutes. Since we simulate traffic flow based on weekday origin-destination demands (Chen et al., 2013), field measurements on Saturdays and Sundays are excluded. Eventually, we obtain 130-weekday measurements of 35 detection stations. Since simultaneous demand–supply calibration is found to be superior to demand-only calibration in accuracy (Balakrishna et al., 2007a, 2007b; Vaze et al., 2009), we will use the lane aggregate traffic data to support the demand and supply calibration.

The mesoscopic traffic simulation in DynusT is based on the anisotropic mesoscopic simulation model, which moves vehicles according to the speed-density relationship (Chiu et al., 2010b). The modified two-regime Greensheild's type equation is used to quantify the relationship as follows

$$v = \begin{cases} v_{\mathrm{f}} & 0 \le \rho \le \rho_{\mathrm{c}} \\ v_0 + (v_{\mathrm{f}} - v_0)(1 - \rho / \rho_{\mathrm{J}})^{\alpha} & \rho_{\mathrm{c}} < \rho \le \rho_{\mathrm{J}} \end{cases} \tag{6.41}$$

where v is the space-mean speed, ρ is the density.

Figure 6-4(a) shows the comparison of the default traffic flow model setting and the calibrated speed-density relationship for one of the detectors, i.e. station ID 3392 that locates at I-270 NB 0.23 Mile North of Grosvenor Ln. The least square error estimations of traffic flow parameters using loop/microwave data are density breakpoint $\rho_{\mathrm{c}} = 31.34$ veh/mile/lane, speed intercept $v_{\mathrm{int}} = 105.73$ mph, minimal speed $v_0 = 5$ mph, jam density $\rho_{\mathrm{J}} = 200$ veh/mile/lane, shape term $\alpha = 3.61$, and

free-flow speed $v_f = 59.44$ mph. The detailed comparisons of other 34 detection stations between simulation results and 6-month field measurements are omitted due to the length limit.
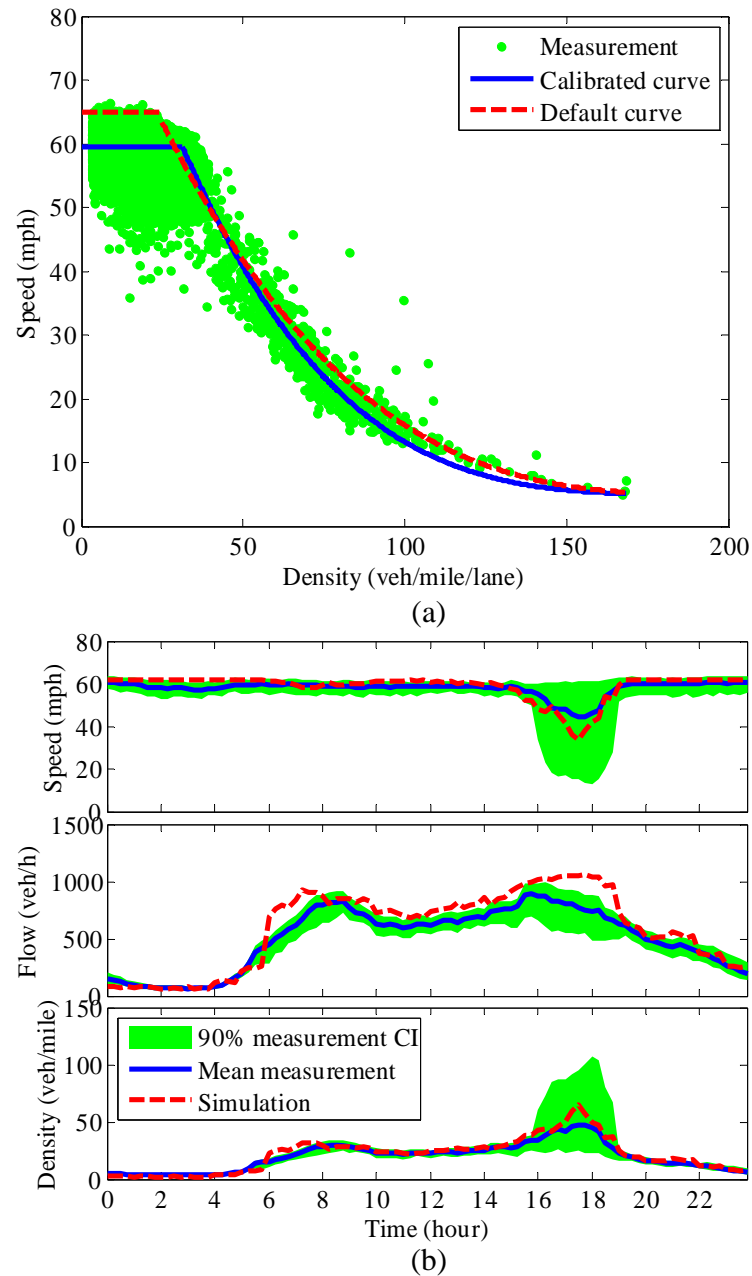

(a)

(b)

Figure 6-4: Supply Calibration Results, (a) Modified Greenshield's Model Calibration for Detector ID 3392; (b) Comparison of Measured and Simulated Traffic Flows on General-Purpose Lanes.

Figure 6-4(b) shows simulation results of the freeway network average speed, density and flow by comparing them with 6-month traffic flow data. The 90% confidence interval (CI) is estimated by the 130-weekday data. We can see the simulation matches well with historical measurements.

To reveal the simulation accuracy from the perspective of network level statistics for both SOV and HOV performance, two separate experiments are performed. First, only the demand is calibrated using our previously proposed method (Zhang et al., 2012) while supply (e.g. number of lanes, traffic flow model parameters, and speed limit) is held using default values of DynusT. In the second experiment, both demand and supply are calibrated. Since there are 35 detector stations collecting traffic flow data in the study network, we first estimate the simulation errors of individual detector stations, and then aggregate the error into weighted average measures of goodness-of-fit using measured traffic volumes as weighted coefficients.

Simulation errors	Demand calibration only			Demand/supply calibration		
	Speed	Flow	Density	Speed	Flow	Density
RMSE	11.14	410.48	48.98	7.83	182.05	15.14
MAE	28.28	852.16	164.83	18.18	447.27	42.52
NRMSE	20.65%	67.48%	242.82%	14.25%	27.69%	48.58%
NMAE	8.96	2.95	12.38	4.50	1.44	2.48

Table 6-5: Calibration Errors of Network-Wide Traffic Flow Quantities on General-Purpose Lanes.

Table 6-5 shows the validation results of the simulation model before and after supply parameters calibration using 6-month field measurements on general-

purpose lanes. The demand/supply calibration provides considerable improvements in simulation accuracy and obtains more satisfactory results than the demand-only calibration.

Segments	*Periods*	*Field detection*	*Demand calibration*	*Demand/supply calibration*
Segment I (I-495 WB to I-270 NB)	AM/PM peaks	1,493	512	217
	Off-peak	5,134	329	2,127
Segment II (I-270 SB to I-495 EB)	AM/PM peaks	1,314	170	4,689
	Off-peak	5,050	403	14,323
Segment III (I-495 EB to I-270 NB)	AM/PM peaks	1,358	56	76
	Off-peak	5,604	65	934
Segment IV (I-270 SB to I-495 WB)	AM/PM peaks	853	57	813
	Off-peak	3,204	263	2,814
All segments	AM/PM peaks	5,018	795	5,795
	Off-peak	18,992	1,060	20,198
	All day	24,010	1,855	25,994
Relative difference absolute	AM/PM peaks	N/A[a]	84.16%	15.50%
	Off-peak	N/A	94.42%	6.35%
	All day	N/A	92.27%	8.26%

[a] N/A: not applicable.

Table 6-6: Comparison of Measured and Simulated VMT on HOV Lanes. (Miles)

Table 6-6 summarizes the HOV lane performance results of the case study with the real-world network. The performance measurement used for HOV lanes is different from general-purpose lanes in this study. In particular, the HOV lane performance may vary significantly with operation hours, so we estimate the vehicle miles traveled on HOV lanes in AM/PM peaks and off-peak periods, respectively. Loop/microwave detectors are installed on HOV lanes of four segments shown in

Table 6-6. The absolutes of relative differences (absolute percentage of simulation and measurement difference to the field measurement) for all HOV segments are reduced to 15.50% and 6.35% for peaks and off-peak, respectively. Figure 6-5 shows the similarity in temporal profiles of simulated and measured HOV-lane VMT.
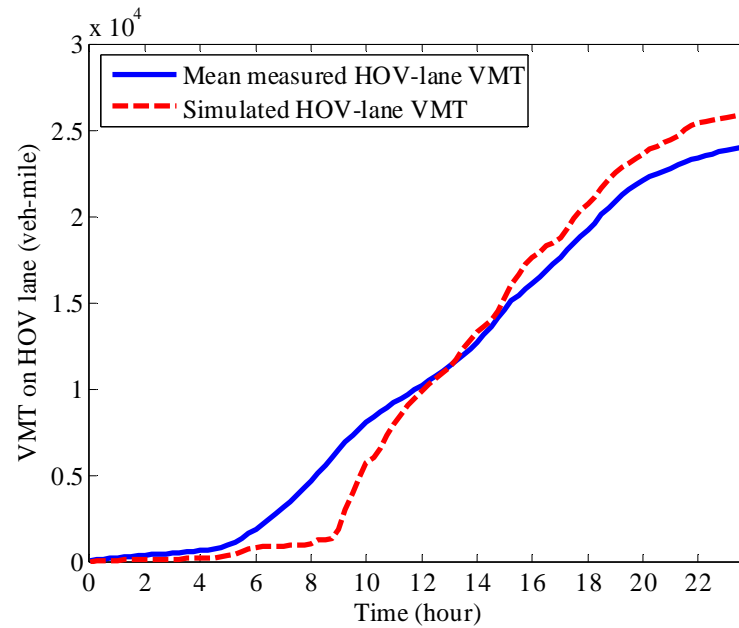
Figure 6-5: Comparison of Mean Measured and Simulated Cumulative HOV-Lane VMT.

6.3.4 SBO Results

We simulate the 5-hour PM peak from 14:00 to 19:00 of the corridor. The HOT rate takes effects from 15:30 to 18:30, while DMS provides congestion warning information to the work zone for the whole simulation period. The initial design points to fit the Bayesian stochastic Kriging surrogate model is generated by LHS. The initial sample $\mathbf{X}$ includes $4^3 = 64$ design points plus upper and lower bounds, and the baseline (neither HOT nor DMS implementations).

Sample	x_1 US\$	x_2 %	x_3 %	$\overline{y}$ min	$\hat{s}_\varepsilon$ min	Sample	x_1 US\$	x_2 %	x_3 %	$\overline{y}$ min	$\hat{s}_\varepsilon$ min
t	3.97	85.71	71.43	12.142	0.10	33	2.22	55.56	69.84	12.111	0.08
2	2.70	69.84	73.02	12.065	0.04	34	0.87	36.51	93.65	12.138	0.08
3	0.71	17.46	15.87	12.062	0.06	35	3.10	82.54	3.17	12.176	0.10
4	1.27	76.19	65.08	12.075	0.12	36	2.78	19.05	95.24	12.037	0.06
5	3.65	46.03	26.98	12.166	0.08	37	1.19	28.57	53.97	12.145	0.21
6	4.29	93.65	96.83	12.188	0.07	38	4.05	63.49	23.81	12.136	0.14
7	2.46	33.33	76.19	12.023	0.13	39	4.21	22.22	28.57	12.196	0.12
8	0.16	77.78	49.21	12.123	0.17	40	1.51	88.89	92.06	12.092	0.12
9	3.57	12.70	0.00	12.171	0.08	41	4.52	23.81	85.71	12.116	0.07
10	4.37	74.60	58.73	12.157	0.08	42	4.13	34.92	50.79	12.167	0.11
11	0.79	6.35	57.14	12.017	0.09	43	0.08	39.68	7.94	12.051	0.11
12	2.30	44.44	17.46	12.044	0.07	44	2.62	71.43	12.70	12.052	0.05
13	4.84	11.11	42.86	12.147	0.16	45	3.89	7.94	80.95	12.121	0.04
14	1.83	96.83	60.32	12.107	0.06	46	1.98	0.00	79.37	12.148	0.10
15	3.49	38.10	98.41	12.145	0.11	47	1.59	41.27	1.59	12.055	0.10
16	3.25	47.62	55.56	12.203	0.19	48	1.90	61.90	4.76	12.086	0.09
17	3.81	73.02	9.52	12.102	0.09	49	0.63	60.32	14.29	12.066	0.10
18	0.48	66.67	34.92	12.059	0.12	50	0.00	92.06	30.16	12.120	0.19
19	2.94	65.08	90.48	12.084	0.10	51	2.86	58.73	25.40	12.083	0.08
20	1.11	68.25	82.54	11.975	0.06	52	3.02	95.24	47.62	12.082	0.05
21	1.43	57.14	100.00	12.156	0.12	53	1.67	15.87	84.13	12.108	0.14
22	2.54	3.17	22.22	12.060	0.09	54	2.38	20.63	38.10	12.136	0.11
23	4.60	80.95	36.51	12.198	0.11	55	3.17	9.52	46.03	12.220	0.03
24	4.68	90.48	61.90	12.184	0.13	56	1.35	98.41	11.11	12.026	0.11
25	3.33	100.00	19.05	12.138	0.04	57	3.73	87.30	39.68	12.212	0.13
26	1.03	25.40	77.78	12.026	0.08	58	0.32	50.79	41.27	12.063	0.10
27	1.75	1.59	6.35	12.127	0.07	59	2.06	79.37	52.38	12.099	0.14
28	0.24	49.21	87.30	12.035	0.13	60	2.14	42.86	44.44	12.101	0.06
29	4.92	52.38	66.67	12.194	0.07	61	5.00	14.29	74.60	12.262	0.14
30	0.56	26.98	63.49	12.009	0.11	62	4.44	4.76	20.63	12.131	0.08
31	3.41	31.75	68.25	12.085	0.05	63	4.76	53.97	88.89	12.110	0.06
32	0.40	30.16	33.33	12.019	0.11	64	0.95	84.13	31.75	12.124	0.08
LB[a]	0.00	0.00	0.00	12.035	0.10	UB[b]	5.00	100	100	12.221	0.12
Baseline	∞	0.00	0.00	12.320	0.08	N/A[c]	N/A	N/A	N/A	N/A	N/A

[a] LB: lower bound; [b] UB: upper bound.

Table 6-7: Space-Filling Latin Hypercube Sampling of Parameters for DoE.

To reduce the influence of random simulation outputs, we run 5 replications for each design point, and each simulation run includes 10 iterations of DTA to

achieve the convergence. DynusT obtains valid results when the convergence is achieved after several times of assignments and vehicular platoon simulations. The average simulation takes around 63 min for each replication, and the relative gaps between two adjacent iterations for the DTA were found to be below 7% for all experiments. So the total computational time spent in Table 6-7 is over 350 hours. Take the baseline as an example, the mean output of the network-wide average travel time is $\bar{y} = 12.320$ min, and the estimated standard deviation is $\hat{s}_\varepsilon = 0.08$ min based on 5 replications.

We compare the mean objective function and the standard deviation of the optima with the baseline to see how much improvement can be achieved after optimization. Figure 6-6 shows the estimation results of the Bayesian stochastic Kriging surrogate model for each design point. Both estimated mean of the average travel time $\hat{\mathbf{f}}(\mathbf{x})$ and its standard error $\hat{\mathbf{s}}(\mathbf{x})$ are presented at these points $\mathbf{X}$ shown in Table 6-7. The dashed lines are bounds with one standard error given by $\hat{\mathbf{f}}(\mathbf{x}) \pm \hat{\mathbf{s}}(\mathbf{x})$. As a comparison, we plot the mean values $\bar{\mathbf{y}}$ and the estimated standard deviations $\hat{\mathbf{s}}_\varepsilon(\mathbf{x})$ of random observations as well. The design points in $\mathbf{X}$ are sorted according to estimated mean values of the average travel time in a descending order.
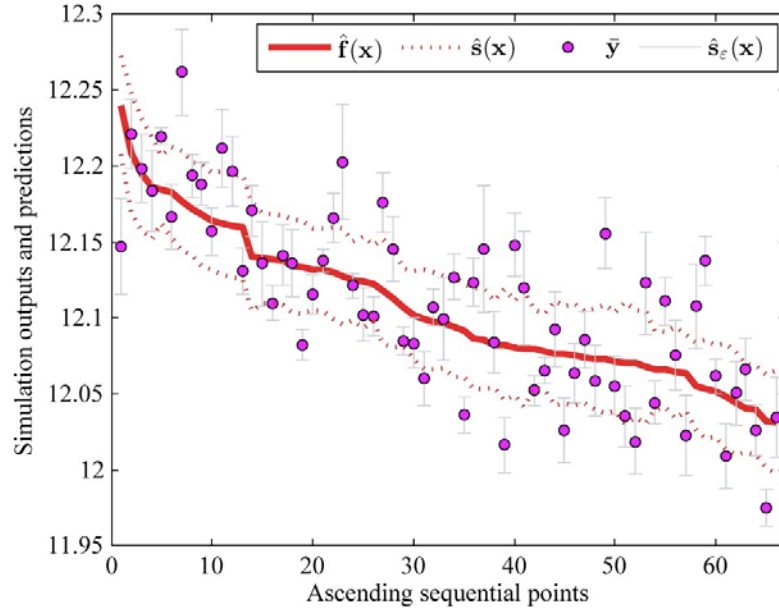
Figure 6-6: Prediction Accuracy of the Bayesian Stochastic Kriging.

To further compare the baseline and the optimal case, we run the simulation for 5 replications for the optima. Figure 6-7 shows the distributions of the baseline and the optimal solution, i.e. $\hat{\mathbf{x}}^* = [\text{US\$}\,0,\, 53.25\%,\, 80.27\%]^{\text{T}}$, which corresponds to the estimated mean value of the average trip travel time. The predictive distribution of the average trip travel time for the optima is $\mathcal{N}_{\text{Optima}}\left(12.020,\, 0.04^2\right)$. Results from the 5-replication simulation of the optimal inputs show that the mean value of outputs is 11.970 min, which is close to the predictive mean value. The standard deviation of the outputs is 0.05 min. So the optimization results suggest the average trip travel time for all users of the I-270 corridor reduces 12.320 min to 11.970 min.
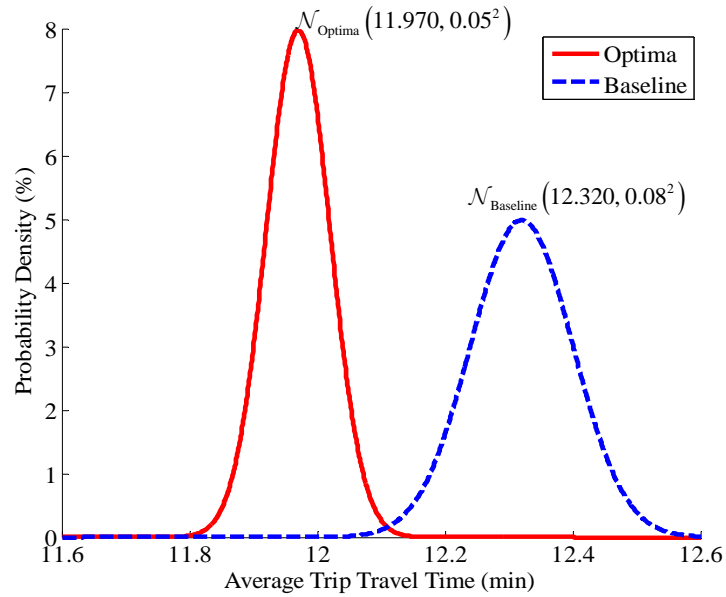
$$\mathcal{N}_{\text{Optima}}\left(11.970, 0.05^2\right)$$

$$\mathcal{N}_{\text{Baseline}}\left(12.320, 0.08^2\right)$$

Figure 6-7: Distributions of Simulation Outputs for the Baseline and Optima.

Scope	*Statistics*	*Baseline*	*Optima*	*Improvement*
Locally impacted vehicles (40,763 vehicles, 12.42%)	Average trip time of impact vehicles (min)	26.11	24.86	4.79%
System-wide impacts (328,314 vehicles, 100%)	Complete trips	302,475	302,918	0.15%
	Average overall trip time (min)	12.32	11.97	2.84%
	Average trip distance (mile)	4.94	4.95	-0.20%[a]
	Average travel speed (mph)	24.07	24.82	3.12%

[a] The negative value indicates no improvement.

Table 6-8: Comparison of the Baseline and Optima for PM Peak Simulation Results.

More network-wide statistics of the 5-hour simulation such as throughput, average overall trip time (including the demand loading time, stop/queueing time, and travelling time) is listed in Table 6-8. Moreover, the vehicles that passed through the work zone links before the work zone was placed in the network are identified as

locally impacted vehicles. The average trip travel time for those locally impacted vehicles under the baseline and optima is also listed in Table 6-8.
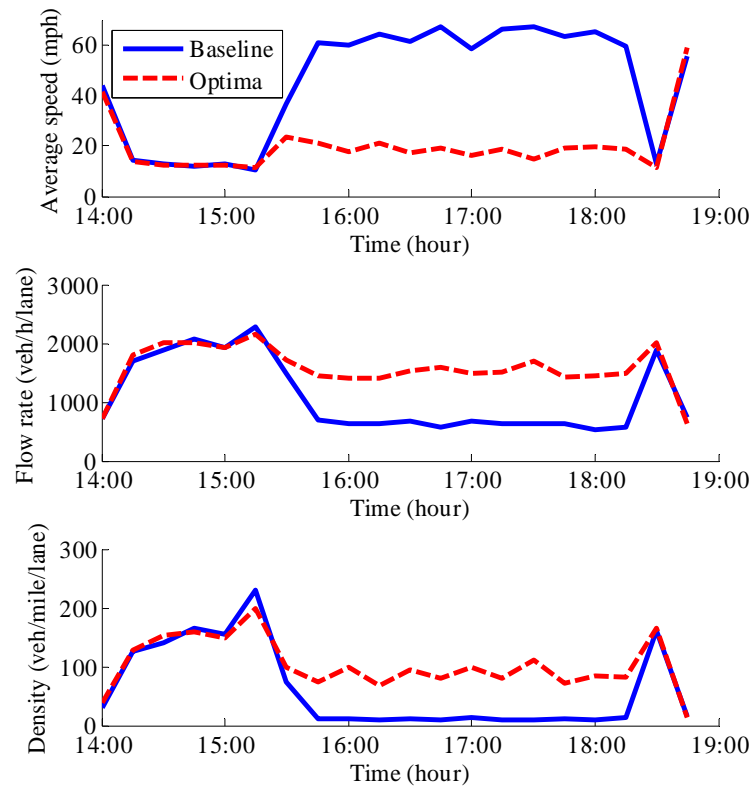

Figure 6-8: Comparisons of Traffic Flow Characteristics on the Work Zone HOV/HOT Lane.

The SOV is allowed to use the HOT lane without paying toll, which is to open the HOT lane to all vehicles during the operational time in the optima case, while SOV is restricted to be on the HOV lane in the baseline case. We further zoom in the HOV link parallel to the freeway mainline work zone (see the layout in Figure 6-3). As shown in Figure 8, the traffic volume on the HOT lane (optima) is larger than the HOV lane of the baseline after 15:30 when the managed lane takes effect. More vehicles are allocated to the HOT lane from the freeway mainline, which can release

the heavy congestion on the remaining three general-purpose lanes next to the work zone.

Then the mean values of statistics of these replications are utilized to demonstrate performance improvements. Figure 6-9(a) illustrates the average trip travel time in every 5 minutes for vehicles that complete their journeys over the entire corridor. The network average travel time is reduced in the optimal case than the baseline. The largest reduction in the average travel time occurs during 17:00 and 18:00, which is a part of the HOV/HOT operation hours. Thus the optimal HOT rate together with DMS implementations successfully help alleviate peak-hour congestion throughout the network. Figure 6-9(b) compares the total corridor throughputs of the optimal solutions and the baseline.
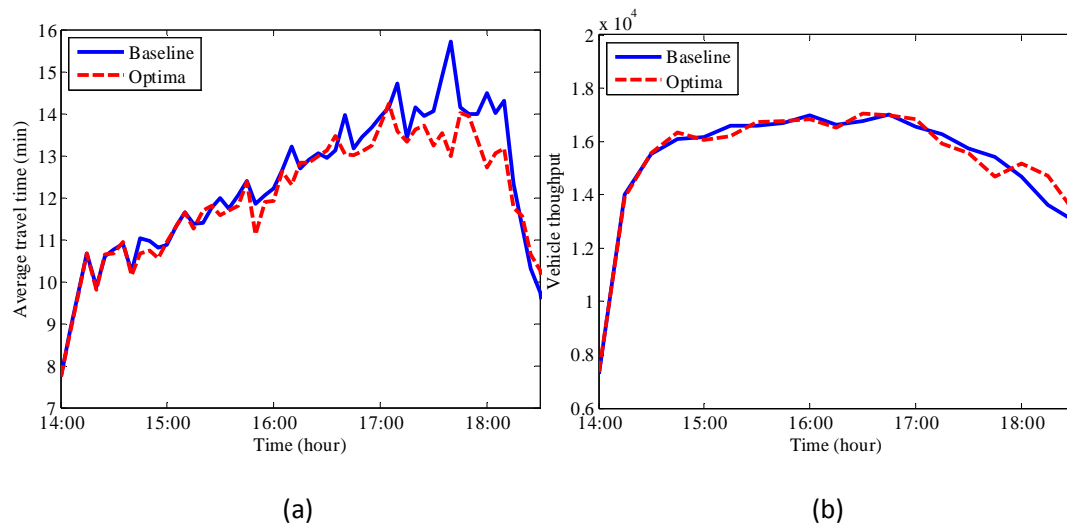

(a)   (b)

Figure 6-9: Comparison of the Baseline and Optima, (a) Average Travel Time; (b) Vehicle Throughput.

154

As shown in Table 6-8, the optima increase the cumulative throughput slightly by 0.15% during the study period. The small improvement in the average travel time of all users in corridor (2.84% reduction) corresponds to more than 26 thousand dollars saved when we use the value of time as US$ 15/hour for the 5-hour PM peak simulations. The benefits can be even larger if we consider 24 hours or a long-term effect.

The 2.84% reduction in travel time represents the average improvement to all corridor travelers caused by the optimal integrated transportation planning and operations strategies. If we focus on all vehicles that originally travel through the work zone link when the work zone was not set up (noted as locally impacted vehicles in Table 6-8), the percentage improvement is even more remarkable, which is 4.79% as shown in Table 6-8.

6.4 Conclusions

On the basis of the regressing Kriging model, a Bayesian stochastic Kriging metamodel is developed in this chapter, which assumes a quadratic form of global trend and different levels of variance for the random simulation error at different locations of the input domain. The heteroscedasticity of the stochastic simulation output is thus taken into account.

A synthetic network is built in DynusT and used to test the performance of the proposed Bayesian stochastic Kriging model. Through comparing the goodness-of-fit of the proposed model with three other surrogate models, i.e. quadratic polynomial

response surface method, ordinary Kriging and regressing Kriging, we find that the Bayesian stochastic Kriging model outperforms the other three models in both estimating the mean values and standard errors for a heteroscedastic simulation input-output relationship.

This proposed method is then applied to jointly optimize a transportation planning strategy (i.e. HOT toll) and a traffic control measure (i.e. DMS) for a freeway-arterial corridor, which is a very difficult problem when simulation is not utilized. The predicted optimal solution is shown to improve the system performance by 2.84%, and the percentage improvement is 4.79% for locally impacted vehicles.

However, to achieve computational convenience, simulation noise is assumed to be normally distributed during the developing of the Bayesian stochastic Kriging model. This assumption may not conform to the simulation outputs. The validity of this assumption will be analyzed in the next chapter.

Chapter 7: Bootstrapping of an Enhanced SVR Method Considering Distribution of Simulation Noise

Observing the asymmetric distribution of simulation noise in the real world applications introduced in the previous chapters, we developed a new enhanced support vector regression model (named distribution-based SVR hereafter) that takes into account the noise distribution in this chapter. The penalty of prediction error is designed to be related to the probability density function of simulation noise. The assumption of normal distribution for simulation error in the Bayesian stochastic Kriging model is thus relaxed. In addition, to improve its capability in supporting global optimization, bootstrapping method is employed to estimate the variance of prediction provided by the distribution-based SVR model. The expected improvement infill strategy can then be applied to search for the new design points.

7.1 Asymmetric Distribution of Simulation Noise

In addition to the characteristics of heteroscedasticity in transportation simulation noise introduced in Chapter 5, we further investigated the distribution of simulation noise for the application illustrated in Chapter 6.

We randomly selected several design points from the initial sample, and replicated the simulation evaluation of those selected points by 75 times. The empirical simulation noise distributions at those design points are presented in Figure 7-1. The unit of simulation noise is minute.
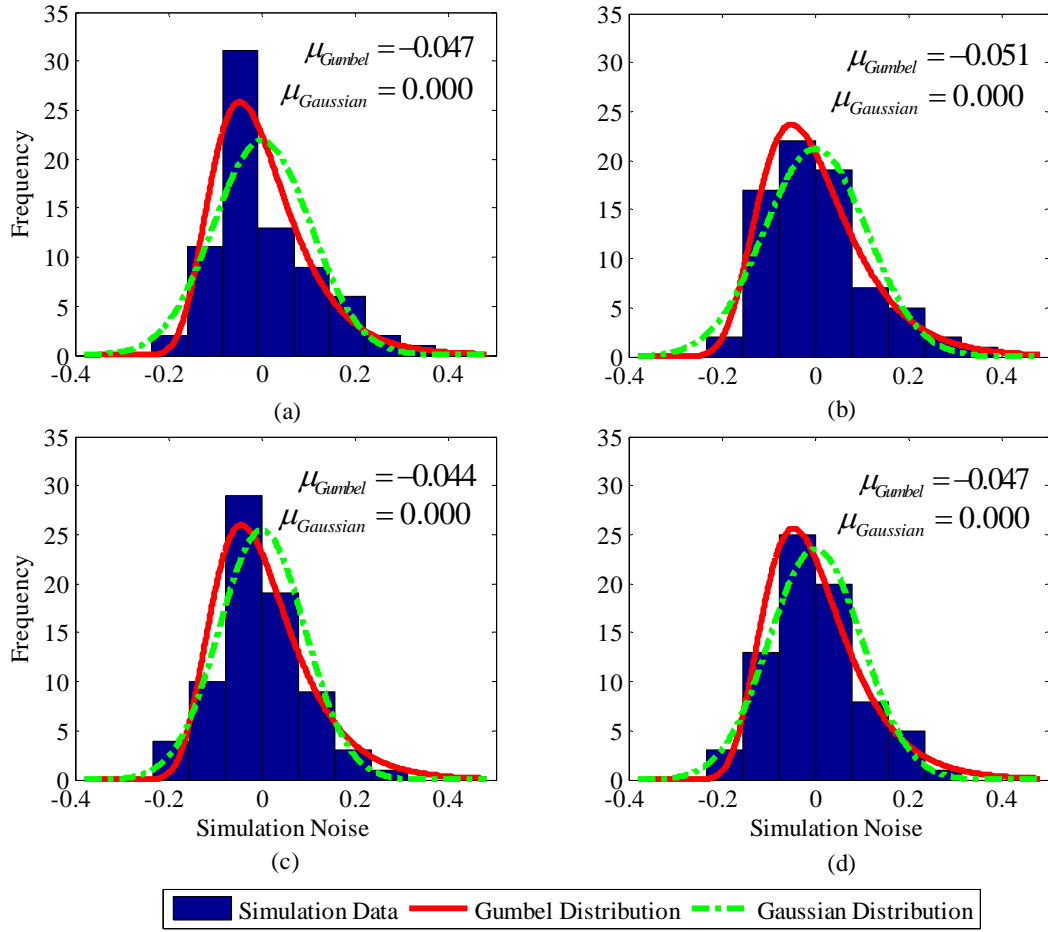
Figure 7-1: Distribution of Simulation Noise at Four Randomly Selected Design Points.

At any design point, the average simulation output is used as the estimate of the true response. Thus the mean of simulation noise is zero. It is clearly shown in Figure 7-1 that the distribution of simulation noise at any of the design points is not symmetric. In general, the distribution is positively skewed and the mode of the distribution is negative.

In order to develop surrogate models that take into account the distribution of simulation noise, the simulation data needs to be fit to appropriate parametric distributions. The candidate parametric distributions should be skewed. We reviewed

several commonly used skewed parametric distributions, and fit the observation at the four design points depicted in Figure 7-1 to those distributions. The asymmetric distributions which are investigated are summarized in Table 7-1.

Design Points	p-value			
	Gaussian Distribution	Gumbel Distribution	Lognormal Distribution	Weibull Distribution
Point 1	0.2385	**0.8109**	0.8069	0.7184
Point 2	0.9380	0.9807	**0.9899**	0.9617
Point 3	0.4449	**0.8358**	0.8341	0.6160
Point 4	0.6525	**0.9706**	0.9669	0.9355

Table 7-1: Chi-Square Goodness-of-Fit Test Results

The p-value for Gumbel distribution and lognormal distribution at the four randomly selected design points is very close, and significantly larger than that for the other two distributions. As small values of p cast doubt on the validity of the null hypothesis, the results in Table 7-1 suggest Gumbel distribution and lognormal distribution can fit the noisy simulation output better than the other two distributions. Gumbel distribution provides comparable performance to lognormal distribution in fitting the noisy data.

One difference between Gumbel distribution and lognormal distribution is that the support of the former distribution is from negative infinity to positive infinity, while the support of the latter distribution is from 0 to positive infinity. Before fitted to lognormal distribution, the simulation noise data needs to be shifted to make sure all values are positive. When the number of observations is large, the optimal offset along with the parameters of the lognormal distribution can be estimated through maximum likelihood estimation. However, when surrogate based optimization is conducted, the number of observations at each design point is usually very limited

(the number of observations is 5 for both the numerical example and real world application introduced in this chapter). In this situation, it is difficult to determine the optimal offset. On the other hand, when we assume the simulation noise follows Gumbel distribution, only two parameters (location parameter and scale parameter) need to be estimated. Based on the estimation of standard deviation from the limited number of observations at the design point, the parameters of the Gumbel distribution can be easily derived.

Therefore, Gumbel distribution is selected in this research and the methodology developed in this chapter is based on the assumption that simulation noise is Gumbel-distributed. The simulation data in Figure 7-1 has been fitted to both Gumbel distribution and Gaussian distribution. The modes of the fitted Gumbel distribution and Gaussian distribution are illustrated in the figure. It is obvious that Gumbel distribution fits the data much better than Gaussian distribution. Thus the common assumption of normally distributed simulation noise in most surrogate models does not hold any more.

In the following section, an enhanced support vector regression model that approximates response surface of simulations with Gumbel-distributed noise will be developed.

7.2 Model Development and Evaluation

The traditional SVR model tolerates prediction error within a pre-specified band $[-\varepsilon, \varepsilon]$. When the prediction error lies outside the region defined by the band,

the penalty (presented as loss hereafter) is assumed to be linearly related to the prediction error. This loss function may not perform well when the distribution of simulation output is not symmetric.

In this section, we investigate the appropriate way to formulate the loss function, and develop the proper SVR model for Gumbel-distributed simulation output data.

7.2.1 Loss Function

To estimate the parameters of an unknown model with noisy data, a widely used method is maximum likelihood estimation (MLE). The likelihood of a particular dataset is defined as the joint probability density of the input-output data conditional on the underlying function f, which can be expressed as follows:

$$L(f) = p(\{\mathbf{x_1}, \cdots, \mathbf{x}_n\}, \{y_1, \cdots, y_n\} \mid f) = \prod_{i=1}^{n} p(\mathbf{x}_i, y_i \mid f) = \prod_{i=1}^{n} p(y_i \mid \mathbf{x}_i, f) p(\mathbf{x}_i) \quad (7.1)$$

Because the choice of input variables $\mathbf{x}_i$ does not depend on the underlying function f, the likelihood can be further simplified as

$$L(f) = \prod_{i=1}^{n} p(y_i \mid \mathbf{x}_i, f) \quad (7.2)$$

For the ease of computation, the maximization of likelihood is usually converted to the minimization of negative log-likelihood in practice, which is

$$\tilde{L}(f) = \sum_{i=1}^{n} -\ln p(y_i \mid \mathbf{x}_i, f) = \sum_{i=1}^{n} -\ln p(y_i - f(\mathbf{x}_i)) \quad (7.3)$$

As introduced in Chapter 5, the traditional SVR is a method that balances between model smoothness and loss due to prediction error. According to Equation 5.4a, the objective function of the traditional SVR model includes two terms. The first term $\frac{1}{2}\|\mathbf{w}\|^2$ represents model complexity, and the second term $\frac{C}{n}\sum_{i=1}^{n}(\xi^{+(i)}+\xi^{-(i)})$ represents the loss due to prediction error. The ideal way to choose the loss function is to make the minimization of negative log-likelihood coincide with the minimization of the total prediction loss at the design points. Thus the loss function can be chosen as follows:

$$loss(\mathbf{x}_i, y_i, f) = -\ln p(y_i - f(\mathbf{x}_i)), \ i=1, 2, \ldots, n \tag{7.4}$$

More specifically, the loss function for the two cases that the prediction $f(\mathbf{x}_i)$ is smaller than the observation y_i and the prediction $f(\mathbf{x}_i)$ is larger than the observation y_i can be expressed in Equation 7.5.

$$\begin{cases} loss(\xi^{+(i)}) = -\ln p(\xi^{+(i)}) & \text{if } y_i - f(\mathbf{x}_i) \geq 0, \ \xi^{+(i)} = y_i - f(\mathbf{x}_i), \ i=1, 2, \ldots, n \\ loss(\xi^{-(i)}) = -\ln p(-\xi^{-(i)}) & \text{if } y_i - f(\mathbf{x}_i) < 0, \ \xi^{-(i)} = f(\mathbf{x}_i) - y_i, \ i=1, 2, \ldots, n \end{cases} \tag{7.5}$$

7.2.2 Distribution-Based Support Vector Regression

There are two types of Gumbel distribution, which are called Gumbel Minimum distribution and Gumbel Maximum distribution. The probability density functions of these two distributions are very similar. The only difference between these two distributions is that Gumbel Minimum distribution is negatively skewed while Gumbel Maximum distribution is positively skewed. According to Figure 7-1, the distribution of the I-270 corridor simulation outputs is positively skewed. Thus

the distribution-based SVR for data with Gumbel Maximum distribution is developed in this section. Distribution-based SVR for data with negatively skewed distribution can be formulated and solved in a similar manner.

The probability density function of a Gumbel Maximum distribution is

$$p(x) = \frac{1}{\beta_{Gumbel}} e^{-(z+e^{-z})}, \ x \in \mathbb{R} \tag{7.6}$$

where $z = \dfrac{x - \mu_{Gumbel}}{\beta_{Gumbel}}$. μ_{Gumbel} is the location parameter, and β_{Gumbel} is the scale parameter. $\mathbb{R}$ is the real set. Thus the loss function corresponding to the Gumbel Maximum distribution at any location $\mathbf{x}_i$, $i = 1, 2, \ldots, n$ is

$$\begin{cases} loss(\xi^{+(i)}) = \dfrac{\xi^{+(i)} - \mu_{Gumbel}^{(i)}}{\beta_{Gumbel}^{(i)}} + e^{-\frac{\xi^{+(i)} - \mu_{Gumbel}^{(i)}}{\beta_{Gumbel}^{(i)}}} + \ln(\beta_{Gumbel}^{(i)}) & \text{if } y_i - f(\mathbf{x}_i) \geq 0, \ \xi^{+(i)} = y_i - f(\mathbf{x}_i) \\[4mm] loss(\xi^{-(i)}) = -\dfrac{\xi^{-(i)} + \mu_{Gumbel}^{(i)}}{\beta_{Gumbel}^{(i)}} + e^{\frac{\xi^{-(i)} + \mu_{Gumbel}^{(i)}}{\beta_{Gumbel}^{(i)}}} + \ln(\beta_{Gumbel}^{(i)}) & \text{if } y_i - f(\mathbf{x}_i) < 0, \ \xi^{-(i)} = f(\mathbf{x}_i) - y_i \end{cases}$$

$$\tag{7.7}$$

In this section, the tolerance band of prediction error (ε) is set to be 0. As the distribution of simulation noise is asymmetric, more studies need to be conducted for the proper selection of non-zero ε. Similar to the traditional SVR presented in Equation 5.4, the Gumbel Maximum distribution-based SVR can be formulated as

$$\min \ \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{n} \sum_{i=1}^{n} (loss(\xi^{+(i)}) + loss(\xi^{-(i)})) \tag{7.8a}$$

$$\text{s.t.} \ \ y^{(i)} - \mathbf{w}^{\mathrm{T}} \boldsymbol{\varphi}^{(i)} - b \leq \xi^{+(i)}, \ i = 1, 2, \ldots, n \tag{7.8b}$$

$$\mathbf{w}^{\mathrm{T}} \boldsymbol{\varphi}^{(i)} + \mu^{(i)} - b \leq \xi^{-(i)}, \ i = 1, 2, \ldots, n \tag{7.8c}$$

$$\xi^{+(i)}, \xi^{-(i)} \geq 0, \ i = 1, 2, \ldots, n \tag{7.8d}$$

Following the same derivation process as described in Chapter 5, the Gumbel
Maximum distribution-based SVR model finally becomes an optimization problem of
dual variables.

$$
\min \quad \frac{1}{2}\sum_{i,j=1}^{n}(\alpha^{+(i)}-\alpha^{-(i)})(\alpha^{+(j)}-\alpha^{-(j)})\varphi(\mathbf{x}^{(i)},\mathbf{x}^{(j)})-\sum_{i=1}^{n}y^{(i)}(\alpha^{+(i)}-\alpha^{-(i)})
$$

$$
-\frac{C}{n}\sum_{i=1}^{n}(loss(\xi^{+(i)})-\xi^{+(i)}\bullet\partial_{\xi}(loss(\xi^{+(i)}))) \tag{7.9a}
$$

$$
-\frac{C}{n}\sum_{i=1}^{n}(loss(\xi^{-(i)})-\xi^{-(i)}\bullet\partial_{\xi}(loss(\xi^{-(i)})))
$$

$$
\text{s.t.} \quad \sum_{i=1}^{n}(\alpha^{+(i)}-\alpha^{-(i)})=0 \tag{7.9b}
$$

$$
\alpha^{\pm(i)}\leq\frac{C}{n}\partial_{\xi}(loss(\xi^{\pm(i)})), \ i=1,2,\ldots,n \tag{7.9c}
$$

$$
\xi^{\pm(i)}=\inf\{\xi\,|\,\frac{C}{n}\partial_{\xi}(loss(\xi^{\pm(i)}))\geq\alpha^{\pm(i)}\}, \ i=1,2,\ldots,n \tag{7.9d}
$$

$$
\alpha^{\pm(i)},\xi^{\pm(i)}\geq0, \ i=1,2,\ldots,n \tag{7.9e}
$$

When the prediction $f(\mathbf{x}_i)$ at the design point $\mathbf{x}_i$ is smaller than the
observation y_i, according to Equation 7.9d, the prediction error $\xi^{+(i)}$ at that point can
be presented as

$$
\xi^{+(i)}=\mu_{Gumbel}^{(i)}-\beta_{Gumbel}^{(i)}\bullet\ln(1-\frac{n\beta_{Gumbel}^{(i)}}{C}\bullet\alpha^{+(i)}) \tag{7.10}
$$

Thus the third term in Equation 7.9a at the specific design point $\mathbf{x}_i$ can be expressed
as

$$-\frac{C}{n}(loss(\xi^{+(i)}) - \xi^{+(i)} \bullet \partial_\xi(loss(\xi^{+(i)})))$$

$$= -\frac{C}{n} \bullet (\frac{n\beta_{Gumbel}^{(i)}\alpha^{+(i)}}{C} \bullet \ln(1 - \frac{n\beta_{Gumbel}^{(i)}\alpha^{+(i)}}{C}) - \ln(1 - \frac{n\beta_{Gumbel}^{(i)}\alpha^{+(i)}}{C}) - \frac{n(\mu_{Gumbel}^{(i)} + \beta_{Gumbel}^{(i)})\alpha^{+(i)}}{C}$$

$$(7.11)$$

The constant terms which are not dependent on the dual variable $\alpha^{+(i)}$ are removed from Equation 7.11, as these terms would not influence the solution of the optimization problem presented in Equation 7.9.

One major advantage of the traditional SVR is its computational efficiency by turning the estimation of the model into a quadratic programming problem. While Equation 7.11 includes the logarithm form of the dual variable $\alpha^{+(i)}$, which makes the objective function presented in Equation 7.9 very complex and the solving of the optimization problem very computationally expensive. Therefore, we approximate the terms with logarithm form of variable $\alpha^{+(i)}$ by their Taylor series expansion to the second order. Equation 7.11 then approximately equals to

$$-\frac{C}{n}(loss(\xi^{+(i)}) - \xi^{+(i)} \bullet \partial_\xi(loss(\xi^{+(i)}))) = \frac{n\beta_{Gumbel}^{(i)}(\alpha^{+(i)})^2}{2 \bullet C} + \frac{n\mu_{Gumbel}^{(i)}\alpha^{+(i)}}{C} \qquad (7.12)$$

Through similarly processing the fourth term in Equation 7.9a, the following optimization problem is derived.

$$\min \quad \frac{1}{2}\sum_{i,j=1}^{n}(\alpha^{+(i)} - \alpha^{-(i)})(\alpha^{+(j)} - \alpha^{-(j)})\varphi(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) - \sum_{i=1}^{n}y^{(i)}(\alpha^{+(i)} - \alpha^{-(i)})$$

$$+ \sum_{i=1}^{n}(\frac{n(\beta_{Gumbel}^{(i)})^2}{2 \bullet C} \bullet (\alpha^{+(i)})^2 + \mu_{Gumbel}^{(i)} \bullet \alpha^{+(i)}) + \sum_{i=1}^{n}(\frac{n(\beta_{Gumbel}^{(i)})^2}{2 \bullet C} \bullet (\alpha^{-(i)})^2 - \mu_{Gumbel}^{(i)} \bullet \alpha^{-(i)})$$

$$(7.13a)$$

$$\text{s.t.} \quad \sum_{i=1}^{n}(\alpha^{+(i)} - \alpha^{-(i)}) = 0 \qquad (7.13b)$$

$$\alpha^{+(i)} \in [0, \frac{C}{n\beta_{Gumbel}^{(i)}}], \alpha^{-(i)} \in [0, \infty), \ i = 1, 2, \ldots, n \qquad (7.13c)$$

The optimization problem presented in Equation 7.13 is still a quadratic programming problem, and it can be conveniently solved using convex optimization methods.

When the distribution of simulation output is negatively skewed, the simulation noise can be fitted to Gumbel Minimum distribution. We omit the derivation of Gumbel Minimum distribution-based SVR here because it is very similar to the derivation of Gumbel Maximum distribution-based SVR introduced in this section.

7.2.3 Numerical Test

Three surrogate models that can deal with heteroscedastic sample data are developed in this dissertation so far. In addition to the characteristics of heteroscedasticity, the noise of sample output is assumed to be Gumbel-distributed. The purpose of the numerical test is to analyze the capability of the newly proposed distribution-based SVR model in approximating response surfaces with skewed noise distribution.

As discussed in section 5.3, the overall prediction accuracy over the entire research domain is of more concern when no evolutionary process is incorporated. This section will only focus on the MoE RMSE, and compare the performance of the three types of surrogate models: heteroscedastic SVR, Bayesian stochastic Kriging and distribution-based SVR.

The test functions and scenario settings are the same with the numerical test in Section 5.3. The two-dimension function in equation 5.16 and the 6-dimension function in Equation 5.20 serve as the unknown underlying function. In the low-variance and high-variance scenarios, the standard deviations are set to be 0.1 and 0.5 times the true function value at each sample point, respectively. The only difference in the numerical test design from Section 5.3 is that the distribution of noise is not Gaussian anymore. Instead, the noise is assumed to be Gumbel Minimum-distributed. Take the low-variance case for the two-dimension function as an example. The observation with Gumbel Minimum-distributed noise at a certain point is obtained by

$$\tilde{f}_1(x_1, x_2) = f_1(x_1, x_2) + Gumbel(\mu_{Gumbel}, \beta_{Gumbel}), \quad x_1, x_2 \in [-2\pi, 2\pi] \tag{7.14}$$

The variance of a Gumbel Minimum distribution is $\dfrac{\pi^2}{6}\beta^2_{Gumbel}$. As the standard deviation is set to be $0.1 f_1(x_1, x_2)$, the scale parameter β_{Gumbel} can be calculated as $\dfrac{\sqrt{0.06} f_1(x_1, x_2)}{\pi}$. Moreover, the mean of a Gumbel Minimum distribution is $\mu_{Gumbel} - \beta_{Gumbel}\gamma$ (γ is Euler's constant, approximately equals to 0.577). As the mean of noise added to the function value is 0, the location parameter μ_{Gumbel} can be computed as $\dfrac{\sqrt{0.06} f_1(x_1, x_2)}{\pi}\gamma$.

When conducting the numerical test, the true standard deviation of noise is not known. We draw 5 random seeds at each of the sample point, and estimate the standard deviation from the sample at each point. The parameters μ_{Gumbel} and β_{Gumbel} that will be used in the estimation of the distribution-based SVR can then be

167

computed. For each of the scenarios, the three surrogate models are estimated with the same sample dataset, and are then validated with a separate while unified validation dataset. In order to investigate the significance of the differences in prediction accuracy among the three models, the whole process including creation of 5-replication dataset, model estimation and validation is repeated 10 times. The performance of the three models under the 8 scenarios is shown in Table 7-1.

Variance Level		low				High			
Test Function		2-variable function		6-variable function		2-variable function		6-variable function	
Sample Size		Small (25)	Large (60)	Small (65)	Large (380)	Small (25)	Large (60)	Small (65)	Large (380)
RMSE	*Heteroscedastic SVR (H-SVR)*	2.922	0.922	0.294	0.172	2.809	**1.903**	**0.256**	**0.206**
	Bayesian stochastic Kriging (BSK)	2.606	1.840	0.380	0.201	**2.691**	2.171	0.349	0.235
	Distribution-based SVR (D-SVR)	**2.414**	**0.897**	**0.239**	**0.170**	2.809	1.914	**0.256**	0.232
ANOVA between D-SVR and H-SVR (p value)		*0.000**	*0.013**	*0.000**	*0.029**	0.997	0.788	0.658	*0.000**
ANOVA between D-SVR and BSK (p value)		*0.000**	*0.000**	*0.000**	*0.000**	0.066	*0.000**	*0.000**	0.082

Note: * indicates the difference is significant at the 95% confidence level.

Table 7-2: Performance of Heteroscedastic SVR, Bayesian Stochastic Kriging and Distribution-Based SVR.

RMSE shown in Table 7-2 represents the average of RMSE for the 10-time model estimation and validation process. Under each of the scenarios, the lowest RMSE provided by the surrogate models is highlighted in bold. The most important finding from Table 7-2 is that the distribution-based SVR always outperforms the other two surrogate models when the variance level is relatively low. In addition, the significance of the differences in RMSE among the three models is tested through ANOVA. Under the scenarios with low variance, the p values for ANOVA between distribution-based SVR and heteroscedastic SVR and for ANOVA between

168

distribution-based SVR and Bayesian stochastic Kriging are all very close to 0, which indicates that the performance of distribution-based SVR is significantly better than that of heteroscedastic SVR and Bayesian stochastic Kriging.

7.3 Bootstrapping and Infill

The capability of accurately approximating the true response surface is crucial to the success in simulation-based optimization. However, the performance of a surrogate model in searching for the optimal solution is heavily dependent on the layout of the initial sample. If the initial sample covers the area around the true optima, it may be easy to find the optimal solution or a near optima solution. Otherwise, successfully finding the optima is very difficult. In this case, it is very necessary to incorporate a global optimal infill strategy to the surrogate model for the optimization task.

The expected improvement-based infill strategy has been recognized as an effective method in searching for the global optima in this dissertation and other existing literature. However, in order to apply the EI infill strategy, both prediction and the variance of the prediction at a new design point should be provided by the surrogate model. Nevertheless, traditional SVR models can only provide the prediction at new points. A supplement method to estimate the variance of prediction is required. In this chapter, we propose to incorporate the bootstrapping method with SVR for the estimation of predictor's variance, and then apply the EI based infill strategy for surrogate-based optimization.

7.3.1 Bootstrapping Method

Defined in Efron and Tibshirani (1993), bootstrapping is a method for assigning measures of accuracy to sample estimates. The measures of accuracy include bias, variance, confidence interval, prediction error, etc. In order to conduct surrogate-based optimization, surrogate models need to be estimated with sample input-output data to approximate the true response surface. According to observations of several real world applications introduced in previous chapters, the simulation used to evaluate the performance of transportation systems is stochastic. As a result, the surrogate model estimated with the sample output should be stochastic, and the prediction at any new points by the surrogate model should also be stochastic.

In the area of surrogate-based optimization, bootstrapping has been implemented in a few studies. Kleijnen (2014) used bootstrapping to estimate the variance of Kriging predictor. In Kleijnen et al. (2012), the Kriging predictor's variance estimated through bootstrapping was further used for the expected improvement-based global optimization. In a paper published earlier (Den Hertog, et al., 2006), the authors argued that the classic Kriging variance formula underestimates the predictor's variance, and used bootstrapping for the estimation of predictor's variance instead. The argument was then tested and verified through several artificial examples and a real-life case study. With regard to the SVR model, De Brabanter et al. (2011) estimated the confidence and prediction intervals for least squares support vector regression using bootstrapping.

Non-parametric bootstrapping is applied in this study. At each of the n initial design points $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}$, R-repetitions of simulations are performed. Thus R observations of output at each design point are retrieved, which can be expressed as $y_r^{(i)}$, $i = 1, 2, \ldots, n$, and $r = 1, 2, \ldots, R$.

Bootstrapping is generally a resampling method. To form a bootstrap sample, at each of the design points $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}$, one out of the R observations is randomly selected as the output, expressed as $y^{(i)*}$, $i = 1, 2, \ldots, n$. This bootstrap sample is used to estimate the D-SVR model as illustrated in Equation 7.13. The estimated D-SVR model based on the bootstrap sample is expressed as $\hat{f}^*(\mathbf{x})$. The whole process of resampling as well as the estimation of the D-SVR model is repeated B times. Therefore, at any new point $\mathbf{x}^{(n+1)}$, B predictions $\hat{y}_b^{(n+1)}$, $b = 1, 2, \ldots, B$ can be given by the B estimated D-SVR model. With these B predictions, the predictor at the new point $\mathbf{x}^{(n+1)}$ can be computed as

$$\hat{\bar{y}}_B^{(n+1)} = \frac{\sum_{b=1}^{B} \hat{y}_b^{(n+1)}}{B} \tag{7.15}$$

and the predictor's variance is

$$(s_B^{(n+1)})^2 = \frac{\sum_{b=1}^{B} (\hat{y}_b^{(n+1)} - \hat{\bar{y}}_B^{(n+1)})^2}{B-1} \tag{7.16}$$

7.3.2 Expected Improvement-Based Infill

The framework of incorporating the bootstrapping method to the D-SVR model for the surrogate-based optimization with the expected improvement infill strategy is illustrated in Figure 7-2.
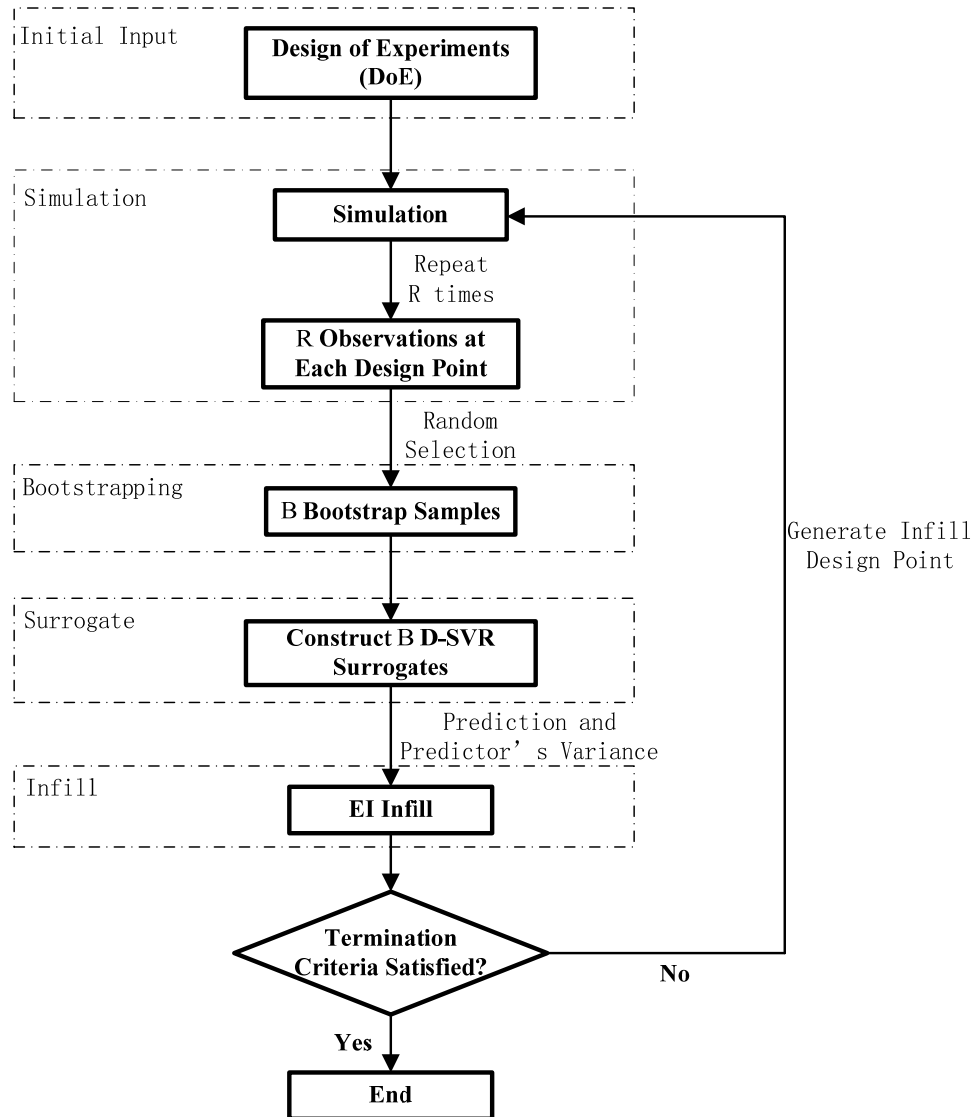

Figure 7-2: Framework of the Expected Improvement-Based Infill with Bootstrapped Distribution-Based SVR Surrogate Model.

The expected improvement-based infill strategy aims to find the new points with the maximal expected improvement compared to the minimal observation at the n old sample points. In the current study, the mean of the observations at each of the old sample points is used as the estimator of the true response, which is $\overline{y}^{(i)} = \dfrac{\sum\limits_{r=1}^{R} y_r^{(i)}}{R}$.

Thus the minimal observation at old sample points is $\min_i(\overline{y}^{(i)})$. Meanwhile, the infill strategy assumes that the distribution of prediction at a new point is Gaussian with mean $\hat{\hat{y}}_B^{(n+1)}$ and variance $(s_B^{(n+1)})^2$. The method of computing the EI can then be slightly revised from Equation 3.20 to the following,

$$\mathrm{EI}(\mathbf{x}^{(n+1)}) = \frac{1}{\sqrt{2\pi}s_B^{(n+1)}} \int_{-\infty}^{\min_i(\overline{y}^{(i)})} (\min_i(\overline{y}^{(i)}) - u)\exp\left(-\frac{(u - \hat{\hat{y}}_B^{(n+1)})^2}{2(s_B^{(n+1)})^2}\right)\mathrm{d}u \qquad (7.17)$$

7.3.3 Numerical Example

The effectiveness of global optimization with the EI infill strategy and bootstrapped D-SVR surrogate model is tested in this section using a numerical example. In addition, as it is verified through numerical examples in Section 7.2.3, when the simulation noise is not symmetrically distributed, the D-SVR model can approximate the true response surface more accurately than other surrogate models that do not consider the distribution of simulation noise. Under the same assumption regarding the simulation noise, it is reasonable to hypothesize that the D-SVR would also be more efficient in searching for the optimal solution when infill is applied than other models. Thus we also compare the performance of the EI infill D-SVR model in

searching for the optima with that of the Bayesian stochastic Kriging model. The infill strategy for the BSK model is also maximizing the EI. Instead of applying bootstrapping, the predictor's variance for the BSK model is estimated through Bayesian analysis as introduced in Chapter 6.
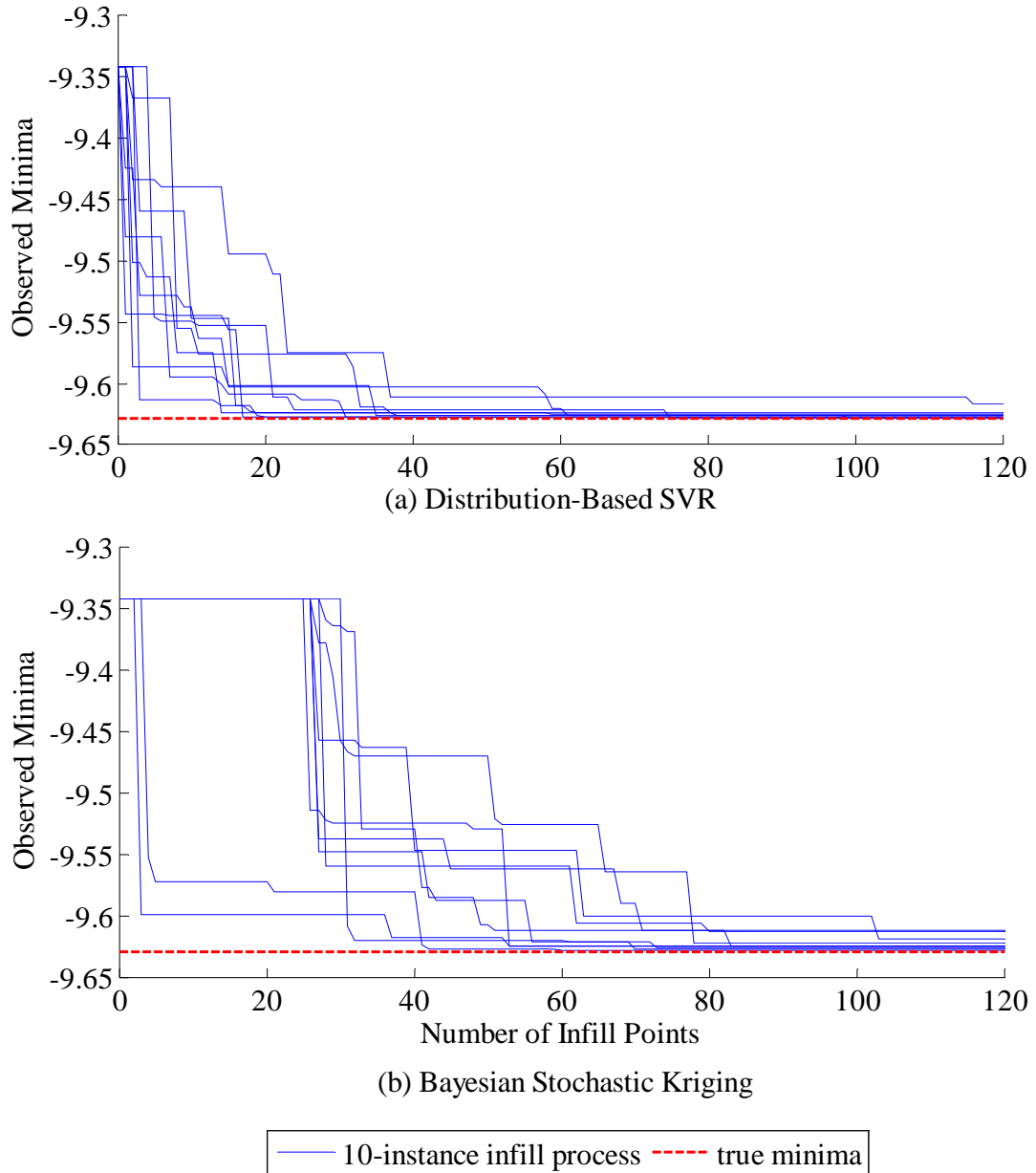

Figure 7-3: Comparison of the EI Infill Global Optimization with Distribution-Based SVR and Bayesian Stochastic Kriging.

Among the 8 scenarios introduced in Section 7.2.3, the case of the 2-dimension function with a low variance level for response noise and a large sample size is used for this infill study. The initial sample size is 60. The infill process will terminate after 120 infill points are generated. For each of the design points, 5 replications of simulations are conducted. The bootstrap sample size B for the D-SVR method is set to be 100. In order to make the comparison between the two methods more robust, the whole procedure as illustrated in Figure 7-2 from the generation of initial sample, construction of surrogate models to the iterative infill process is repeated 10 times. The comparison of the performance in global optimization with the two methods is shown in Figure 7-3.

Both methods are shown to be effective in finding the optimal solution through the EI global infill strategy. Due to the simulation noise, significant variation exhibits among the 10 instances of the optimization process for both methods. Overall, along with the infill of new points, the optimal solution identified by both methods converges to the true optima. However, the minimal observed value found by the D-SVR model declines faster than the BSK model along the infill process, which indicates that the D-SVR model is more efficient in finding the optimal solution than the BSK model.

The infill process with both the D-SVR model and the BSK model is averaged over the 10 instances and compared in Figure 7-4. It shows more clearly that the D-SVR model can find the optimal solution much faster than the BSK model. With less than 20 iterations of infill, the D-SVR model is successful in finding solutions with

objective value smaller than -9.6, while the BSK model takes around 60 iterations of infill to find comparable solutions. In other words, with very tight computational budget, the D-SVR model will be able to provide much better solutions than the BSK model.
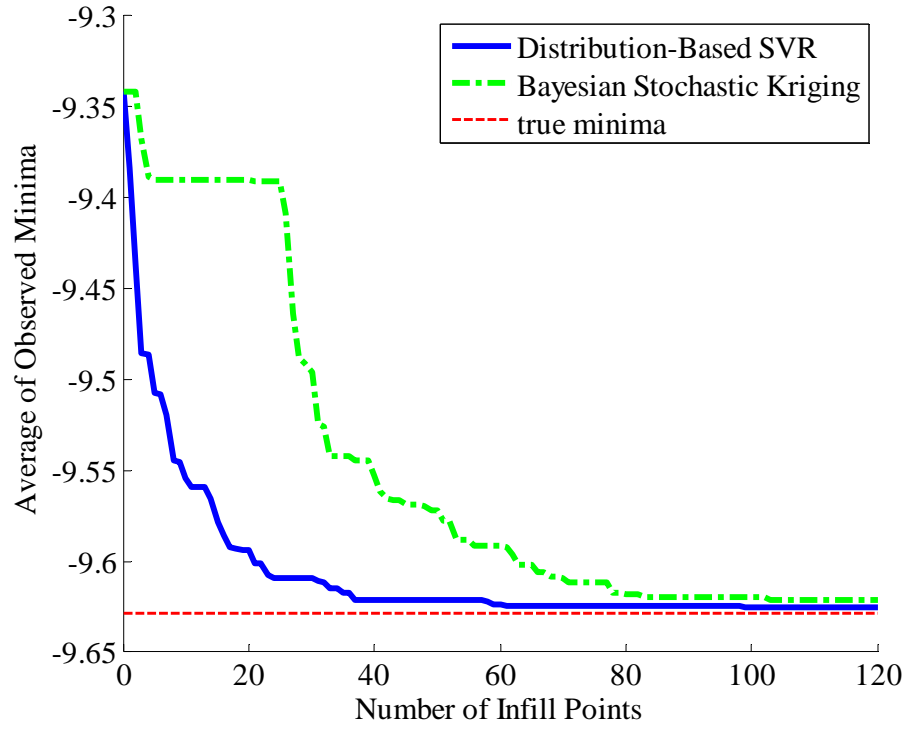

Figure 7-4: Comparison of the EI Infill Global Optimization with Distribution-Based SVR and Bayesian Stochastic Kriging (Averaged over 10 Instances).

7.4 Application

The case study investigated in this section is the same as that introduced in Section 6.3. The purpose of the case study is to jointly optimize the HOT toll and the freeway diversion control for a freeway-arterial corridor network, with the objective of minimizing the average trip travel time for all users of the network.

Similar to the numerical test, both the distribution-based SVR model and the Bayesian stochastic Kriging model are applied for this case study. The same 67 initial sample points are used for the estimation of the surrogate models. In addition, 20 iterations of infill will be conducted to search for the optimal solution. One thing should be noted is that the simulation noise is Gumbel Maximum-distributed as shown in Figure 7-1, thus the D-SVR model used for this study is Gumbel Maximum-distribution-based SVR.

After 20 iterations of infill, the optimal solution provided by the BSK model is $\hat{\mathbf{x}}_{BSK}^{*} = [\$\,0.00, 69.22\%, 7.36\%]^{\mathrm{T}}$. We then run 10 replications of simulations for this optimal input. The average trip travel time retrieved from the output from the 10 replications is

$$y_{BSK}^{*} = (12.005, 11.930, 11.943, 11.945, 11.977, 12.032, 11.916, 12.058, 11.932, 11.946)$$

minutes. The mean average trip travel time is 11.968 min, and the standard deviation of the output is 0.05 min.

The optimal solution provided by the Gumbel Maximum-distribution-based SVR model after 20 infill is $\hat{\mathbf{x}}_{D-SVR}^{*} = [\$\,1.30, 76.19\%, 65.08\%]^{\mathrm{T}}$, and the average trip travel time got from the 10-replication simulation is

$$y_{D-SVR}^{*} = (11.960, 11.893, 11.917, 11.918, 11.939, 12.002, 11.883, 11.929, 11.898, 12.028)$$

minutes. The mean of the responses is 11.937 min, and the standard is 0.05 min.
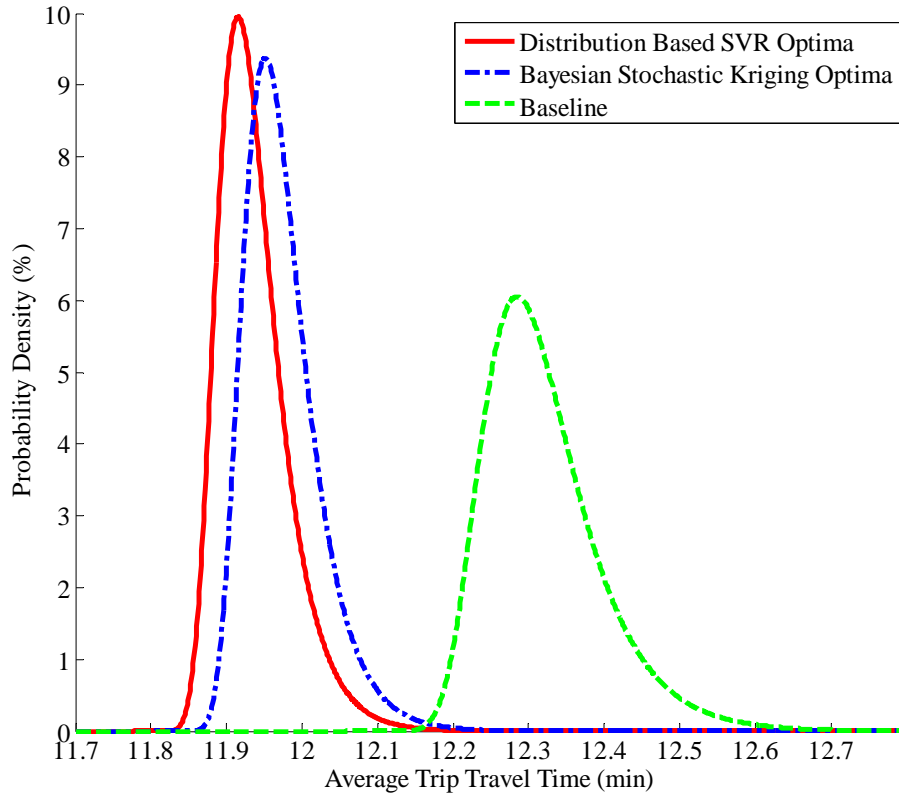
Figure 7-5: Distributions of the Baseline and Optima Predicted by the D-SVR and BSK Models.

As the distribution of the output is not Gaussian, it is not appropriate to test the statistical significance of the difference between the optimal objective values computed from the two models. With the assumption that the noise follows Gumbel Maximum distribution and the mean of noise is 0, we plot the distribution of the optima provided by the two models and compare them visually in Figure 7-5.

It shows clearly that both of the optimal solutions predicted by the D-SVR model and the BSK model are significantly better than the baseline. To test the statistical significance of the difference between the mean output provided by the D-SVR optima and the BSK optima, the Mann-Whitney U test is conducted. Results

show that the average trip travel time under the D-SVR optima is significantly lower

than that under the BSK optima at the 90% confidence level.

The performances of the entire corridor network and the locally impacted

vehicles of the baseline and the optima predicted by the two surrogate methods are

summarized and compared in Table 7-3.

Scope	Statistics	Baseline	BSK Optima		D-SVR Optima	
			Performance	Improvement	Performance	Improvement
Locally impacted vehicles	Average trip time of impact vehicles (min)	26.110	24.393	6.58%	24.110	7.66%
System-wide impacts	Complete trips	302,475	304,859	0.79%	305,084	0.86%
	Average overall trip time (min)	12.320	11.968	2.86%	11.937	3.11%
	Average trip distance (mile)	4.94	4.95	-0.20%[a]	4.94	0
	Average travel speed (mph)	24.07	24.82	3.12%	24.83	3.16%

[a] The negative value indicates no improvement.

Table 7-3: Comparison of the Baseline, BSK Optima and D-SVR Optima for PM
Peak Simulation Results.

The optimal solutions provided by both methods improve the performance of

the corridor network compared to the baseline, which refers to the case that no effort

is made in HOT lane management or freeway diversion control. The improvement in

terms of average trip travel time to all of the corridor users made by the D-SVR

optima is a bit larger than that made by the BSK optima. Although the difference

looks small in this case study, the total saved travel time will become very large when

accounting the large number of vehicles and the duration of the work zone. Moreover,

if the traffic condition of the network gets even worse due to more incidents (e.g.

more work zones are placed), the total travel time saved by the optimal integrated corridor planning and operational strategy would become even larger.

Concerning the locally impacted vehicles, their performance has been improved much more significantly by the joint optimization. Their average trip travel time is reduced by 6.58% when the BSK predicted optimal solution is implemented, and the D-SVR predicted optimal solution creates a 7.66% decrease in average trip travel time for those locally impacted vehicles.
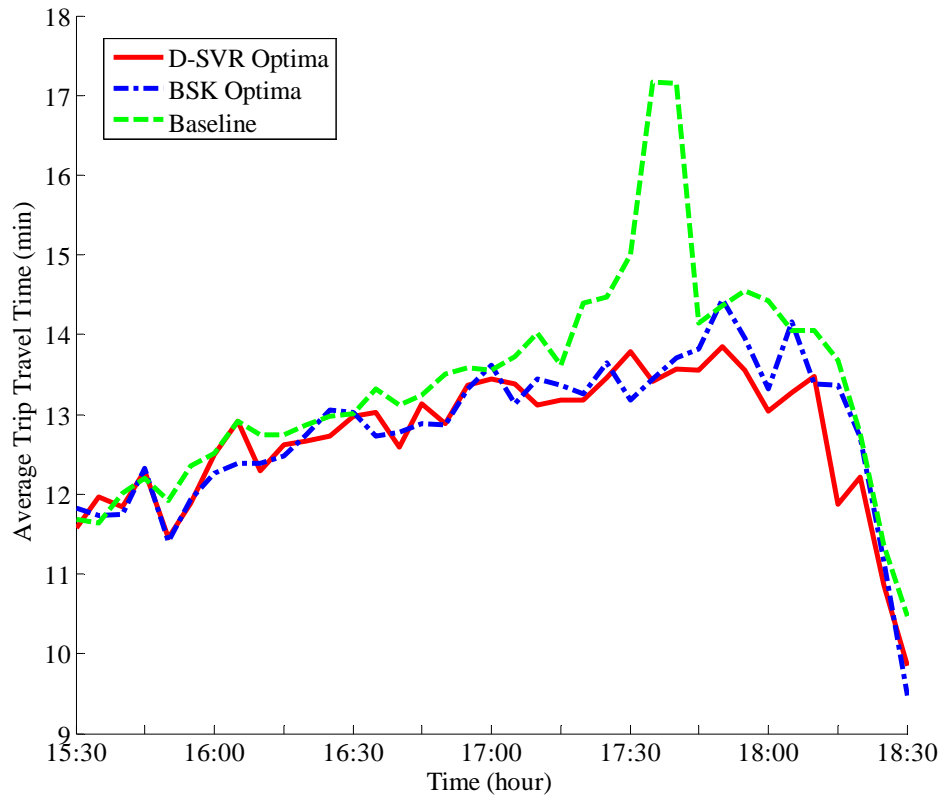

Figure 7-6: Network Wide Average Trip Travel Time along Hour of the Day of the Baseline, BSK Optima and D-SVR Optima Cases.

To present more details of the effect on system wide performance caused by the optimal solutions provided by the two surrogate models, the changes of average trip travel time along time of the day for the three scenarios are plotted in Figure 7-6.

During the most congested period (17:15-18:00), the optima predicted by both methods reduced the average trip travel time significantly by around 25%. After the optimal policy is implemented, the average trip travel time distributes more evenly within the afternoon peak period, and there is no severely congested period any more. During the period 17:45-18:15, the performance of the D-SVR optima is significantly better than that of the BSK optima.

7.5 Conclusions

Due to the observation of asymmetrically distributed output from the I-270 freeway-arterial corridor simulation, an enhanced support vector regression model that takes into account the distribution of simulation noise is developed in this chapter. The newly developed surrogate model is named distribution-based support vector regression (D-SVR). This model relaxes the assumption of normality in simulation noise in most surrogate models. The prediction loss is assumed to be dependent on the actual distribution of simulation noise.

Through a numerical test that sets the distribution of noise to be Gumbel Minimum, the developed D-SVR model is compared to the heteroscedastic SVR and the Bayesian stochastic Kriging in terms of the performance on prediction accuracy. The results show that when the signal-to-noise ratio is relatively high, the D-SVR model can predict the true response surface more accurately and always outperforms the heteroscedastic SVR and the BSK model.

In addition, bootstrapping is introduced and incorporated into the D-SVR surrogate-based optimization framework. Applying bootstrapping to provide prediction and predictor's variance at new points and adopting the maximization of the expected improvement as the global optimal infill strategy, we tested the performance of the bootstrapped D-SVR in searching for the global optimal solution with a numerical example. It is found that the bootstrapped D-SVR is more efficient in global optimization, and it can find a better solution much faster than the BSK model. This finding suggests that with a fixed computation budget, the D-SVR model can provide a better solution than the BSK model, which is of great value to real world applications.

The proposed EI infill bootstrapped D-SVR method is then utilized in the real world application of a joint optimization of HOT toll and freeway diversion control, which is introduced in Chapter 6. With 20 iterations of infill, the D-SVR predicted optima can reduce the network wide average trip travel time by 3.11% than the baseline, and the reduction for locally impacted vehicles is 7.66%. The comparison between the D-SVR method and the BSK method is also conducted. Results suggest the improvement made by the D-SVR optima is significantly higher than that made by the BSK optima.

Chapter 8: Conclusions

Although simulation has shown its value in informing decision makers about the influence of their proposed policies on network performance, it is often used for scenario evaluation instead of optimization, mainly due to its high computation cost. There is clearly a gap between the promising evaluation tool of simulation and the goal of policy optimization in the transportation research. This dissertation tries to fill this gap by integrating rigorous mathematical optimization techniques with simulation evaluation. Efforts have also been made to reduce the required computation time for the optimization process.

The remainder of this chapter summarizes the contributions of this research work, and proposes future research directions to further extend the developed methodologies.

8.1 Contributions

This research is one of the first to introduce simulation-based optimization methods into large-scale transportation network research. In addition, based on observations of specific characteristics related to transportation simulation, existing simulation-based optimization methods are improved to enhance their efficiency. Overall, this dissertation has contributed to both methodology developments from the academic perspective and a number of real world applications from the industry perspective.

First, based on review of existing simulation-based optimization methods, surrogate-based optimization method is selected as the focus of this study. The surrogate-based optimization method enjoys both the advantages of simulation in policy evaluation and the efficiency of mathematical optimization. It can be used to solve transportation problems of imperative needs for practitioners. Various types of optimization problems have been proposed and solved in this dissertation, which include single-objective optimization, multi-objective optimization and constrained optimization problems.

Second, the surrogate-based optimization method is capable of solving problems on the optimization of combinations of different types of policies, which is very difficult to be dealt with when simulation is not utilized. For instance, travel demand management policies (e.g. toll, managed lanes, etc.) mainly concern the macro-level travel behavior such as destination choice, departure time choice and route choice, while traffic operational strategies (e.g. freeway diversion control, traffic signal, etc.) mainly concern the micro-level traffic flow dynamics. Developing mathematical models that can capture the impact of both types of policies is very challenging. Existing analytical models evaluating the impact of one type of policies usually ignore the influence of measures from the other category. On the other hand, simulation models travel behavior at the individual level. The response of each individual to any policy can be captured, and thus the impact of combinations of policies to the system can be conveniently measured.

Third, the optimization methods developed in this research are general and can be applied to many other optimization problems. As the simulation mainly serves

as a black-box function in the framework, any simulator can be utilized coupled with the developed methods in application. Moreover, as long as the simulation is responsive to the particular policy, any policy can be optimized with the developed methods. There is no transferability issue associated with this method. Thus the proposed method can be used by transportation management agencies at all levels of government.

Fourth, existing surrogate-based optimization methods are significantly improved from the methodological point of view. Characteristics of transportation simulation noise have been investigated, and the theoretical foundation of the surrogate models is revised accordingly. Based on the observation of heteroscedasticity in simulation noise, two advanced surrogate models that adapts for the heteroscedastic noise are developed: the heteroscedastic support vector regression (H-SVR) model and the Bayesian stochastic Kriging (BSK) model. Furthermore, the transportation simulation noise is found to be asymmetrically distributed, violating the assumption of normality in noise in most surrogate models. A distribution-based support vector regression (D-SVR) model that takes into account the asymmetric distribution of noise is then developed. Overall, it is found that these enhanced models can improve the accuracy in approximating the true response surface. More importantly, they can also improve the converging speed of optimization, which leads to significant savings of computational cost in practice.

Fifth, two real world problems have been investigated using the developed methods. The first problem is to optimize the dynamic pricing of a toll road in the State of Maryland. This problem is formulated into three forms: with single-objective,

185

with multi-objective and with complicated constraints. The solutions of the problem in all of the three forms are investigated in this study. The second problem is to jointly optimize the HOT toll rate and the freeway diversion control for a freeway-arterial corridor in the State of Maryland under a work zone scenario. The solutions provided by the developed methods to both of the two problems are shown to significantly improve the network performance compared to the baseline. The cost of adjusting these policies is minor, while the benefit in terms of travel time savings is significant.

8.2 Recommendations for Future Research

This dissertation undertakes the research on integrating simulation evaluation with mathematical optimization, improving surrogate models for transportation simulation and solving real world transportation related problems. To better serve the needs of transportation management agencies, this work can be extended through several directions from the application point of view and the methodological point of view.

The methods developed in this dissertation mainly deal with continuous optimization problems. While lots of transportation problems involve discrete decision variables (e.g. discrete/mixed integer network design problems, optimal allocation of traffic detectors, etc.), it is valuable to extend the ability of surrogate-based optimization methods in solving discrete or mixed discrete-continuous problems.

The simulator used in all of the case studies is DynusT, which is essentially a dynamic traffic assignment model. Travel adjustment behavior along other dimensions (e.g. departure time, mode, destination, etc.) than the route choice in response to any particular policies is generally ignored. Integrating other modules that model the travel behavioral change due to the policies would definitely enhance the simulation's capability in predicting the traffic realism. More reliable optimization results can thus be provided when this type of fully integrated activity/agent-based models is used as the simulator for the surrogate-based optimization.

In theory, if the number of simulation evaluations becomes infinite, the global optimal solution can always be found. However, the theoretical proof of convergence rate of the developed methods needs further investigation. In practice, the selection of stopping rule for the optimization process is also important. The rule applied in existing studies usually specifies a total number of simulation replications.

Improving the method of generating the initial sample for the surrogate model estimation is also an interesting topic. A good initial sample that covers promising regions of the domain can greatly enhance the efficiency of the method in searching for the optimal solution. Furthermore, for the surrogate-based optimization methods with infill, the appropriate allocation of computational budget between the initial samples and infill samples also needs additional research.

The efficiency of an optimization method can be significantly improved when parallel computing techniques are applied. Regarding surrogate-based optimization methods, the multiple-replication simulations on both initial and infill sample points can be easily distributed to multiple computer resources. Developing global optimal

infill strategies that can take advantage of parallel computing is a topic that deserves further investigation.

Bibliography

Adeli, H. and Jiang, X. (2009). Intelligent Infrastructure: Neural Networks, Wavelets, and Chaos Theory for Intelligent Transportation Systems and Smart Structures, CRC Press, Taylor & Francis, Boca Raton, FL.

Adeli, H. and Karim, A. (2000). Fuzzy-Wavelet RBFNN Model for Freeway Incident Detection. *Journal of Transportation Engineering*, 126(6): 464–471.

Andradóttir, S. (1998). A Review of Simulation Optimization Techniques. *In Proceedings of the 30th Winter Simulation Conference*: 151-158.

Andradóttir, S. (2005). An Overview of Simulation Optimization via Random Search. *Handbooks in Operations Research and Management Science*, 13: 617-631.

Ankenman, B., Nelson, B. and Staum, J. (2010). Stochastic Kriging for Simulation Metamodeling. *Operations Research*, 58(2): 371–382.

Aziz, H. and Ukkusuri, S. (2012). Integration of Environmental Objectives in a System Optimal Dynamic Traffic Assignment Model. *Computer-Aided Civil and Infrastructure Engineering*, 27(7): 494–511.

Balakrishna, R., Antoniou, C., Ben-Akiva, M., Koutsopoulos, H. and Wen, Y. (2007a). Calibration of Microscopic Traffic Simulation Models: Methods and Application. *Transportation Research Record*, 1999: 198–207.

Balakrishna, R., Ben-Akiva, M. and Koutsopoulos, H. (2007b). Offline Calibration of Dynamic Traffic Assignment: Simultaneous Demand-And-Supply Estimation. *Transportation Research Record*, 2003: 50–58.

Balakrishna, R., Morgan, D., Slavin, H. and Yang, Q. (2009). Large-Scale Traffic Simulation Tools for Planning and Operations Management. *In Proceedings of the 12th IFAC Symposium on Control in Transportation Systems*: 117-122.

Banks, J., Carson, J., Nelson, B. and Nicol, D. (2000). Discrete Event Systems Simulation. 3rd Edition. Prentice Hall, Englewood Cliffs, NJ.

Baraldi, P., Canesi, R., Zio, E., Seraoui, R. and Chevalier, R. (2011). Genetic Algorithm-Based Wrapper Approach for Grouping Condition Monitoring Signals of Nuclear Power Plant Components. *Integrated Computer-Aided Engineering*, 18(3): 221–234.

Barceló, J. and Casas, J. (2005). Dynamic Network Simulation with AIMSUN. *Simulation Approaches in Transportation Analysis*: 57-98. Springer US.

Barton, R. and Meckesheimer, M. (2006). Metamodel-Based Simulation Optimization. *Handbooks in Operations Research and Management Science*, Amsterdam, Elsevier B.V.

Bellemans, T., De Schutter, B. and De Moor, B. (2006). Model Predictive Control for Ramp Metering of Motorway Traffic: A Case Study. *Control Engineering Practice*, 14(7): 757-767.

Ben-Akiva, M., Bierlaire, M., Bottom, J., Koutsopoulos, H. and Mishalani, R. (1997). Development of a Route Guidance Generation System for Real-time Application. *In Proceedings of 8th IFAC/IFIP/IFORS Symposium on Transportation Systems*: 433-439.

Ben-Akiva, M., Bierlaire, M., Koutsopoulos, H. and Mishalani, R. (2002). Real Time Simulation of Traffic Demand-Supply Interactions within DynaMIT. *Transportation and Network Analysis: Current Trends*: 19-36. Springer US.

Ben-Akiva, M. and Lerman, S. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand. The MIT Press.

Bettonvil, B. (1989). A Formal Description of Discrete Event Dynamic Systems Including Infinitesimal Perturbation Analysis. *European Journal of Operational Research*, 42(2): 213-222.

Björkman, M. and Holmström, K. (2000). Global Optimization of Costly Nonconvex Functions Using Radial Basis Functions. *Optimization and Engineering*, 1(4): 373–397.

Burris, M., Ungemah, D., Mahlawat, M. and Pannu, M. (2009). Investigating the Impact of Tolls on High-Occupancy-Vehicle Lanes Using Managed Lanes. *Transportation Research Record*, 2099: 113–122.

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

CATT Lab. (2013). The Center for Advanced Transportation Technology Laboratory. http://www.cattlab.umd.edu/

Chabuk, T., Reggia, J., Lohn, J. and Linden, D. (2012). Causally-Guided Evolutionary Optimization and Its Application to Antenna Array Design. *Integrated Computer-Aided Engineering*, 19(2): 111–124.

Chan, K. and Lam, W. (2002). Optimal Speed Detector Density for the Network with Travel Time Information. *Transportation Research Part A*, 36(3): 203-223.

Chen, C., Schonfeld, P. and Paracha, J. (2005). Work Zone Optimization for Two-Lane Highway Resurfacing Projects with an Alternate Route. *Transportation Research Record*, 1911: 51–66.

Chen, O., Hotz, A. and Ben-Akiva, M. (1997). Development and Evaluation of a Dynamic Ramp Metering Control Model. *In Proceedings of 8th IFAC/IFIP/IFORS Symposium on Transportation Systems*: 1162-1168.

Chen, X., Ankenman, B. and Nelson, B. (2012). The Effects of Common Random Numbers on Stochastic Kriging Metamodels. *ACM Transactions on Modeling and Computer Simulation*, 22 (2): 7.

Chen, X., Zhang, L., He, X., Xiong, C.and Li, Z. (2013). Surrogate-Based Optimization of Expensive-to-Evaluate Objective for Optimal Highway Toll Charges in Transportation Network. *Computer-Aided Civil and Infrastructure Engineering*, DOI: 10.1111/mice.12058.

Chiu, Y., Bottom, J., Mahut, M., Paz, A., Balakrishna, R., Walle, T. and Hicks, J. (2010a). A Primer for dynamic Traffic Assignment. 89th Annual Meeting of Transportation Research Board, Washington, D. C.

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

Chiu, Y., Zhou, L. and Song, H. (2010b). Development and Calibration of the Anisotropic Mesoscopic Simulation Model for Uninterrupted Flow Facilities. *Transportation Research Part B*, 44(1): 152–174.

Ciuffo, B., Casas, J., Montanino, M., Perarnau, J. and Punzo, V. (2013). From Theory to Practice: Gaussian Process Metamodels for the Sensitivity Analysis of Traffic Simulation Models. A Case Study of the Aimsum Mesoscopic Model. 92nd Annual Meeting of Transportation Research Board, Washington, D. C.

Ciuffo, B., Punzo, V. (2013). "No Free Lunch" Theorems Applied to the Calibration of Traffic Simulation Models. *IEEE Transactions on Intelligent Transportation Systems*. In Press.

Ciuffo, B., Punzo, V. and Quaglietta, E. (2011). Kriging Metamodelling to Verify Traffic Micro-Simulation Calibration Methods. 90th Annual Meeting of Transportation Research Board, Washington, D. C.

Dai, D. and Schonfeld, P. (1998). Metamodels for estimating Delays through Series of Waterway Queues. *Transportation Research Part B*, 32(1): 1–19.

De Brabanter, L., De Brabanter, J., Suykens, J. and De Moor, B. (2011). Spproximate Confidence and Prediction Intervals for Least Squares Support Vector Regression. *IEEE Transactions on Neural Networks*, 22(1): 110-120.

De Palma, A., Kilani, M. and Lindsey, R. (2005). Congestion Pricing on a Road Network: A Study Using the Dynamic Equilibrium Simulator METROPOLIS. *Transportation Research Part A*, 39(7): 588-611.

Den Hertog, D., Kleijnen, J. and Siem, A. (2006). The Correct Kriging Variance Estimated by Bootstrapping. *Journal of the Operational Research Society*, 57(4): 400-409.

Eaton, M. (1983). Multivariate Statistics: A Vector Space Approach. Eaton, M.L. (Ed.). New York: Wiley.

Efron, B. and Tibshirani, R. (1994). An Introduction to the Bootstrap. Vol. 57. CRC Press.

Eglese, R. (1990). Simulated Annealing: A Tool for Operational research. *European Journal of Operational Research*, 46(3): 271-281.

Federal Highway Administration and Federal Transit Administration (2007). The Transportation Planning Process: Key Issues. Retrieved February 20, 2014 from http://www.planning.dot.gov/documents/BriefingBook/BBook.htm#1BB

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

Feng, C. and Wu. H. (2006). Integrating fmGA and CYCLONE to Optimize the Schedule of Dispatching RMC Trucks. *Automation in Construction*, 15: 186-199.

Flötteröd, G., Bierlaire, M. and Nagel, K. (2011). Bayesian Demand Calibration for Dynamic Traffic Simulations, *Transportation Science*, 45(4): 541–561.

Forrester, A. and Keane, A. (2009). Recent Advances in Surrogate-Based Optimization. *Progress in Aerospace Sciences*, 45(1): 50–79.

Forrester, A., Keane, A., and Bressloff, N. (2006). Design and Analysis of "Noisy" Computer Experiments, *AIAA Journal*, 44(10): 2331–2339.

Forrester, A., Sobester, A. and Keane, A. (2008). Engineering Design via Surrogate Modelling: A Practical Guide, Wiley.

Friedmann, J. (1987). Planning in the Public Domain – From Knowledge to Action. Princeton: Princeton University Press.

Fu, M. (1994). Optimization via Simulation: A Review. *Annals of Operations Research*, 53: 199-247.

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

Fu, M. (2006). Gradient Estimation. Chapter 19 in *Handbooks in Operations Research and Management Science: Simulation*. Henderson, S. and Nelson, B. edition, Elsevier.

Fu, M. and Howell, W. (2003). Application of Perturbation Analysis to Traffic Light Signal Timing. *In Proceedings of the 42nd IEEE Conference on Decision and Control*: 4837-4840.

Fu, M., Glover, F. and April, J. (2005). Simulation Optimization: A Review, New Developments, and Applications. *In Proceedings of the 37th Winter Simulation Conference*: 83-95.

Fu, M. and Hu, J. (1994). Smoothed Perturbation Analysis Derivative Estimation for Markov Chains. *Operations Research Letters*, 15(5): 241-251.

Gao, L., Liu, M., Sun, Z. and Mao, B. (2008). Simulation on Impact of Information Guidance on Regional Traffic Flow. *Journal of Transportation Systems Engineering and Information Technology*, 8(4): 63-69.

Gardner, L., Bar-Gera, H. and Boyles, S. (2013). Development and Comparison of Choice Models and Tolling Schemes for High-Occupancy/Toll (HOT) Facilities. *Transportation Research Part B*, 55: 142–153.

Gartner, N., Pooran, F. and Andrews, C. (2001). Implementation of the OPAC Adaptive Control Strategy in a Traffic Signal Network. *Intelligent Transportation Systems, Proceedings of 2001 IEEE*: 195-200.

Gengembre, E., Ladevie, B., Fudym, O., Thuillier, A. (2012). A Kriging Constrained Efficient Global Optimization Approach Applied to Low-Energy Building Design Problems. *Inverse Problems in Science and Engineering*, 20(7): 1101-1114.

Gibbs, M. (1997). Bayesian Gaussian Processes for Regression and Classification, Doctoral dissertation, University of Cambridge.

Glasserman, P. (1991). Gradient Estimation via Perturbation Analysis. Kluwer Academic Publishers, Boston, Massachusetts.

Glover, F. and Laguna, M. (1997). Tabu Search. Boston: Kluwer Academic.

Glynn, P. (1987). Likelihood Ratio Gradient Estimation: An Overview. *In Proceedings of the 19th Winter Simulation Conference*: 366-375.

Glynn, P. (1989). Likelihood Ratio Derivative Estimators for Stochastic Systems. *In Proceedings of the 21st Winter Simulation Conference*: 374-380.

Goel, R. and Burris, M. (2012). HOT Lane Policies and Their Implications. *Transportation*, 39(6): 1019–1033.

Goldberg, D. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning, 412. Reading Menlo Park: Addison-wesley.

Goldsman, D., Nelson, B. and Schmeiser, B. (1991). Methods for Selecting the Best System. *In Proceedings of the 23rd Winter Simulation*: 177-186.

Gutmann, H. (2001). A Radial Basis Function Method for Global Optimization. *Journal of Global Optimization*, 19(3): 201–227.

Hachicha, W., Ammeri, A., Masmoudi, F. and Chachoub, H. (2010). A Comprehensive Literature Classification of Simulation Optimisation Methods. Retrieve March 2, 2014 from http://mpra.ub.uni-muenchen.de/27652/

Halati, A., Lieu, H. and Walker, S. (1997). CORSIM-Corridor Traffic Simulation Model. *Traffic Congestion and Traffic Safety in the 21st Century: Challenges, Innovations, and Opportunities*.

Hegyi, A., De Schutter, B. and Hellendoorn, H. (2005). Model Predictive Control for Optimal Coordination of Ramp Metering and Variable Speed Limits. *Transportation Research Part C*, 13(3): 185-209.

Heidergott, B. (2001). Option Pricing via Monte Carlo Simulation — A Weak Derivative Approach. *Probability in the Engineering and Informational Sciences*, 15(3): 335-349.

Highway Statistics. (1980). Office of Highway Policy Information. Federal Highway Administration.

Highway Statistics. (2008). Office of Highway Policy Information. Federal Highway Administration.

Ho, Y. and Deng, M. (1994). The Problem of Large Search Space in Stochastic Optimization. *In Proceedings of the IEEE Conference on Decision and Control*: 1470-1475.

Ho, Y., Sreenivas, R. and Vakili, P. (1992). Ordinal Optimization of DEDS. *Discrete Event Dynamic Systems: Theory and Applications*, 2(1): 61-88.

Horowitz, A., Weisser, I. and Notbohm, T. (2003). Diversion from a Rural Work Zone with Traffic-Responsive Variable Message Signage System. *Transportation Research Record*, 1824: 23–28.

Hsu, J. and Nelson, B. (1988). Optimization over a Finite Number of System Designs with One-Stage Sampling and Multiple Comparisons with the Best. *In Proceedings of the 20th Winter Simulation Conference*: 451-457.

Huang, D., Allen, T., Notz, W. and Zeng, N. (2006). Global Optimization of Stochastic Bblack-Box Systems via Sequential Kriging Meta-models. *Journal of Global Optimization*, 34(3): 441–466.

Hussain, M., Barton, R. & Joshi, S. (2002). Metamodeling: Radial Basis Functions, versus Polynomials. *European Journal of Operational Research*, 138(1): 142–154.

Improta, G. and Cantarella, G. (1984). Control Systems Design for an Individual Signalized Junction. *Transportation Research Part B*, 18: 147-167.

Jacob, C., Abdulhai, B., Hadayeghi, A. and Malone, B. (2006). Highway Work Zone Dynamic Traffic Control Using Machine Learning. IEEE Intelligent Transportation Systems Conference, ITSC 06, 262-272.

Jakobsson, S., Patriksson, M., Rudholm, J. and Wojciechowski, A. (2010). A Method for Simulation-based Optimization Using Radial Basis Functions. *Optimization and Engineering*, 11(4): 501-532.

Jones, D. (2001). A Taxonomy of Global Optimization Methods Based on Response Surfaces. *Journal of Global Optimization*, 21(4): 345–383.

Jones, D., Schonlau, M. and Welch, W. (1998). Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4): 455–492.

Kazemi, A. and Zarandi, M. (2008). An Agent-Based Framework for Building Decision Support System in Supply Chain Management. *Journal of Applied Sciences*, 8: 1125-1137.

Kessaci, A., Farges, J. and Henry, J. (1990). Upper Level for Real Time Urban Traffic Control Systems. *In Preprints 11th IFAC World Congress*, 10: 226-229.

Khattak, A., Schofer, J. and Koppelman, F. (1995). Effect of Traffic Information on Commuters' Propensity to Change Route and Departure Time. *Journal of Advanced Transportation*. 29(2): 193-212.

Kim, H. and Adeli, H. (2001). Discrete Cost Optimization of Composite Floors Using a Floating Point Genetic Algorithm. *Engineering Optimization*, 33(4): 485–501.

Kleijnen, J. (2009). Kriging Metamodeling in Simulation: A Review. *European Journal of Operational Research*, 192(3): 707–716.

Kleijnen, J. (2014). Simulation-Optimization via Kriging and Bootstrapping: a Survey. *Journal of Simulation*, in press, DOI: 10.1057/jos.2014.4..

Kleijnen, J., Van Beers, W. and Van Nieuwenhuyse, I. (2012). Expected Improvement in Efficient Global Optimization through Bootstrapped Kriging. *Journal of Global Optimization*, 54(1): 59-73.

Kochel, P., Kunze, S. and Nielander, U. (2003). Optimal Control of a Distributed Service System with Moving Resources: Application to the Fleet Sizing and Allocation Problem. *International Journal of Production Economics*, 81-82: 443-459.

Kotsialos, A., Papateorgiou, M., Mangeas, M. and Haj-Salem, H. (2002). Coordinated and Integrated Control of Motorway Networks via Non-linear Optimal Control. *Transportation Research Part C*, 10(1): 65-84.

Kullback, S. and Leibler, R. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22 (1): 79–86.

Laguna, M. and Marti, R. (2002). Scatter Search. Boston: Kluwer Academic Publishers.

Law, A. and Kelton, W. (1991). Simulation Modeling and Analysis, 2nd edtion. McGraw-Hill, New York.

LeBlanc, L. and Boyce, D. (1986). A Bi-Level Programming for Exact Solution of the Network Design Problem with User-Optimal Flow. *Transportation Research Part B*, 20(3): 259-265.

Lee, L., Lau, T. and Ho, Y. (1999). Explanation of Goal Softening in Ordinal Optimization. *IEEE Transactions on Automatic Control*, 44(1): 94-99.

Levinson, D., Liu, H., Garrison, W., Hickman, M., Danczyk, A. and Corbett, M. (2013). Fundamental of Transportation. *WIKI Book*.

Li, D., Xu, L., Goodman, E., Xu, Y. and Wu, Y. (2013). Integrating a Statistical Background-Foreground Extraction Algorithm and SVM Classifier for Pedestrian Detection and Tracking. *Integrated Computer-Aided Engineering*, 20(3): 201–216.

Li, A., Xu, N., Nozick, L. and Davidson, R. (2011). Bilevel Optimization for Integrated Shelter Location Analysis and Transportation Planning for Hurricane Events. *Journal of Infrastructure Systems*, 17(4): 184-192.

Li, Z., Madanu, S., Zhou, B., Wang, Y. and Abbas, M. (2010). A Heuristic Approach for Selecting Highway Investment Alternatives. *Computer-Aided Civil and Infrastructure Engineering*, 25(6): 427–439.

Lou, Y., Yin, Y. and Laval, J. (2011). Optimal Dynamic Pricing Strategies for High-Occupancy/Toll Lanes. *Transportation Research Part C*, 19(1): 64–74.

Mahmassani, H. (2002). Dynamic Network Traffic Assignment and Simulation Methodology for Advanced System Management Applications. 81st Annual Meeting of Transportation Research Board, Washington, D. C.

Melouk, S., Keskin, B., Armbrester, C. and Anderson, M. (2010). A Simulation Optimization-Based Decision Support Tool for Mitigating Traffic Congestion, *Journal of the Operational Research Society*, 62(11): 1971–1982.

Meng, Q. and Wang, X. (2008). Sensitivity Analysis of Logit-Based Stochastic User Equilibrium Network Flows with Entry–Exit Toll Schemes. *Computer-Aided Civil and Infrastructure Engineering*, 23(2): 138–156.

Mesbah, M., Sarvi, M., Ouveysi, I. and Currie, G. (2011). Optimization of Transit Priority in the Transportation Network Using a Decomposition Methodology. *Transportation Research Part C*, 19(2): 363-373.

Messmer, A. and Papageorgiou, M. (1995). Route Diversion Control in Motorway networks via Nonlinear Optimization. *IEEE Transactions on Control Systems Technology*, 3(1): 144-154.

Montgomery, D. (2008). Design and Analysis of Experiments, Wiley.

Morris, M. and Mitchell, T. (1995). Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, 43(3): 381–402.

Mosseri, G., Hall, M. and Meyers, J. (2004). VISSIM Micro-simulation Modeling of Complex Geometry and Traffic Control: A Case Study of Ocean Parkway, NY. In Institute of Transportation Engineers Annual Meeting.

Murray, P., Mahmassani, H. and Abdelghany, K. (2001). Methodology for Assessing High-Occupancy Toll-lane Usage and Network performance. *Transportation Research Record*, 1765(1): 8-15.

Nava, E. and Chiu, Y. (2012). A Temporal Domain Decomposition Algorithmic Scheme for Large-Scale Dynamic Traffic Assignment, *International Journal of Transportation Science and Technology*, 1(1): 1–24.

Omrani, R. and Kattan, L. (2012). Demand and Supply Calibration of Dynamic Traffic Assignment Models. *Transportation Research Record*, 2283: 100–112.

Osorio, C. (2010). Mitigating Network Congestion: Analytical Models, Optimization Methods and Their Applications, Lausanne, Ecole Polytechnique Fédérale de Lausanne, PhD Dissertation.

Papageorgiou, M. (1980). A New Approach to Time-of-Day Control Based on a Dynamic Freeway Traffic Model. *Transportation Research Part B*, 14(4): 349-360.

Paz, A. and Peeta, S. (2009). On-line Calibration of Behavior Parameters for Behavior-Consistent Route Guidance. *Transportation Research Part B*, 43(4): 403–421.

Queipo, N., Haftka, R., Shyy, W., Goel, T., Vaidyanathan, R. and Tucker, P. (2005). Surrogate-Based Analysis and Optimization. *Progress in Aerospace Sciences*, 41(1): 1–28.

Ramadurai, G. and Ukkusuri, S. (2011). B–Dynamic: An Efficient Algorithm for Dynamic User Equilibrium Assignment in Activity-Travel Networks. *Computer-Aided Civil and Infrastructure Engineering*, 26(4): 254–269.

Ramanathan, V. and Schonfeld, P. (1994). Approximate Delays Caused by Lock Service Interruptions. *Transportation Research Record*, 1430: 41-49.

Rasmussen, C. and Williams, C. (2006). Gaussian Processes for Machine Learning, MIT Press.

Regis, R. and Shoemaker, C. (2005). Constrained Global Optimization of Expensive Black Box Functions Using Radial Basis Functions. *Journal of Global Optimization*, 31(1): 153–171.

Retherford, J. and McDonald, M. (2011). Estimation and Validation of Gaussian Process Surrogate Models for MEPDG-Based Sensitivity Analysis and Design Optimization. 90th Annual Meeting of Transportation Research Board, Washington, D. C.

Robinson, S. (1996). Analysis of Sample-Path Optimization. *Mathematics of Operations Research*, 21(3): 513-528.

Rubinstein, R. and Shapiro, A. (1993). Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method. 346, New York, Wiley.

Sarma, K. and Adeli, H. (2001). Bi-Level Parallel Genetic Algorithms for Optimization of Large Steel Structures. *Computer-Aided Civil and Infrastructure Engineering*, 16(5): 295–304.

Schmitt, L. (2001). Theory of Genetic Algorithms. *Theoretical Computer Science*, 259(1): 1-61.

Schölkopf, B. and Smola, A. (2002). Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond. Cambridge, MIT Press.

Schrank, D., Eisele, B. and Lomax, T. (2012). TTI's 2012 Urban Mobility Report. Texas A&M Transportation Institute.

Schwartz, J., Wang, W. and Rivera, D. (2006). Simulation-Based Optimization of Process Control Policies for Inventory Management in Supply Chains. *Automatica*, 42(8): 1311-1320.

Simon, G., Volgyesi, P., Maróti, M. and Lédeczi, Á. (2003). Simulation-Based Optimization of Communication Protocols for Large-Scale Wireless Sensor Networks. *In proceedings of the IEEE Aerospace Conference.*

Simpson, T., Mauery, T., Korte, J. and Mistree, F. (2001). Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization. *AIAA Journal*, 39(12): 2233-2241.

Smola, A. and Schölkopf, B. (2004). A Tutorial on Support Vector Regression. *Statistics and Computing*, 14(3): 199–222.

Song, M., Yin, M., Chen, X., Zhang, L. and Li, M. (2013), A Simulation-Based Approach for Sustainable Transportation Systems Evaluation and Optimization: Theory, Systematic Framework and Applications, *In Proceedings of the 13th COTA International Conference of Transportation Professionals*, Shenzhen, China.

Spall, J. (1992). Multivariate stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation. *IEEE Transactions on Automatic Control*, 37(3): 332-341.

Stamatiadis, C. and Gartner, N. (1996). MULTIBAND-96: A Program for Variable Bandwidth Progression Optimization of Multiarterial Traffic Networks. *Transportation Research Record*, 1554: 9-17.

Stevanovic, J., Stevanovic, A., Martin, P. and Bauer, T. (2008). Stochastic Optimization of Traffic Control and Transit Priority Settings in VISSIM. *Transportation Research Part C*, 16(3): 332-349.

Sumalee, A. (2004). Optimal Road User Charging Cordon Design: A Heuristic Optimization Approach, *Computer-Aided Civil and Infrastructure Engineering*, 19(5): 377–392.

Sundaram, S., Koutsopoulos, H., Ben-Akiva, M., Antoniou, C. and Balakrishna, R. (2011). Simulation-Based Dynamic Traffic Assignment for Short-Term Planning Applications. *Simulation Modelling Practice and Theory*, 19(1): 450–462.

Swisher, J., Hyden, P., Jacobson, S. and Schruben, L. (2004). A Survey of Recent Advances in Discrete Input Parameter Discrete-Event Simulation Optimization. *IIE Transactions*, 36(6): 591-600.

Szeto, W., Solayappan, M. and Jiang, Y. (2011). Reliability–Based Transit Assignment for Congested Stochastic tTransit Networks. *Computer-Aided Civil and Infrastructure Engineering*, 26(4): 311–326.

Tao, X. and Schonfeld, P. (2005). Lagrangian Relaxation heuristic for Selecting Interdependent Transportation Projects under Cost Uncertainty. *Transportation research record*, 1931: 74-80.

Tao, X. and Schonfeld, P. (2006). Selection and Scheduling of Interdependent Transportation Projects with Island Models. *Transportation research record*, 1981: 133-141.

Tao, X. and Schonfeld, P. (2007). Island Models for a Stochastic Problem of Transportation Project Selection and Scheduling. *Transportation research record*, 2039: 16-23.

Tekin, E. and Sabuncuoglu, I. (2004). Simulation Optimization: A Comprehensive Review on Theory and Applications. *IIE Transactions*, 36(11): 1067-1081.

Teklu, F., Sumalee, A. and Watling, D. (2007). A Genetic Algorithm Approach for Optimizing Traffic Control Signals Considering Routing, *Computer-Aided Civil and Infrastructure Engineering*, 22(1): 31–43.

Tian, X., Mahut, M., Jha, M. and Florian, M. (2007). Dynameq Application to Evaluating the Impact of Freeway Reconstruction. 86th Annual Meeting of Transportation Research Board, Washington, D. C.

Ting, C. and Schonfeld, P. (1998). Optimization through Simulation of Waterway Transportation Investments. *Transportation Research Record*, 1620: 11–16.

Torday, A. and Dumont, A. (2003). Floating Car Data Acquisition and Dynamic Route Guidance Strategies Evaluation with AIMSUN

Tresidder, T., Zhang, Y. and Forrester, A. (2012). Acceleration of Building Design Optimization through the Use of Kriging Surrogate Models. *In Proceedings of the Building Simulation and Optimization Conference*: 1-8.

Unnikrishnan, A. and Lin, D.(2012). User Equilibrium with Recourse: Continuous Network Design Problem. *Computer-Aided Civil and Infrastructure Engineering*, 27(7): 512–524.

Vaze, V., Antoniou, C., Wen, Y. and Ben-Akiva, M. (2009). Calibration of Dynamic Traffic Assignment Models With Point-To-Point Traffic Surveillance. *Transportation Research Record*, 2090: 1–9.

Villemonteix, J., Vazquez, E. and Walter, E. (2009). An Informational Approach to the Global Optimization of Expensive-to-Evaluate Functions. *Journal of Global Optimization*, 44(4): 509–534.

Vincent, R. and Young, C. (1986). Self-Optimizing Traffic Signal Control Using Microprocessor: The TRRL "MOVA" Strategy for Isolated Intersections. *Traffic Engineering and Control*, 27: 385-387.

Wandekokem, E., Mendel, E., Fabris, F., Valentim, M., Batista, R., Varejao, F. and Rauber, T. (2011). Diagnosing Multiple Faults in Oil Rig Motor Pumps Using Support Vector Machine Classifier Ensembles. *Integrated Computer-Aided Engineering*, 18(1): 61–74.

Wang, S. and Schonfeld, P. (2005). Scheduling Interdependent Waterway Projects through Simulation and Genetic Optimization. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 131(3): 89-97.

Wang, S. and Schonfeld, P. (2008). Scheduling of Waterway Projects with Complex Interrelations. *Transportation research record*, 2062: 59-65.

Wang, S. and Schonfeld, P. (2012). Simulation-based Scheduling of Mutually Exclusive Projects with Precedence and Regional Budget Constraints. *Transportation research record*, 2273: 1-9.

Wang, S., Yang, N. and Schonfeld, P. (2009). Simulation-based Network Maintenance Planning and Scheduling. *Transportation research record*, 2100: 94-102.

Wang, Y. and Prevedouros, P. (1996). Synopsis of Traffic Simulation Models. The 75th Annual Meeting of Transportation Research Board, Washington, D.C.

Wolpert, D. and MacReady, W. (1997). No Free Lunch Theorems for Opimization. *IEEE Transaction on Evolutionary Computation*, 1: 67-82.

Woodbury, M. (1950). Inverting Modified Matrices, Memorandum Rept. 42, Statistical Research Group, Princeton University, Princeton, NJ.

Xiong, C. and Zhang, L. (2013). A Descriptive Bayesian Approach to Modeling and Calibrating Drivers' En Route Diversion Behavior. *IEEE Transactions on Intelligent Transportation Systems*, 14(4): 1817–1824.

Xiong, C., Zhu, Z., He, X., Chen, X., Zhu, S. and Zhang, L. (2014). A 24-hour Large-Scale Microscopic Simulation Case Study of Inter-County Connector in Maryland. The 93rd Annual Meeting of Transportation Research Board, Washington, D.C.

Yang, H. (1999). System Optimum, Stochastic User Equilibrium and Optimal Link Tolls. *Transportation Science*, 33(4): 354-360.

Yang, H. and Huang, H. (1998). Principle of Marginal-Cost Pricing: How Does It Work in a General Network? *Transportation Research Part A*, 32: 45-54.

Yang, H. and Huang, H. (2004). The Multiclass, Multicriteria Traffic Network Equilibrium and System Optimum Problem. *Transportation Research Part B*, 38(1):1–15.

Yang, H. and Huang, H. (2005). Mathematical and Economic Theory of Road Pricing. Elsevier, Oxford.

Yang, N. (2010). Optimization of Highway Work Zone Decisions Considering Short-term and Long-term Impacts. Doctoral dissertation, University of Maryland.

Yang, Y., Lu, H., Yin, Y. and Yang, H. (2013). Optimization of Variable Speed Limits for Efficient, Safe, and Sustainable Mobility. *Transportation Research Record*, 2333(1): 37-45.

Yim, K., Wong, S., Chen, A., Wong, C. and Lam, W. (2011). A Reliability-based Land Use and Transportation Optimization Model. *Transportation Research Part C*, 19(2): 351-362.

Yin, Y. and Lou, Y. (2009). Dynamic Tolling Strategies for Managed Lanes. *Journal of Transportation Engineering*, 135(2): 45–52.

Yoo, T., Cho, H. and Yücesan, E. (2010). Hybrid Algorithm for Discrete Event Simulation-based Supply Chain Optimization. *Expert Systems with Applications*, 37: 2354-2361.

Zhang, L., Chang, G.L., Zhu, S., Xiong, C., Du, L., Mollanejad, M. Hopper, N., and Mahapatra, S. (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.

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

Zhou, L., Yan, G. and Ou, J. (2013). Response Surface Method Based on Radial Basis Functions for Modeling Large-Scale Structures in Model Updating. *Computer-Aided Civil and Infrastructure Engineering*, 28(3): 210–226.

Zhu, L, Schonfeld, P., Kim, Y., Flood, I. and Ting, C. (1999). Queuing Network Analysis for Waterways with Artificial Neural Networks. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 13(05): 365-375.