

## ABSTRACT

Title of Document: INTEGRATED TRAFFIC STATE UNCERTAINTY  
MODELING, PROTRABLE TRAFFIC SENSOR  
NETWORK PLANNING, AND MANAGEMENT

XUECHI ZHANG, DOCTOR OF PHILOSOPHY, 2018

Directed by: ALI HAGHANI, PROFESSOR  
  
DEPARTMENT OF CIVIL AND ENVIRONMENTAL  
ENGINEERING

Over the past few years, traffic congestion has become a genuine nightmare to most of the urban commuters. Providing real-time traffic information is of key significance to Intelligent Transportation System (ITS). Accurate travel time or traffic speed information through Advanced Traveler Information System (ATIS) can provide guidance for travelers who make decisions every day on travel mode, route choice and departure time. Meanwhile, travelers' anxiety can be reduced with a better understanding of their current and future travel time. With a well-organized and reliable traffic surveillance network, ITS can not only assist travelers in understanding their travel time and planning their trips via ATIS but also detect traffic incident and dispatch a patrol team in a timely manner. Therefore, comprehensive and reliable traffic network surveillance is of fundamental significance in building a smart transportation network. This dissertation deals with three major issues about the highway system.

Traffic state such as travel time or traffic speed serves as a key parameter to reflect the highway system operation efficiency. Understanding the real-time traffic information is

useful in helping travelers make smart route choice and schedule proper departure times. Lots of efforts have been made to improve traffic state prediction performance with advanced real-time prediction models. But there is limited work studying the intrinsic prediction uncertainty of such data-driven based predictions. This dissertation developed an entropy-based uncertainty estimation model to evaluate system state predictability under any given measurement space from a stochastic evolution perspective. Then we considered the highway network as a stochastic system and applied the proposed model to evaluate travel time prediction uncertainty under both temporal and spatial measurement spaces. Moreover, the quantitative relationships between data-driven based prediction errors and the proposed uncertainty measurements are analyzed based on a real-world case study.

Second, we developed a sensor network optimization model aiming to provide network-level real-time traffic information surveillance. The proposed model has two advantages compared with traditional traffic sensor planning models. Conventionally, people only focus on the surveillance benefit at the location where sensors are placed while ignoring the surveillance benefit improvement inferred from the spatial traffic state correlations. Moreover, the proposed network optimization model provides one with the flexibility to come up with optimal sensor relocation strategies. Specifically, when traffic demand and travel time uncertainty are heterogeneously distributed in a highway network for a given time period, appropriately relocating sensors can fully make use of the surveillance resources and enhance the network surveillance. The proposed model was applied to plan a travel time surveillance network for Washington D.C.-Baltimore commute network. Optimal sensor placement strategies and relocation operations with respect to the surveillance benefits are analyzed and discussed for the study area.

Last, we consider the sensor placement problem from a different perspective given the a priori information is completely missing. For a highway network with complete

unknown historical traffic data and unknown GPS coverage, the question that how operators should plan a sensor network to evaluate these a priori traffic information is answered. Specifically, a multistage stochastic optimization model with endogenous uncertainty is presented, and a Monte Carlo simulation-based approach is designed to evaluate the optimal solution. The proposed optimization model was applied to the same Washington D.C.-Baltimore commute network and serves as a supplement to the real-time surveillance based dynamic sensor network model.

INTEGRATED TRAFFIC STATE UNCERTAINTY MODELING, PORTABLE  
TRAFFIC SENSOR NETWORK PLANNING, AND MANAGEMENT

By

Xuechi Zhang

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

Doctor of Philosophy

2018

Advisory Committee:

Professor Ali Haghani, Chair

Professor Paul Schonfeld

Associate Professor Cinzia Cirillo

Assistant Professor Barton Forman

Professor Michael C. Fu

© Copyright by

Xuechi Zhang

2018

## **Dedication**

To my beloved mother and father.

## **Acknowledgements**

First of all, I would like to express my sincere gratitude to my advisor Prof. Haghani. I would never have finished my graduate career and this dissertation without his guidance and support. Dr. Haghani not only provided continuous financial support and excellent research guidance to me but is also consistently advising me how to do things from various aspects of the life. I really appreciate his help and support during these years. I will firmly remember his guidance whenever and wherever I am in the future.

I would like to sincerely thank all my committee members, Prof. Paul Schonfeld, Prof. Cinzia Cirillo, Prof. Barton Forman and Prof. Michael Fu. I would not be able to improve and finalize this dissertation without your valuable suggestions and comments since the date of my dissertation proposal. Most importantly, the academic skills I have learned from all of your classes during the past six years will benefit me forever in my future professional career.

Moreover, special gratitude to Prof. Saini Yang from Beijing Normal University. It was you that inspired my interests in the field of transportation engineering and recommended me to pursue a Doctoral study. Your consistent advice and encouragement helped me overcome many tough obstacles during my graduate study. I also want to thank all my colleagues at the University of Maryland. Because of you guys, my graduate life turned to interesting and fruitful.

Last but not least, I want to express my deepest gratitude to my parents and sister for their continuous support and warmest encouragement in past years. Your selfless support is the biggest motivation for me to finish this dissertation and obtain my Ph.D. degree.

Xuechi Zhang

April 13, 2018

# Table of Contents

Table of Contents .....	iv
List of Tables .....	vii
List of Figures .....	viii
Chapter 1 Introduction.....	1
1.1 Background and Motivation .....	1
1.2 Problem Statement and Objective.....	3
1.3 Research Contributions .....	5
1.4 Dissertation Organization .....	6
Chapter 2 Literature Review.....	8
2.1 Traffic Sensor Introduction.....	8
2.1.1 Fixed Point Sensor .....	9
2.1.2 Fixed Paired Sensor .....	11
2.1.3 Probe Sensor .....	13
2.1.4 Portable Point Sensor.....	15
2.1.5 Portable Paired Sensor .....	17
2.1.6 Summary.....	20
2.2 Traffic Sensor Location Optimization .....	23
2.2.1 Traffic Flow Measurement and Estimation .....	23
2.2.2 Traffic Time Collection and Estimation .....	27
2.2.2.1 Point Sensor Based SLP.....	28
2.2.2.2 Paired Sensor Based SLP.....	31
2.2.2.3 Probe Sensor Based SLP.....	34
2.2.3 Summary.....	36
Chapter 3 Real-Time Traffic Surveillance Efficiency Modeling: A Stochastic System Perspective 40	
3.1 Variance and Covariance Based Approach.....	40
3.2 Travel Time Prediction Accuracy Improvement .....	44
3.2.1 Travel Time Prediction Necessity.....	44
3.2.2 Travel Time Prediction Error without Real-time Surveillance.....	45
3.2.3 Travel Time Prediction Error with Real-time Surveillance .....	47
3.2.4 Surveillance Benefit Based on Travel Time Prediction Error Reduction.....	51
3.3 Entropy-Based Model to Quantify Traffic Condition Prediction Uncertainty ..53	
3.3.1 Background and Introduction .....	53
3.3.2 System State Uncertainty.....	56
3.3.3 Temporal System State Prediction Uncertainty .....	58
3.3.4 Spatial System State Prediction Uncertainty .....	61
3.4 Application to Real-time Travel Time Surveillance and Prediction.....	63
3.4.1 Introduction.....	63
3.4.2 State Space Definition.....	63
3.4.3 Numerical Experiments .....	65

3.4.3.1	Preliminaries and Objective Description .....	65
3.4.3.2	Travel Time Prediction Error Evaluation .....	67
3.4.3.3	Travel Time Prediction Uncertainty Evaluation .....	76
3.4.3.4	Empirical Analysis of Prediction Error Reduction and Uncertainty Reduction .....	84
3.5	Summary and Conclusions .....	90
Chapter 4	Traffic Sensor Network Planning Based on A Prior Information.....	92
4.1	Static Planning Model for Sensor Location Optimization.....	92
4.1.1	Introduction.....	92
4.1.2	Preliminaries .....	92
4.1.3	Mathematical Formulation.....	94
4.1.3.1	Decision Variables .....	94
4.1.3.2	Model Objective.....	95
4.1.3.3	Model Constraints.....	98
4.1.4	Model and Solution Method Discussion.....	99
4.2	Dynamic Planning Model for Sensor Location-Relocation Optimization.....	100
4.2.1	Background and Introduction .....	100
4.2.2	Preliminaries .....	101
4.2.2.1	A Time-Space Network Representation .....	102
4.2.2.2	Notations.....	103
4.2.3	Mathematical Formulation.....	105
4.2.3.1	Model Objective.....	105
4.2.3.2	Surveillance Coverage Constraints .....	106
4.2.3.3	Facility Routing Feasibility Constraints .....	107
4.2.3.4	Budget Constraints.....	109
4.2.3.5	Decision Variables .....	111
4.2.4	Model Complexity Analysis .....	112
4.2.5	Valid Integer Cuts.....	112
4.3	Applications to DC-Baltimore Commuting Network.....	114
4.3.1	Real-world Case Study Preliminaries .....	114
4.3.2	Travel Time Data Description and Statistical Analysis.....	115
4.3.3	Traffic Volume Data Extraction and Imputation .....	119
4.3.4	Time-space Network Specification.....	123
4.3.5	Numerical Experiments and Solution Time Investigation.....	124
4.3.6	Monetary Savings Estimation and Practical Policy Implications.....	134
4.3.7	Applications with the Existence of a Second Data Source .....	142
4.4	Summary and Conclusions .....	149
Chapter 5	Data Validation Oriented Sensor Network Planning: A Multistage Stochastic Optimization Approach.....	152
5.1	Introduction and Problem Statement .....	152
5.2	Preliminaries .....	155
5.2.1	Notations.....	155
5.2.2	Time-space Network Representation.....	158
5.3	Generalized Mathematical Formulation .....	160
5.4	Solution Approach .....	165

5.4.1	Scenario Decomposition-Based Reformulation.....	165
5.4.2	Non-Anticipativity Constraints Generation .....	168
5.4.3	Monte Carlo Scenario Aggregation Based Estimation .....	172
5.4.3.1	Sampling-based Solution Evaluation.....	172
5.4.3.2	Probabilistic Scenario Generation.....	176
5.5	Case Study and Numerical Experiments.....	178
5.5.1	Preliminaries and Experimental Design.....	178
5.5.2	Results Summary and Analysis .....	180
5.6	Summary and Conclusions .....	203
Chapter 6	Summary and Future Research .....	206
6.1	Contributions.....	206
6.2	Key Findings .....	207
6.3	Practical Concerns and Application.....	209
6.4	Future Research .....	211
References	213	

## List of Tables

Table 1: Comparison of Various Typical Traffic Sensing Technologies .....	21
Table 2: Summary of Selected Milestone Studies.....	39
Table 3: Travel Time Magnitude Classification.....	77
Table 4: Travel Time Trend Classification.....	77
Table 5: Travel Time Reliability Classification. ....	77
Table 6: Notations for the Dynamic Sensor Location Model.....	103
Table 7: Weekday Average Traffic Volume Distribution Probabilities at Different Time of Day under Different Roadway Types (Schrank et al. 2015). ....	122
Table 8: Weekend Average Traffic Volume Distribution Probabilities at Different Time of Day under Different Roadway Types (Schrank et al. 2015). ....	122
Table 9: Summary of Computational Results under Different Model Configurations.....	128
Table 10: Empirical Relationship between Relative Travel Time Prediction Error Reduction and Stochastic System Uncertainty by Introducing Real-time Measurements. ....	136
Table 11: Monetary Saving based Optimization Results Summary.....	137
Table 12: Notation Preparation for the Type-II Sensor Network Planning Model. ....	157
Table 13:Notations for Non-anticipativity Constraints Generation.....	170
Table 14: Monte Carlo Simulation Based MR-SSA Algorithm.....	176
Table 15: Random Scenario Generation Algorithm.....	177
Table 16: Solutions Comparisons of Problems under Different Model Parameters Configurations. .....	198

## List of Figures

Figure 1: Two Types of Fixed Point Traffic Sensors (a): In-roadway Sensor (b): Roadside Sensor .....	10
Figure 2: Two Examples of Fixed Paired (AVI) Traffic Sensors (a): RFID Sensing Based Toll Station (b): Camera Based License Plate Recognition Sensors.....	13
Figure 3: Illustration of Two Portable Traffic Detection Systems (a): Bluetooth Detection (b): WiFi Detection. ....	18
Figure 4: Illustration of Segment Travel Time Measurement by AVI Sensors.....	45
Figure 5: Mean Travel Time Estimation with Historical Observations Based on t-Student Distribution.....	47
Figure 6: Illustrative Examples of Travel Time Prediction Error Reduction: (a) MAPE Based Error Reduction; (b) RMSE Based Error Reduction .....	53
Figure 7: Entropy of A Single Event with respect to the Occurrence Probability.....	57
Figure 8: Case Study Network: (a) Map View of the Target Network (i.e., Commuting Network between Washington D.C. and Baltimore); (b) Extracted and Abstracted Network. ....	66
Figure 9: Spatial Distributions of Travel Time Coefficient of Variations in the Study Network: (a) Weekday Morning Peak Period; (b) Weekday Afternoon Peak Period; (c) Weekend Morning Peak Period; (d) Weekend Afternoon Peak Period. ....	68
Figure 10: Real-time Information Based Travel Time Prediction MAE Reduction Distributions Across Different Locations at Various Time Periods.....	70
Figure 11: Real-time Information Based Travel Time Prediction RMSE Reduction Distributions Across Different Locations at Various Time Periods.....	71

Figure 12: Real-time Information Based Travel Time Prediction MAPE Reduction Distributions Across Different Locations at Various Time Periods.....	72
Figure 13: Empirical Relationship Between Travel Time Prediction RMSE and MAPE: (a) Real- time Temporal Information Based Prediction, (b) Real-time Spatial Information Based Prediction. .....	74
Figure 14: Empirical Relationship Between Travel Time Prediction MAE and MAPE: (a) Real- time Temporal Information Based Prediction, (b) Real-time Spatial Information Based Prediction. .....	75
Figure 15: Empirical Relationship Between Travel Time Prediction MAE and RMSE: (a) Real- time Temporal Information Based Prediction, (b) Real-time Spatial Information Based Prediction .....	75
Figure 16: Travel Time Prediction Uncertainty Comparison Without/With Real-time Temporal Information for Segment 0706 at Different Periods.....	79
Figure 17: Travel Time Prediction Uncertainty Comparison Without/With Real-time Temporal Information for Segment 0605 at Different Periods.....	79
Figure 18: Travel Time Prediction Uncertainty Comparison Without/With Real-time Immediate Downstream Segment (i.e., 0605) Information for Segment 0706 at Different Periods. ....	80
Figure 19: Travel Time Prediction Uncertainty Comparison Without/With Real-time Second Immediate Downstream Segment (i.e., 0504) Information for Segment 0706 at Different Periods. .....	81
Figure 20: Travel Time Prediction Uncertainty Comparison Without/With Real-time Immediate Upstream Segment (i.e., 0807) Information for Segment 0706 at Different Periods. ....	81
Figure 21: Travel Time Prediction Uncertainty Comparison Without/With Real-time Second Immediate Upstream Segment (i.e., 0908) Information for Segment 0706 at Different Periods..	82

Figure 22: Historical Inference based Travel Time Prediction Uncertainty V.S. Real-time Temporal Information based Travel Time Prediction Uncertainty. ....	83
Figure 23: Historical Inference based Travel Time Prediction Uncertainty V.S. Real-time Spatial Information based Travel Time Prediction Uncertainty. ....	84
Figure 24: Travel Time MAE Reductions by RF Predictions V.S. Travel Time Uncertainty Reductions Estimated by the Proposed Model: (a) Conditioning on Real-time Temporal Information; (b) Conditioning on Real-time Spatial Information. ....	87
Figure 25: Travel Time RMSE Reductions by RF Predictions V.S. Travel Time Uncertainty Reductions Estimated by the Proposed Model: (a) Conditioning on Real-time Temporal Information; (b) Conditioning on Real-time Spatial Information. ....	88
Figure 26: Travel Time MAPE Reductions by RF Predictions V.S. Travel Time Uncertainty Reductions Estimated by the Proposed Model: (a) Conditioning on Real-time Temporal Information; (b) Conditioning on Real-time Spatial Information. ....	89
Figure 27: Segment with Direct Travel Time Measurement System .....	95
Figure 28: Two Typical Layout Configurations for Indirect Traffic State Estimation .....	97
Figure 29: A Time-Space based Traffic Monitoring Network with Finite Sensors .....	103
Figure 30: Illustration of Valid Cuts Regarding Relocation Operations. ....	113
Figure 31: Case Study Network: (a) Map View of the Target Network (i.e., Commuting Network between Washington D.C. and Baltimore); (b) Extracted and Abstracted Network. ....	115
Figure 32: Spatial Distributions of Travel Time Coefficient of Variations in the Study Network: (a) Weekday Morning Peak Period; (b) Weekday Afternoon Peak Period; (c) Weekend Morning Peak Period; (d) Weekend Afternoon Peak Period. ....	117
Figure 33: Time-dependent Spatial Distributions of Travel Time Coefficient of Variations in Weekday: (a) 8:00-9:00 AM; (b) 9:00-10:00 AM; (c) 10:00-11:00 AM; (d) 11:00-12:00 AM..	118

Figure 34: 2016 Annual Average Daily Traffic Counts Distribution in the Case Study Network: (a) Annual Average Weekday Daily Traffic (AAWDT); (b) Annual Average Weekend Daily Traffic (AAWNDT). .....	120
Figure 35: Time-dependent Weekday Average Traffic Volume Distribution w.r.t. Different Roadway Types (Schrank et al. 2015).....	121
Figure 36: Time-dependent Weekend Average Traffic Volume Distribution w.r.t. Different Roadway Types (Schrank et al. 2015).....	121
Figure 37: Time-space Network Specification of the Case Study.....	124
Figure 38: Solution Time Comparison of Each Optimization Case.....	129
Figure 39: Overall Solution Time Comparison between the Original Formulation and the Cut based Formulation. ....	129
Figure 40: Objective Value (Total Prediction Uncertainty Reduction) Comparison for Different Cases.....	131
Figure 41: Effects of Relocation Operations on Surveillance Benefit Improvement: (a) 4-sensor Case, (b) 8-sensor Case, (c) 12-sensor Case, (d) 16-sensor Case, (e) 20-sensor Case, (f) 24-sensor Case, and (g) 28-sensor Case.....	134
Figure 42: Combinatory Impacts of Sensor Budget and Relocation Limits on Total Network Travel Time Error Savings. ....	140
Figure 43: Impacts of Relocation Limits on Total Travel Time Error Savings for Different Fleet- Size Sensor Network: (a) 4-Sensor Network; (b) 8-Sensor Network; (c) 12-Sensor Network; (d) 16-Sensor Network; (e) 20-Sensor Network; (f) 24-Sensor Network; (g) 28-Sensor Network. .	142
Figure 44: Total Travel Time Error Savings with 4 Sensors under Different Levels of 2 <sup>nd</sup> Data Source Coverage: (a) Policy 1; (b) Policy 2.....	146
Figure 45: Total Travel Time Error Savings with 8 Sensors under Different Levels of 2 <sup>nd</sup> Data Source Coverage: (a) Policy 1; (b) Policy 2.....	147

Figure 46: Total Travel Time Error Savings with 12 Sensors under Different Levels of 2 <sup>nd</sup> Data	
Source Coverage: (a) Policy 1; (b) Policy 2.....	147
Figure 47: Total Travel Time Error Savings with 16 Sensors under Different Levels of 2 <sup>nd</sup> Data	
Source Coverage: (a) Policy 1; (b) Policy 2.....	148
Figure 48: Total Travel Time Error Savings with 20 Sensors under Different Levels of 2 <sup>nd</sup> Data	
Source Coverage: (a) Policy 1; (b) Policy 2.....	148
Figure 49: Total Travel Time Error Savings with 24 Sensors under Different Levels of 2 <sup>nd</sup> Data	
Source Coverage: (a) Policy 1; (b) Policy 2.....	149
Figure 50: Total Travel Time Error Savings with 28 Sensors under Different Levels of 2 <sup>nd</sup> Data	
Source Coverage: (a) Policy 1; (b) Policy 2.....	149
Figure 51: Data Validation based Network Optimization .....	154
Figure 52: Time-space Representation.....	160
Figure 53: Non-anticipativity Constraint Generation Example in the Study.....	170
Figure 54: Solutions for 2-Stage Optimization by MR-SAA-N2 Algorithm: (a) Sensor/Relocation Cost Ratio=1.0; (b) Sensor/Relocation Cost Ratio=2.0; (c) Sensor/Relocation Cost Ratio=5.0; and (d) Sensor/Relocation Cost Ratio=10.0.....	185
Figure 55: Solutions for 3-Stage Optimization by MR-SAA-N2 Algorithm: (a) Sensor/Relocation Cost Ratio=1.0; (b) Sensor/Relocation Cost Ratio=2.0; (c) Sensor/Relocation Cost Ratio=5.0; and (d) Sensor/Relocation Cost Ratio=10.0.....	187
Figure 56: Solutions for 4-Stage Optimization by MR-SAA-N2 Algorithm: (a) Sensor/Relocation Cost Ratio=1.0; (b) Sensor/Relocation Cost Ratio=2.0; (c) Sensor/Relocation Cost Ratio=5.0; and (d) Sensor/Relocation Cost Ratio=10.0.....	189
Figure 57: Solutions for 2-Stage Optimization by MR-SAA-N1 Algorithm: (a) Sensor/Relocation Cost Ratio=1.0; (b) Sensor/Relocation Cost Ratio=2.0; (c) Sensor/Relocation Cost Ratio=5.0; and (d) Sensor/Relocation Cost Ratio=10.0.....	192

Figure 58: Solutions for 3-Stage Optimization by MR-SAA-N1 Algorithm: (a) Sensor/Relocation Cost Ratio=1.0; (b) Sensor/Relocation Cost Ratio=2.0; (c) Sensor/Relocation Cost Ratio=5.0; and (d) Sensor/Relocation Cost Ratio=10.0.....	194
Figure 59: Solutions for 4-Stage Optimization by MR-SAA-N1 Algorithm: (a) Sensor/Relocation Cost Ratio=1.0; (b) Sensor/Relocation Cost Ratio=2.0; (c) Sensor/Relocation Cost Ratio=5.0; and (d) Sensor/Relocation Cost Ratio=10.0.....	196
Figure 60: Optimal Sensor & Relocation Numbers with respect to Different S/R Cost Ratios for 2-Stage Optimization.....	200
Figure 61: Optimal Sensor & Relocation Numbers with respect to Different S/R Cost Ratios for 3-Stage Optimization.....	200
Figure 62: Optimal Sensor & Relocation Numbers with respect to Different S/R Cost Ratios for 4-Stage Optimization.....	201
Figure 63: Costs of Sensor Purchase and Relocation Operations of Different Optimization Cases. .....	202
Figure 64: Total Cost w.r.t. Optimal Decisions of Different Cases. ....	203
Figure 65: Flow Chart Demonstrating the Application of the Developed Models in this Dissertation.....	211

# Chapter 1 Introduction

## *1.1 Background and Motivation*

Over the past few years, traffic congestion has become a genuine nightmare to most of the urban commuters. According to the latest report published by Texas A&M Transportation Institute and INRIX Inc. (Schrank et al. 2015), motorists in Washington D.C. waste an average of 82 hours a year stuck in traffic, and an average of around 80 hours in Los Angeles and San Francisco Bay Area. In other cities like Beijing and Sao Paulo, commuters can sometimes get stuck in traffic for several hours per day given severe bad traffic congestion. Providing real-time traffic information is of key significance to Intelligent Transportation System (ITS). Accurate travel time or traffic speed information through Advanced Traveler Information System (ATIS) can provide guidance for travelers who make decisions every day on travel mode, route choice and departure time. Meanwhile, travelers' anxiety can be reduced with a better understanding of their current and future travel time. With a well-organized and reliable traffic surveillance network, ITS can not only assist travelers in understanding their travel time and planning their trips via ATIS but also detect traffic incident and dispatch a patrol team in a timely manner. Therefore, comprehensive and reliable traffic network surveillance is of fundamental significance in building a smart transportation network.

Different kinds of traffic monitoring sensors have emerged in the past several years. The goal of developing more and more advanced traffic monitoring sensors is to collect and report the real-world traffic state, like traffic speed and travel time, with higher accuracy and lower cost. Since the sensor resources are always limited due to high cost and finite budget, the allocation of a given number of sensors to a set of candidate locations on a particular highway corridor or network becomes an optimization problem. The general objective of such optimization problem is to maximize the total surveillance benefit. Specifically, surveillance benefit can be considered as

network flow coverage, Origin-Destination (OD) demand estimation accuracy, and network-level travel time estimation and prediction accuracy. Due to limited number and type of sensors in the planning phase, as well as some installation feasibility concerns in the deployment phase, one is not always able to come up with a full coverage configuration for the network to be monitored. Thus, the decision variables are the locations at which the limited traffic sensors should be installed. A desirable installation configuration should yield satisfactory benefits given all the constraints are met.

Surveillance benefit of a highway network is determined by many factors such as traffic demand and real-time traffic information prediction performance. The inner relationship between a particular deployment configuration and its corresponding surveillance benefit should be explicitly studied and considered. For a particular highway segment monitored by one or several traffic sensors, the surveillance benefit can be directly estimated as, for example, the travel time estimation variance or the expected travel time prediction error. Those types of monitoring benefits can be derived based on the historical data by different methods, like regression and descriptive statistics. In other words, with a highway segment monitored by sensors, we can have a better knowledge about the current and future traffic state on this segment with the data directly collected from the sensors. The most research adopts this assumption in the literature. Since each highway segment in the target monitoring region is not isolated, one question should also be answered. That is, can traffic state detection on one site benefit the traffic state inference on another site? We believe that the spatial patterns of the traffic state across a highway network can increase our knowledge of the network-level information even when some parts are not monitored. Therefore, the spatial pattern of traffic information should not be ignored when a traffic sensor deployment is being planned. It is promising to get a desirable surveillance benefit with limited sensors if the spatial traffic pattern is considered.

Another promising direction in traffic sensor location optimization is to consider the monitoring system as a dynamic sensor network. Traditional traffic count sensors like inductive loop detectors, toll station recorders and license plate recognition cameras are location fixed. Various low-cost portable traffic sensors have emerged in the past several years, such as Bluetooth detectors, WIFI detectors, and some other removable magnetic identification devices. The cost of these newly emerged sensors is decreasing continuously while their performance is improving. With the convenient movement of these sensors, extending the traditional sensor planning model to a dynamic operational model might bring additional surveillance benefits due to the time-dependent traffic fluctuations across the highway network. For example, when a large-scale traffic event is anticipated at some time in a specific area, relocating sensors from the existing network to that area can bring additional surveillance benefits for real-time traffic controls and smart guidance.

## *1.2 Problem Statement and Objective*

This dissertation deals with three major issues about the highway system. They are (1) highway state evolution uncertainty modeling and estimation, (2) static and dynamic sensor network planning model for real-time traffic state surveillance (i.e., online purpose), and (3) multistage sensor placement model for data collection and validation (i.e., offline purpose).

Traffic state at a specific location evolves with some recurrent patterns. This makes it possible to be predicted based on the knowledge of past information. Moreover, the transportation system is an inner correlated system, in which traffic patterns correlate with each other across different locations. Thus state prediction on one location can be inferred based on the data measured at other locations, which is named as the spatial information based traffic state prediction. However, the predictability is not the same for all different highway segments. The state (i.e., travel time) predictability is affected by many factors. For example, non-recurrent traffic event occurrence rate makes the highway segment a stochastic system. Thus the evolution

of traffic parameters such as travel time and traffic speed behaves like stochastic variables, and no model can perfectly predict the state variable. Moreover, state predictability is also related to the choice of measurement variables and the measurement data quality. The problem of how to generally and quantitatively estimate the state prediction uncertainty of a stochastic system with respect to a given measurement space will be studied and answered by this research. Also, the relationship between prediction uncertainty and empirical prediction errors is investigated by this study.

Traffic sensor location optimization problem can generally be defined as, given a number of functioning sensors, determine installation locations, whereby deployment of the sensors at those locations will maximize the surveillance benefit. In this research, we consider the surveillance benefit as providing travelers with accurate real-time travel time information. Specifically, the optimization problem aims to figure out the installation locations at which the sensors can better collaborate to accurately estimate and predict the travel time or travel speed information. The optimization model adopts both the concepts of temporal and spatial information based predictions.

Moreover, the traditional traffic sensor location optimization model is always for planning purpose. Due to the emergence of various portable traffic sensors, the convenience and low cost of relocation operations make it possible to develop a dynamic operational model regarding the sensor network. Since traffic flow is highly time-dependent within a highway network, a dynamic traffic sensor network with limited sensors might bring additional surveillance benefits in comparison with a static one. This research targets to develop a sensor location optimization framework both from a planning and operational perspective. Whether a dynamic sensor network can bring additional surveillance benefit against a static one will be investigated with a real-world case study.

Offline traffic data collection and data quality validation are two other applications of traffic sensor network. Studying temporal-spatial traffic state patterns can help transportation

planners to understand the bottleneck and state predictability of the highway system. For example, to plan a real-time surveillance-based sensor network, one should have some knowledge about the traffic state fluctuations across the network to determine and compare the surveillance effectiveness at different locations. Also, the emergence of various traffic data providers makes it necessary to validate the data quality against ground truth measurements.

For such offline purpose, it is useful to plan a sensor placement strategy to meet different data collection and validation goals. Two issues should be considered for this type of study. First, for large-scale highway networks, how to determine a cost-effective sensor fleet size and come up with a proper stage-wise placement strategy to meet all the requirements. Second, joint usage of a second independent data source may reduce the operational cost of physical sensors. For example, when we have access to a second independent data source and the data quality is validated to be reliable, we can take advantage of this data source to collaboratively collect and verify the temporal-spatial traffic state patterns. Hence, this research will tackle this problem as well and develops a multistage stochastic optimization model with uncertainty on second data source reliability. This optimization model can serve as a supplement to the real-time surveillance-based sensor network planning model.

### *1.3 Research Contributions*

This dissertation contributes three advanced models to the field of the highway traffic information system.

First, we present a general probabilistic model to estimate temporal-spatial system state evolution uncertainty. The proposed model can be specified and applied to any stochastic system to evaluate the surveillance (i.e., measurement) effectiveness for prediction of a particular system state. The concept of conditional entropy is adopted to model the system state evolution uncertainty. The advantage of the proposed model is that the prediction uncertainty evaluation process does not require one to specify the prediction model structure.

Second, we present both a static and a dynamic optimization model to plan traffic sensor placement strategies to improve real-time network surveillance. The proposed model has two advantages compared to existing traffic sensor network models. First, the impact of traffic state spatial correlations on real-time surveillance improvement is explicitly considered. Based on a real-world case study, we found surveillance benefit improvement induced by spatial information-based predictions is not a trivial part. Second, the proposed optimization model provides one with the flexibility to come up with optimal sensor relocation strategies. Especially when traffic demand and travel time uncertainty are heterogeneously distributed in a highway network for a given time period, appropriately relocating some sensors to different locations can further enhance the network surveillance.

Lastly, we presented a sensor placement optimization model with the goal of efficiently collecting and validating traffic information. To the best of the author's knowledge, there are very few studies dealing with traffic sensor network design with the purpose of effectively collecting and validating temporal-spatial traffic state patterns. The basic concept of the optimization model is multi-stage network link coverage model. Moreover, the existence of independent data source providing the same type of traffic state information is considered by the planning model. But data reliability is considered as a stochastic variable and can only be revealed after data validation process. Particularly, a Monte Carlo simulation-based scenario decomposition algorithm is designed to solve the optimization model with endogenous uncertainty.

#### *1.4 Dissertation Organization*

The rest of this dissertation is organized as follows. The next chapter provides a comprehensive literature review related to this research. Firstly, various traffic sensors including traditional sensors and recently emerging portable sensors are summarized and compared. Secondly, existing research in the field of the traffic sensor location optimization problem is reviewed and discussed. In Chapter 3, a probabilistic model is developed to evaluate system state prediction uncertainty

under given measurement space. Further, a real-world case study is conducted to investigate the relationship between empirical prediction errors and prediction uncertainty. Chapter 4 presents a static and a dynamic traffic sensor network optimization model with the objective of providing real-time traffic state surveillance. The proposed model is applied to a real-world commute network and the marginal surveillance benefit with respect to sensor relocation operations is analyzed and discussed. In Chapter 5, a multistage stochastic sensor placement optimization model with the objective of efficiently collecting and validating spatial-temporal traffic data is developed. A case study with a real-world highway network is conducted to demonstrate the proposed model. Practical implications are given based on the case study. Finally, Chapter 6 summarizes the overall work and points out several interesting future research directions.

## Chapter 2 Literature Review

In this chapter, we give a comprehensive review on development of sensor technologies, applications, location optimization problems in the field of the highway system. Section 2.1 briefly discusses the traffic sensor technology development in past years. Sensor technologies are introduced and described according to their functionalities. Section 2.2 gives a literature review on application of those developed traffic sensors based on their monitoring purpose. The review in this subsection is divided into two clusters. They are, deploying sensors for network OD demand estimation and traffic state (i.e., travel time or traffic speed) inference. In section 2.3, we review the studies in terms of the methodology to solve the traffic sensor location optimization problem. Methodologies belonging to different categories are discussed and compared accordingly.

### 2.1 *Traffic Sensor Introduction*

Based on the type of measurement data, traffic sensors can be divided into three clusters. They are point sensors, point-to-point sensors and probe sensors (Xing 2012). Point sensors are those collecting traffic information, such as instantaneous speed, traffic volume, and occupancy, at fixed locations of a highway segment. Point-to-point sensors, also named as paired sensors, are those collaboratively collecting traffic information, such as experienced path travel time and travel speed, by identifying and re-identifying partial of the vehicles within the traffic. With the emergence of Automatic Vehicle Location (AVL) technologies, like Global Positioning System (GPS), real-time location information of some in-vehicle passengers enriches the traffic database with information such as individual vehicle trajectories, travel time and speed data. This type of sensor is named as probe sensor. Based on the taxonomy approach proposed by Xing (2012), we further classify the traffic sensors into five categories according to both their measurement functionalities and installation properties. These are fixed point sensors, fixed paired sensors,

probe sensors, portable point sensors and portable paired sensors. Unlike traditional fixed sensors (e.g., inductive loops), portable sensors can be placed on the side of the road, and do not require the stop of traffic during installation. Moreover, with the development of wireless technologies, portable sensors can be easily installed and relocated. The next following five subsections will review and discuss those traffic sensors according to the above classification.

### 2.1.1 Fixed Point Sensor

Point traffic sensors are the first type of sensors developed and used to monitor and collect traffic data in the world. In the 1920s, Charles Adler Jr., a railway signal engineer, first developed a sensor that was activated when a driver sounded his car horn at an instrumented location. The invention of Charles' sensor declared the birth of traffic sensors (Klein, Milton, and Gibson 2006). Interested readers can refer to the report of Klein, Milton, and Gibson (2006) for more detail on the historical development of traffic sensor in early years. The measurement traffic data by fixed point sensors are usually traffic volume, instantaneous traffic speed, and roadway occupancy. Typical traffic sensing technologies belonging to this category are inductive loop, magnetometer, microwave radar, active/passive infrared, ultrasonic, acoustic and video image processor (Koerner 1976; Caruso and Withanawasam 1999; Sergent 1981; Ahmed, Hussain, and Saadawi 1994; Matsuo, Kaneko, and Matano 1999; Kuhn, Bui, and Pieper 1998; Michalopoulos 1991). Once installed at a fixed location along a highway segment, the fixed point sensor can continuously count the traffic passing it. Many of the well-adopted fixed point sensors such as inductive loop detectors, magnetometer, and magnetic sensors, are installed under the roadway to detect the passing traffic. While other fixed point sensors like acoustic, microwave radar and video-based sensors can be installed relatively easily along the roadside. Figure 1 shows two examples of the in-roadway sensor and the roadside sensor. The main advantage of in-roadway sensors is their insensitivity to inclement weather since those sensors are installed under the pavement. Consequently, they can have better performance compared against the roadside

sensors, the detection accuracy of which is sensitive to the weather (Mimbela and Klein 2000; Klein, Milton, and Gibson 2006; Leduc 2008). However, the installation and maintenance cost of those in-roadway sensors are significantly high, since the installation usually requires pavement cut and is intrusive to traffic. We refer interested readers to the work of Mimbela and Klein (2000), and Klein, Milton, and Gibson (2006) for more detailed and comprehensive comparisons of those well adopted fixed point traffic sensors.

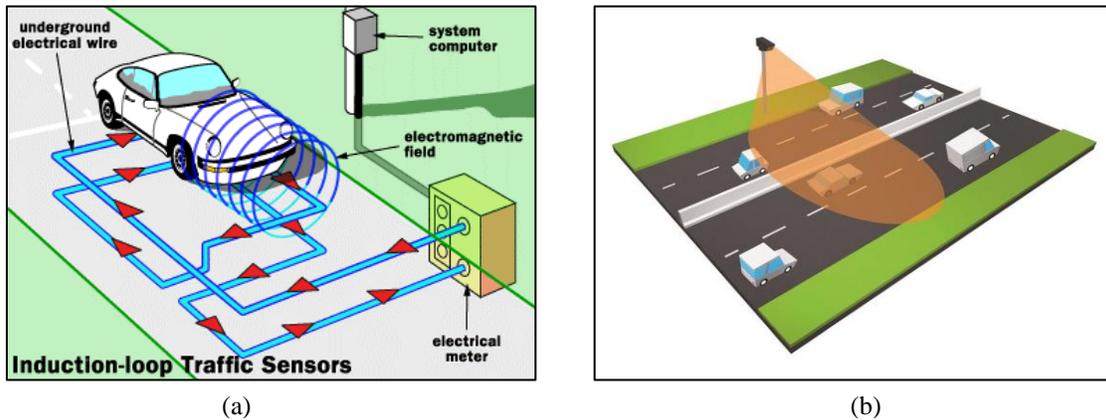


Figure 1: Two Types of Fixed Point Traffic Sensors (a): In-roadway Sensor (b): Roadside Sensor

In addition to those above traditional fixed-point sensors, which are usually used to count traffic volume, measure instantaneous traffic speed and estimate the roadway occupancy, there are scores of newly emerged point sensors and technologies that are capable of providing additional traffic information of interest. For instance, Cheung et al. (2005) proposed to use an advanced magnetic sensor to classify the vehicle type in higher resolution, i.e., passenger car, SUV, van, bus, MT, truck and other. Sen, Siriah, and Raman (2011) developed an acoustic sensing-based technique to classify and report the real-time traffic congestion level. The robust noise filtering technique indicated a promising application of acoustic-based sensors in traffic condition detection. For more introductions to recent success and design of various advanced point traffic sensors, readers are referred to the work of Haoui, Kavalier, and Varaiya (2008), and Losilla et al. (2011).

### 2.1.2 Fixed Paired Sensor

Different from point sensors, paired sensors are used to collect traffic information, such as experienced travel time and average travel speed by collaboratively tracking trajectories of individual vehicles. Paired sensors rely on Automatic Vehicle Identification (AVI) technology (Foote 1974). Passive tags attached to vehicles and electronic interrogators (or readers) are two key components of such a detection system. The system detects the passage of a vehicle at each fixed reading location by monitoring the signal received by the electronic reader. Since each vehicle returns a unique identification signal, the system can calculate the travel time of each responded vehicle by matching those raw detections at each reading location. Automatic License-Plate Recognition (ALPR) were developed and used to monitor the traffic information first in 1976 at the Police Scientific Development Branch in the UK (“Automatic Number Plate Recognition” 2016). Turner et al. (1998) summarized and listed the travel time collection issues arising from the license plate matching technologies. Standards on sample size, observation location distance, matching and screening algorithms were discussed in their work. For additional issues and concerns on license plate recognition systems, readers are referred to the surveys of Gilly and Raimond (2013), and Lad and Patel (2015). Another branch of AVI based travel time data collection technique is Radio Frequency Identification (RFID) technology. RFID emerged as a new approach to monitor traffic during the 1980s (Walton 1983) and has become a mature traffic surveillance technology. Figure 2 gives two illustrative examples of RFID sensing technique and camera-based license plate recognition technique which can be used to collect vehicle travel time data.

Typical fixed paired sensor systems used to monitor and collect travel time data are Toll Stations with RFID transponder, and camera-based license plate recognition system (Hassett 1998; Lindveld et al. 2000; Toppen and Wunderlich 2003; Tanaka 1992; Washburn and Nihan 1999). Scores of advanced travel time estimation algorithms were developed to filter and estimate

the travel time data to improve the travel time estimation accuracy based on such sensing system. Liu et al. (2006) took advantage of neural network and Kalman filter algorithms to predict real-time travel time information for arterials monitored with paired camera recognition sensors. Soriguera, Thorson, and Robusté (2007) developed a simple filtering algorithm to estimate freeway section travel time with data collected by toll infrastructures. Other more advanced algorithms emerged in recent years for corridor travel time estimation with data collected by such AVI based detection system can be seen in the works of Park et al. (2009), Haghani et al. (2010), Lu (2013), and Yang, Ozbay, and Xie (2015). Unlike the point detectors, the paired sensors can only detect the travel time information of individual vehicle equipped with interactive tag or transponder. Therefore, the traffic state information estimated from those detections highly depends on the sample size. As for RFID based sensing system (e.g. toll station), the sample size depends on the number of vehicles passing through the highway section, while the sample size of a camera based license plate recognition not only depends on the number of vehicles passing through the sensing section but also is related to the recognition accuracy of the sensors. Many successful attempts have been provided to improve the identification accuracy and reliability of such license plate recognition systems starting from this point (Chang et al. 2004; Anagnostopoulos et al. 2006; Guo and Liu 2008; Abolghasemi and Ahmadyfard 2009; Wen et al. 2011). State-of-art reviews on automatic license plate recognition techniques and algorithms were given by S. Du et al. (2013), and Ye and Doermann (2015). Interested readers can refer to these studies for further references.

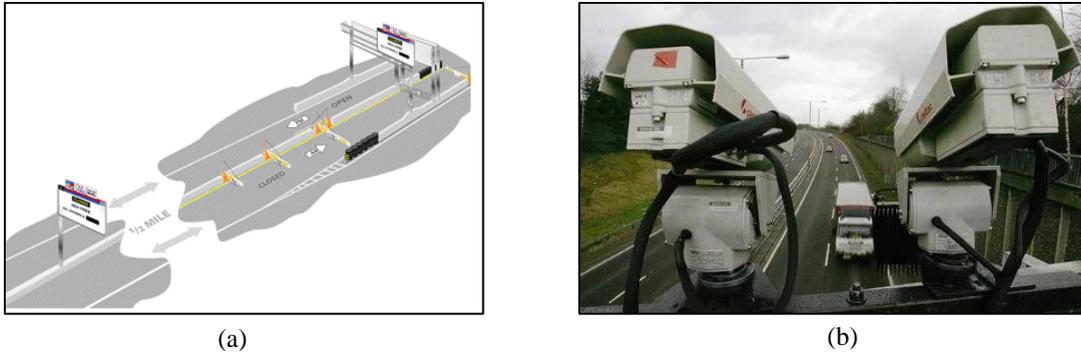


Figure 2: Two Examples of Fixed Paired (AVI) Traffic Sensors (a): RFID Sensing Based Toll Station (b): Camera Based License Plate Recognition Sensors

### 2.1.3 Probe Sensor

In early years, researchers ran equipped vehicles to record the traffic information in highway sections of interest. Based on the positioning and driving speed information of the running vehicle, average travel time and traffic speed information in those particular time windows can be roughly estimated (Von Tomkewitsch 1982; T. K. Liu 1994). This type of traffic data collection method is the prototype of probe sensor concept and is still currently used in particular measurement scenarios due to its higher flexibility compared to installing fixed traffic detectors. Applications of Automatic Vehicle Location (AVL) technologies in the transportation system, such as Global Positioning System (GPS) and electronic Distance Measuring Instruments (DMI's), provide new possibilities for real-time traffic state data measurement and collection. With the rapid growth of in-vehicle route guidance devices and smart cell phones in personal navigation markets, those real-time operating devices have become the main data source providing large-scale travel time and travel speed data, and are generally named as probe sensors.

Based on the operating type of probe vehicles, probe sensors can be classified into three categories, i.e., freight based probe sensors, public transit and taxi based probe sensors and passenger car based probe sensors. Freight vehicles are usually running on the interstate freeways and report their traveling locations to the central management system. Most buses and taxies are operating within a fixed urban region and can report speed and travel time information for those

urban highway sections. Passenger cars make it possible to provide vehicle trajectory information in a large-scale network. But the data availability highly depends on how many vehicles are equipped with functioning navigators and GPS-based mobile phones, i.e., penetration rate. In recent years, numerous studies have been done to estimate and predict the traffic information from data collected by those probe sensors. Zhao et al. (2012) developed a systematic methodology for identifying and ranking bottlenecks using GPS probe data collected from a fleet of 6000 traveling trucks in Washington State. Uno et al. (2009) took advantage of GPS data reported by buses to evaluate travel time reliability and the level of service (LOS) of the road in Hirakata City. Other interesting studies on traffic state estimation using the bus as probe vehicle are done by Bertini and Tantiyanugulchai (2004), Pu, Lin, and Long (2009), and Vanajakshi, Subramanian, and Sivanandan (2009). Compared to freight and bus-based probe data, taxis based probe data seems to be more popular in urban network travel time estimation due to its traveling homogeneity and high sample size. Herring et al. (2010) proposed a probabilistic modeling framework for estimating and predicting arterial travel time distributions using sparsely observed probe data from a fleet of 500 taxis in San Francisco, CA. Jenelius and Koutsopoulos (2013) presented a parametric statistical model for urban road network travel time estimation with low-frequency taxi probe vehicles. Spatial and temporal variations in speed data were considered to improve the estimation accuracy. The sampling frequency in their study is around one report per 2 minute and 780 meters, which is significantly lower than that of previous related studies (Hunter et al. 2009; Hofleitner et al. 2012; Westgate et al. 2013). For the past few years, commercial data companies such as INRIX, HERE and TomTom have been continuously collecting and fusing vehicle trajectory data from various probe sources (e.g., freight vehicles, public transit systems, and passenger cars) to construct network-level travel time and speed database in the United States. Abundant of studies have been done based using those large-scale probe data. For instance, Haghani, Hamed, and Sadabadi (2009) used the Bluetooth ground truth data to validate the quality of INRIX probe data on both freeways and arterials through the I-95

Corridor Coalition Vehicle Probe project. Later in 2014, the validation project was extended by incorporating another two probe data sources, i.e., HERE and TomTom. Fusing data from both physical sensors and probe sensors to improve travel time data quality or comprehensively estimate real-time traffic states also has become an interesting topic in recent years (Bhaskar, Chung, and Dumont 2011; J.-Q. Li et al. 2013; X. Zhang, Hamed, and Haghani 2015).

The main advantage of probe sensors is the data coverage scale. Unlike those physical traffic sensors installed at limited highway sections, probe sensors are capable of reporting the travel time and travel speed data anywhere they are located. Thus, this sensing technique gives the possibilities to monitor real-time traffic state information across the entire highway network. However, the drawback of this traffic sensing technique is the unstable sample size, which highly depends on the penetration rate of the vehicles equipped with such GPS related devices. In other words, the traffic data is not guaranteed to be continuously collected. The sampling and reporting frequency issues in terms of probe sensors have been studied and discussed in the literature (Turner et al. 1998; M. Chen and Chien 2000; Herrera et al. 2010; Jenelius and Koutsopoulos 2015).

#### 2.1.4 Portable Point Sensor

The rapid development of Wireless Sensor Network (WSN) technology casts light on the invention of more advanced and convenient traffic surveillance technologies in past few years. The exceptional features of WSN technology, such as flexibility, cost-effectiveness, and simple installation, enable the development of portable traffic surveillance sensors. Based on the measurement data, we also divide the portable traffic sensors into two clusters, i.e., portable point sensors and portable paired sensors. In this subsection, we briefly introduce and discuss some recently invented portable point traffic sensors. In the next subsection, some typical portable paired traffic sensors will be introduced.

The traffic state data measured by point sensors are instantaneous traffic speed, traffic volume, and occupancy. Unlike the traditional fixed-point sensors, portable point sensors provide the flexibility of temporary installation and monitoring on a highway segment. Pneumatic tubes are one of the prototypes of portable point traffic sensors in early years used to collect traffic direction, speed and volume information (Turner et al. 1998). However, the traffic is required to be stopped during the installation and configuration of such a detection system. One notable advantage of modern portable traffic sensors is the non-intrusive property. In other words, those portable traffic sensors are usually installed at the roadsides. Thus the installation and uninstallation operation will not disrupt the traffic. Kotzenmacher, Minge, and Hao (2005) developed a portable non-intrusive traffic detection system to collect the traffic speed and volume data as an alternative to conventional point sensors, such as inductive loops and road tube counters. The traffic sensing system they developed can be quickly and safely deployed at the roadside. Thus it can temporarily and quickly collect the traffic data on the target highway segments. Additional calibration efforts need to be made to classify the vehicle types by using their portable sensing system. Almorox-Gonzalez et al. (2007) presented a Linear Frequency Modulated Continuous Wave (LFM-CW) based radar sensor for vehicle speed detection. The developed portable sensor was evaluated in real-world traffic surveillance and proved to be a good alternative to conventional intrusive point sensors. A wireless anisotropic magnetic driven traffic sensor was developed by Taghvaeeyan and Rajamani (2014). The presented sensor can just be placed next to the adjacent lane during surveillance and can count traffic volume, measure traffic speed and classify vehicle type. Due to its modular, compact and lightweight properties, the developed sensor can be easily applied to portable traffic surveillance at either intersections or highway segments. Another study by Wahlström et al. (2014) proposed using a portable two-axis magnetometer sensor for detecting vehicle driving direction. Balid, Tafish, and Refai (2015) took advantage of WSN technology and developed a portable sensor system for real-time traffic surveillance. The experimental study indicated 98% accuracy for vehicle counting and detection.

The estimated cost of the proposed portable detection system is less than 40 dollars. The major drawback for their detection system is that the setup does not work for roads with more than two lanes since the magnetic sensor is placed at the roadside and can only detect vehicles passing through the most adjacent lane. This is the common issue for all roadside magnetic traffic sensors. Another concern for the portable traffic sensors is the energy consumption issue due to their common wireless property. With this concern, Komguem et al. (2014) proposed to use classical battery-equipped wireless sensor nodes but having energy harvesting capabilities for their WSN based queue length estimation system, named as WARIM. Even though the sensors can be simply placed on the road surface, they are still intrusive to the traffic during the deployment. Thus, it cannot be completely classified as a portable sensor. But the energy harvesting capability concept they presented inspires the future development of portable sensors.

#### 2.1.5 Portable Paired Sensor

In this subsection, portable paired sensors emerging in recent ITS applications are reviewed and discussed. In common with portable point sensors mentioned above, portable paired sensors also have properties such as lightweight, easy-to-install, and low cost. Compared against portable point sensors, portable paired sensors are usually used to temporarily collect traffic data such as experienced travel time and mean segment travel speed. Due to its collaborative functionality for data collection, portable paired sensors are also called as portable point-to-point sensors (Xing 2012). Figure 3 presents two typical applications of portable paired sensors in travel time data collection (i.e., Bluetooth and WiFi detection techniques).

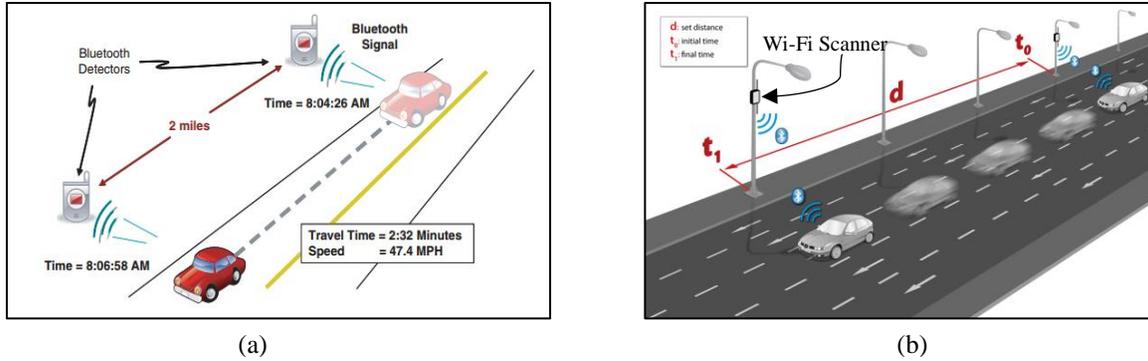


Figure 3: Illustration of Two Portable Traffic Detection Systems (a): Bluetooth Detection (b): WiFi Detection.

Bluetooth Media Access Control Scanner (BMS) has become an attractive alternative to collect travel time data by tracking individual vehicles equipped with Bluetooth devices in recent years. The concept behind BMS based travel time collection is simple. Discoverable Bluetooth (BT) devices, such as mobile phones, headphones, and vehicle navigation systems, can be scanned and identified by a nearby BMS with their unique Media Access Control Identifier (MAC-ID). Thus the travel time of a vehicle with a discovered MAC-ID can be easily calculated as the difference between the two timestamps recorded by two particular BMS devices. Interested readers for a fundamental understanding of Bluetooth travel time collection technique are referred to the work by Haghani et al. (2010), and Bhaskar and Chung (2013). In recent past few years, BT based traffic monitoring technique has been proved to be a success with their applications and continues drawing transportation researchers' attention in the development of ITS. For instance, Haseman, Wasson, and Bullock (2010) proposed quantifiable metrics for a state transportation agency to evaluate work zone mobility performance with 1.4 million travel time records collected by a set of BT sensors temporarily deployed at the target region. They concluded that the flexibility of the real-time monitoring technique might enable future contracts for other innovative travel time reliability analysis. Martchouk, Mannering, and Bullock (2010) used BT detections to analyze both the overall travel time variability and individual vehicle's travel time pattern on freeways. Since BT sensors can detect individual vehicle trajectory, Hainen et al. (2011) proposed a route choice estimation framework with BT detection samples as a surrogate

for license plate matching data. BT detection as a flexible surveillance alternative can also be applied to other travel modes, such as bicycle travel time estimation (Mei, Wang, and Chen 2012), pedestrian travel pattern analysis (Malinovskiy and Wang 2012), and crowd density estimation (Schauer, Werner, and Marcus 2014).

On the other hand, many works have been done to investigate the detection accuracy and effectiveness. For example, field experiments were conducted by Malinovskiy et al. (2010 a) to investigate the effects of antenna selection on travel time collection reliability. Key conclusions were drawn that omnidirectional antenna yields in larger detection zone but are subject to more noise and bigger spatial errors, while the directional antenna results in a smaller detection zone but is subject to a lower sampling rate. By conducting field experiments under various configurations, Malinovskiy et al. (2011) suggested a larger detection zone is desirable despite the apparent loss of accuracy as larger sample size will reduce random error rates. Moreover, Brennan et al. (2010) investigated the influence of vertical placement on data collection efficiency with Bluetooth collection devices. Based on a 24-hour empirical dataset collected from I-65 in Indianapolis, the authors found 7.4% of the vehicles within 30 feet and 6.6% of the vehicles between 102 and 114 feet had a discoverable MAC address.

Similarly, Wi-Fi Media Access Control Scanners (WMS) which can scan MAC-ID used in Wi-Fi communication can also be applied in traffic monitoring by identifying vehicles equipped with corresponding discoverable devices. Compared against BT detection technique, WMS based traffic surveillance technique has not been widely used, and its usage is still being explored. To the best of our knowledge, very few studies have been done in terms of using WMS to monitor and collect traffic information. In this limited literature, Lubber et al. (2011) proposed an additional Wi-Fi based vehicle identification and re-identification approach to measure travel times and mean travel speeds. DeZani et al. (2012) presented an ATIS developed on Android platform to provide travelers with real-time traffic congestion information detected by a preinstalled Wi-Fi network. Geographic locations of each vehicle with discoverable Wi-Fi

identifier can be detected and aggregated into the central server. The server then calculates the traffic congestion level of each roadway and reports this information to the travelers.

BMS and WMS as two promising alternatives for traffic data collection sensors have similar operation concepts. However, the communication difference in BT and Wi-Fi technology, the penetration rate of BT and Wi-Fi enabled devices might yield a significant difference in the quality of travel time data being collected and matched. Benefits, challenges and future enhancements in terms of BT and Wi-Fi-based crowd data collection techniques were thoroughly discussed by Abedi, Bhaskar, and Chung (2013). Empirically, Abbott-Jard, Shah, and Bhaskar (2013) evaluated the BMS and WMS based travel time collection reliability and quality. As is indicated by their empirical analysis, BT sensors seem to collect more samples than Wi-Fi sensors due to the larger usage of BT enabled devices in vehicles. But this did not imply WMS based sensing technology will be completely outperformed by BMS based sensing technology in the future due to the rapid growth of Wi-Fi enabled devices in the mobile and vehicle market.

#### 2.1.6 Summary

A summary of the above traffic sensors is presented in Table 1. The major strengths and weaknesses regarding each type of sensor are listed.

Table 1: Comparison of Various Typical Traffic Sensing Technologies

Class	Typical Technology	Strengths	Weaknesses
Fixed Point Sensors	Inductive Loop	<ul style="list-style-type: none"> <li>• Insensitive to inclement weather like, fog, rain and snow</li> <li>• Mature, well-understood technology</li> </ul>	<ul style="list-style-type: none"> <li>• Installation requires pavement cut</li> <li>• High maintenance cost</li> <li>• Multiple units required to monitor a location</li> </ul>
	Magnetometer	<ul style="list-style-type: none"> <li>• Insensitive to inclement weather like, fog, rain and snow</li> <li>• Less susceptible than loops to stresses of traffic</li> </ul>	<ul style="list-style-type: none"> <li>• Installation requires pavement cut</li> <li>• High maintenance cost</li> <li>• Multiple units required to monitor a location</li> </ul>
	Microwave Radar	<ul style="list-style-type: none"> <li>• Insensitive to inclement weather</li> <li>• Multiple lane operations</li> </ul>	<ul style="list-style-type: none"> <li>• Cannot detect stopped vehicles</li> </ul>
	Ultrasonic	<ul style="list-style-type: none"> <li>• Multiple lane operations</li> <li>• Capable of detecting over height vehicle</li> </ul>	<ul style="list-style-type: none"> <li>• Sensitive to temperature change and extreme air turbulence</li> </ul>
	Acoustic	<ul style="list-style-type: none"> <li>• Multiple lane operations</li> <li>• Insensitive to precipitation</li> </ul>	<ul style="list-style-type: none"> <li>• Not good at monitoring slow-moving vehicles</li> </ul>
Fixed Paired Sensors	RFID (e.g., toll Station)	<ul style="list-style-type: none"> <li>• Multiple lane operations</li> <li>• High penetration rate and matching accuracy</li> </ul>	<ul style="list-style-type: none"> <li>• Insensitive to inclement weather</li> <li>• Limited surveillance coverage</li> </ul>
	License Plate Matching	<ul style="list-style-type: none"> <li>• Multiple lane operations</li> <li>• Higher flexibility than RFID based fixed paired sensors</li> </ul>	<ul style="list-style-type: none"> <li>• Sensitive to inclement weather like, fog, snow and rain</li> <li>• High calculation burden</li> </ul>
Probe Sensors	Freight Probe	<ul style="list-style-type: none"> <li>• The high penetration rate in freeway</li> <li>• Large range of temporal coverage</li> </ul>	<ul style="list-style-type: none"> <li>• Limited spatial coverage (i.e., freeway)</li> </ul>
	Bus/Taxi Probe	<ul style="list-style-type: none"> <li>• The high penetration rate in urban area</li> </ul>	<ul style="list-style-type: none"> <li>• Limited spatial coverage (i.e., urban network)</li> </ul>

			<ul style="list-style-type: none"> <li>Limited temporal coverage (i.e., day time)</li> </ul>
	Passenger Car Probe	<ul style="list-style-type: none"> <li>Large-scale spatial coverage</li> <li>Large range of temporal coverage</li> </ul>	<ul style="list-style-type: none"> <li>Penetration rate is of high variance (i.e., highly depends on the number of GPS users)</li> </ul>
Portable Point Sensors	Magnetic	<ul style="list-style-type: none"> <li>Lightweight and high portability</li> <li>Insensitive to inclement weather such as snow, rain, and fog</li> </ul>	<ul style="list-style-type: none"> <li>Multiple lane operations unavailable</li> <li>Battery consuming for wireless devices</li> </ul>
	Radar	<ul style="list-style-type: none"> <li>Lightweight and high portability</li> <li>Multiple lane operations available</li> </ul>	<ul style="list-style-type: none"> <li>Only insensitive to short ranges of inclement weather</li> <li>Battery consuming for wireless devices</li> </ul>
Portable Paired Sensors	Bluetooth	<ul style="list-style-type: none"> <li>Low cost, lightweight, and high portability</li> <li>High privacy protection</li> <li>Multiple lane operations available</li> <li>Insensitive to inclement weather</li> <li>Mature and large experience base</li> </ul>	<ul style="list-style-type: none"> <li>Not suitable for the short-distance roadway segment</li> <li>Sample size highly depends on the number of in-vehicle BT devices</li> </ul>
	Wi-Fi	<ul style="list-style-type: none"> <li>Low cost, lightweight, and high portability</li> <li>High privacy protection</li> <li>Insensitive to inclement weather</li> <li>Multiple lane operations available</li> </ul>	<ul style="list-style-type: none"> <li>Not suitable for the short-distance roadway segment</li> <li>Sample size highly depends on the number of in-vehicle Wi-Fi devices</li> <li>Lack of practical experience</li> </ul>

## 2.2 *Traffic Sensor Location Optimization*

Locating traffic sensors on a transportation network aims to measure various traffic-related information, such as traffic volume, OD flow, travel time and travel speed. That information cannot only provide insights into transportation planning and operation agencies but can also serve as understandable traveling metrics to travelers. Since different sensors have different functionalities, locating each type of sensors mainly depends on the surveillance purpose. For example, point sensors are good options to measure traffic volume and spot speed information, while paired sensors, as well as probe sensors, are ideal tools to measure the experienced travel time information either for a short segment or a long route. In the abundant literature, traffic sensor location problem can be divided into two major clusters based on the deployment purpose. The first one is about traffic flow measurement, and the other one is about real-time traffic state information (e.g., travel time and travel speed) collection and provision. In next following two subsections, we will review and discuss the most relevant studies regarding each research track.

### 2.2.1 Traffic Flow Measurement and Estimation

Measuring traffic flows with sensors has gain growing interests in the past few years due to its relevance to transportation management and traffic control. Based on the monitoring purpose, flow measurement-based sensor location problems (SLP) can be divided into two tracks. The first one is directly using traffic flow as benefit quantification index. For example, by looking at the historical traffic volume distribution, one can identify the most critical links and deploy traffic counting sensors on such links to retrieve traffic volume information of interest. The second type SLP talks about how to optimally place flow counting sensors in part of the network to observe or estimate the network-level flows (e.g., route flow and OD-pair flow) as accurate as possible. In literature, the second type flow-based SLP is more popular and complicated than the first one due to its practical application in highway traffic surveillance.

For the aforementioned first type flow based SPL problem, one key assumption is that the total number of trips or trip variances are known as deterministic quantities. The intrinsic concept behind this problem is to monitor trips and traffic flows greedily. Teodorovic et al. (2002) presented a bi-objective model to determine the locations of AVI sensors used to monitor traffic flows. The two objectives are maximizing the total number of readings along a route for each OD pair, and the total number of OD pairs covered by the installed sensors. One assumption that an OD pair is covered only if its shortest path is covered is adopted in this model. This may limit the OD demand estimation in post-deployment data collection process since there are usually multiple paths traveling by vehicles belonging to the same OD pair. A. Chen, Chootinan, and Pravinongvuth (2004) extended the models proposed by Teodorovic et al. (2002) by considering the coverage of multiple paths for each OD pair. Prior knowledge of the traffic flow on each path is used to weight and quantify the route and OD pair coverage benefits. Also, a third objective was incorporated, that is, minimizing the number of readers used to monitoring those routes. Mirchandani, Gentili, and He (2009) considered the sensor location problem as a vehicle-miles monitoring problem (VMMP) and presented a formulation to determine optimal AVI sensors locations to maximize the total vehicle-miles. The arc-based VMMP formulation is also known as the constrained covering problem and was previously proved to be NP-complete (Plesník 1999). Subsequently, a greedy Heuristic is proposed to solve the proposed VMMP. Asudegi and Haghani (2013) proposed integer programming models for determining optimal number and location of Bluetooth sensors. Although the main purpose for such a deployment is to collect travel time data in a more reliable manner, they additionally considered two objectives: coverage of a high percentage of total traffic volume in the network, and covering as many as OD pairs as possible. This further implies the distribution of traffic volume and OD topologies of a particular network play key roles in determination of sensor locations. Essentially, the above SPLs have a similar objective, which is monitoring as many of the traveling vehicles as possible. However,

monitoring as many vehicles as possible may not guarantee the maximization of surveillance benefit (e.g. route flow or OD flow estimation accuracy).

Flow estimation based SLPs usually assume the traffic flows over the roads can be described by a network flow function with respect to the observations from a set of segments. Based on the argument that traffic information provided by the sensors can be suitably used both for traffic flow derivation and for OD matrix estimation, Bianco, Confessore, and Reverberi (2001) defined and solved the sensor location problem (SLP) with the objective to infer all traffic flows of a network with minimum number of counting sensors installed. And the OD matrix estimation error was proved to be bound by the proposed greedy Heuristic approaches. In a similar concept, Chung (2001) presented an optimal network sampling framework for estimating trip matrices, which was viewed as a prototype of the network count location problem (NCLP). Ehlert, Bell, and Grosso (2006) reformulated the original NCLP by incorporating two extensions: (1) using original detector counts to update the link choice proportions, and (2) taking the prior OD flows into account. The proposed formulation was applied to a real network with moderate size with satisfactory solution quality. Nevertheless, large size networks require a more efficient algorithm to guarantee a solution within a reasonable time. Theoretically speaking, prior information of particular OD pair or route flows is of significant importance to estimate the network-level OD-pair flows. This is mainly because the number of independent OD-pair or route flows is much larger than the number of independent link flows, even if the flow of each link is known, the solution set of OD-pair flow estimation problem is still infinite (E. Castillo et al. 2002). Enrique Castillo et al. (2010) presented three formulations dealing with the estimation of route flows based on the subsets of monitored links. The first one is about minimizing the number of vehicle scanning cameras to be used to estimate a given subset of route flows. The second one is to figure out the subset of links to be monitored for a given number of scanning sensors. Finally, they took the scanning error issue into account and reconsidered the previous two problems. An application advantage is that when not enough cameras are available to solve the

overall estimation problem in one run, the proposed model can be applied multiple times to improve the information gain, given the sensors are portable and can be relocated for different runs. Zhou and List (2010) derived analytical formulations to describe OD demand estimation variance propagation by explicitly taking into account several important error sources, such as historical data uncertainty, sensor measurement errors and approximation errors. Based on the derived estimation variance scheme, a scenario-based stochastic optimization procedure and a beam search algorithm were developed to find the suboptimal locations of point and point-to-point (AVI) sensors. Unlike previous deterministic flow-estimation based sensor location problem (SLP), Fei, Mahmassani, and Murray-Tuite (2013) considered the SLP under traffic flow uncertainty. In particular, the occurrence of random events (e.g., accident) may redistribute the traffic flow to a large extent, and the placement of traffic counting sensors with the objective to maximize the OD coverage and information gain should be robust to those random impacts. Therefore, a nonlinear two-stage stochastic model was developed. The first stage generates sensor placement strategy to maximize OD coverage and information gain before considering any random events. The second stage deals with stochastic events and calculates the recourse function by incorporating the cost of vehicular flow changes under random events. Due to the high nonlinearity of the objective function as well as an extremely large number of second-stage realizations, an iterative heuristic called Hybrid Greedy Randomized Adaptive Search Procedure (HGRASP) was presented to find near-optimal solutions.

In theory, Gentili and Mirchandani (2012) classified traffic flow estimation based SLPs into two categories: the Sensor Location Flow-Observability Problem and the Sensor Location Flow-Estimation Problem. The major difference between those two problems is the solution space. Sensor Location Flow-Observability problem mainly answers two questions: (1) whether the flow on each link or path within can be exactly determined given partial observations from the network? (2) Where are the best locations to install counting sensors to exactly infer the flow of the entire network? The intrinsic concept behind this type of problem is flow conservation law.

Based on the flow conservation rule, optimal locations for a given number of counting sensors or a minimum number of counting sensors as well as the corresponding installation locations are found with the objective of uniquely determining the network flow distribution. On the other side, Sensor Location Flow-Estimation Problem mainly deals with the situation in which the unique observability of flows is not possible. This situation may happen in two cases: either when the sensor coefficient matrix associated to the network does not have full rank or when the budget constraints limit the total number of sensors. Prior estimate information regarding the route or OD flow must be considered as a reference of the estimation error or variance to figure out the most suitable sensor locations in this situation (Enrique Castillo et al. 2010). For a comprehensive summary of SLP formulations in terms of Flow-Observability and Flow-Estimation problem, we recommend the readers to refer to the work of Gentili and Mirchandani (2012).

### 2.2.2 Traffic Time Collection and Estimation

Another important application of locating sensors on a highway network is to provide traffic state information helping travelers to understand and plan their trips. Travel time information and average travel speed information are two typical metrics index depicting the traffic state in transportation engineering. Since travel time and average travel speed are convertible, and travel time can be directly used to understand the travel cost, it is commonly generated as the major deliverable to travelers in ATIS. Therefore, in sensor location problem (SLP) with the objective of providing real-time traffic state information, people always refer to the measurement and estimation of travel time information. There are mainly two steps with respect to locating sensors to provide travel time information. The first step is to use the deployed sensors to collect traffic-related data, named as the raw data collection process. The type of the raw data depends on the type of sensors used. For example, if two point sensors are used, the raw data collected are instantaneous traffic speed at two fixed locations, while the raw data would directly be experienced travel time if the AVI sensors are used. The second step is about travel time

estimation or prediction based on the raw data collected, such as, converting point speed data to travel time data or predicting the real-time travel time information based on the most recently collected travel time data. Due to the significance of travel time information in large-scale applications, locating various sensors to retrieve and provide travel time information attracts more interests in the field of SLP research compared with the flow-estimation based SLP. The sensor location optimization problems proposed and resolved in this dissertation are also from the travel time information collection and estimation perspectives. Therefore, we will present a thorough review of the existing methodologies and techniques related to travel time provision based SLP in this subsection. Moreover, since different types of sensors collect and estimate travel time in different ways, we further classify the existing studies according to the sensor types (i.e., point sensor, paired sensor, and probe sensor).

#### 2.2.2.1 Point Sensor Based SLP

Point sensor can be used to measure spot speed data at a particular highway location. Since the interest of estimation is travel time, spot speed data collected by point sensors need to be converted to travel time by some specific methods, such as regression, mid-point estimation, and flow-density model. For traffic surveillance, the speed detector density on the road is very important because it affects the precision of the measurement of the travel time to a large extent. Therefore, for point sensor location problem, there are usually two important aspects. One is to find the optimal installation segments; the other one is to determine the optimal density of the speed detectors in those segments.

For a particular highway segment where speed detectors are needed to be installed, the choice of detector density is highly related to the investment cost and the travel time estimation error. Thus, it is important to determine an appropriate deployment density for the different purpose under various situations. Chan and Lam (2002) considered the tradeoff between investment cost and travel time measurement error and proposed a bi-level programming model to determine the optimal speed detector density to minimize both investment cost and travel time

measurement error. They assumed travel time error variance as the product of the link travel time and the measurement travel time error dispersion function, which is further assumed to be a function of the speed detector density, volume/capacity ratio and the scaling factors of investment cost. Empirical data was used to verify and calibrate the proposed error variance model. The presented travel time error variance function with respect to speed detector density is in a general form and can be easily used as a reference to determine speed detector density in other highway segments (Edara et al. 2008). Ozbay, Bartin, and Chien (2004) also investigated the impact of sensor density on travel time estimation error with empirical data from the South Jersey Real-Time Motorist Information System project. They found that increasing sensor numbers density did not necessarily improve the travel time estimation accuracy both in recurrent and non-recurrent events. This phenomenon was also found and discussed in the study of Bartin, Ozbay, and Iyigun (2007), in which a conclusion was drawn that the marginal surveillance benefit decreased as the of the number of point sensors within a segment increased. Similarly, Chaudhuri et al. (2010) used field data to investigate how the inaccuracy of the travel time estimates was affected by increased sensor spacing. Hypothetical uniform point sensor spacing cases, like 0.5, 1, 1.5, 2, 2.5, and 3 miles, were examined. The analysis showed that the actual location of the sensors is the key element in the estimation of travel time, even though the sensor spacing did affect the estimation accuracy. Rather, it was essential to increase sensor density in major bottleneck areas to improve the estimation accuracy.

The studies mentioned above mainly talked about the optimal point sensor density issue. As is indicated by Chaudhuri et al. (2010), location plays a key role in the error of the travel time estimates. Therefore, only considering point sensor density is not enough for a network-level traffic monitoring, especially when the sensor resource is very limited. Ban et al. (2009) formulated the problem of determining optimal sensor locations as a dynamic programming (DP) model, with the objective defined based on link travel time mean square errors (MSEs). Two important implications were obtained from their numerical experiments: (1) it is optimal to place

many sensors in bottleneck areas and deploy a few in free-flow segments; (2) There should be an optimal number of sensors to use, beyond which installing more sensors is not beneficial. Based on the historical spatial-temporal speed and travel time profiles, Kianfar and Edara (2010) proposed a new clustering-based methodology to identify optimal point sensors installation locations with the objective to minimize travel time estimation errors. The basic concept is to group freeway sections with identical or similar traffic speed patterns together and then determine the sensors locations based on those final grouped clusters. Rather, a specific travel time estimation approach using speed data detected from those final clusters was developed. With the proposed location clustering method and corresponding travel time estimation approach, optimal locations for speed detectors were identified. Compared to the conventional mid-point sensor placement strategy, the optimal placement method proposed can not only save sensor resources but also can produce better travel time estimates. In the same concept, Kim et al. (2011) presented a genetic algorithm-based optimization framework to determine locations for speed detectors in freeways. For a particular freeway corridor, point speed and travel time information were obtained from tremendous simulation runs. Based on the summarized speed and travel time profiles, different combinations of sensor locations were evaluated based on the fitness function, which is calculated as the mean absolute relative error (MARE) between estimated and actual travel time. Numerical experiments indicated the travel time estimate error could be guaranteed within 10% in various traffic conditions. Previous studies were all formulated as nonlinear programs, and only heuristic approaches could be used to seek close to optimal solution. Under the same objective (i.e., minimizing travel time estimate errors), Danczyk and Liu (2011) proposed an approach to transform the nonlinear program into an equivalent mixed-integer linear model, which can be easily solved to optimality using resource constrained shortest path algorithms. A common critical element in the above three studies is that travel time, and speed profile should be known in a high resolution within the target region. Otherwise, it is not possible to comprehensively evaluate all of the candidate locations. Moreover, one inexplicit assumption

should be confirmed before applying those empirical data based optimization approaches. That is, the future temporal-spatial speed and travel time patterns should be relatively consistent with the historical ones. An interesting research direction in this field should be investigating the monitoring benefit (i.e., travel time estimate errors) deviations from a before-and-after study perspective.

#### 2.2.2.2 Paired Sensor Based SLP

As is introduced in section 2.1, paired sensors are used to directly collect travel time data by tracking the trajectories of individual vehicles. Successful identifications of a vehicle by both component sensors produce a valid detection, which can be viewed as a sampling point to infer the average travel time. Statistically speaking, the number of valid detections is of key importance for estimating the actual traffic state of the monitoring region. For a particular highway segment with relatively consistent daily traffic flow, the number of valid detections is basically determined by the penetration rate. Therefore, penetration rate, as well as traffic dynamics, are two key elements in dealing with paired sensor based SLPs.

Yang and Miller-Hooks (2002) proposed a binary programming model as well as a Heuristic to select information critical arcs (ICAs) within a traffic network given a priori information on the travel time variance and covariance. The basic interpretation of finding ICA is to find highway segments with high traffic dynamics (e.g., high travel time variance). Even though they did not explicitly claim the problem as a sensor location optimization problem, the selected ICAs can be optimal locations to install travel time collection sensors. Sherali, Desai, and Rakha (2006) used coefficient of variations (CV) of traffic demand as objective parameters describing the benefit of travel time measurements over various paths, and proposed a quadratic binary programming model to determine the optimal locations for AVI tag readers. Simulations were run to generate the associated benefit factor coefficients for each path between any two candidate reader installation locations. This benefit factor is similar to that in the ICAs selection study by Yang and Miller-Hooks (2002), in which travel time variance is used as benefit

coefficient. Depicting travel time collection benefit in this way does make sense since the surveillance benefit intrinsically comes from garnering as much information about the variability in travel time as possible. In other words, if the travel time of a roadway is nearly the same throughout the day, it is of no value to install sensors to collect its travel time information, and inference from GPS probe vehicles might be enough to tell the whole story. Asudegi and Haghani (2013) presented two formulations for Bluetooth sensor locations optimization. Maximizing total travel time CV values across the network was considered as one of their objectives. They argued that since mean value of travel time may fall into different ranges across all the links, CV value was more suitable than a single variance when quantifying travel time collection benefit. Also, they pointed out that maximizing network-level travel time CV value and minimizing travel time prediction error were not equivalent and should be considered separately according to the user's preference.

Deploying sensors with travel time CV value or historical variance as benefit factors aims to capture as many travel time changes as possible. To guarantee satisfactory travel time prediction errors, one should consider three issues: (1) data quality, (2) fitness of prediction model, and (3) travel behavior predictability. Mirchandani, Gentili, and He (2009) presented two binary programming models to determine AVI sensor locations. In their second model, travel time prediction reliability was chosen as the deployment benefit. Considering the mean travel time along a roadway changes dynamically and follows a normal distribution with a priori information, they proposed to update the prior distribution with samples from detection according to Bayesian theory. Subsequently, optimization model was used to determine from which segments or routes to sample travel times so as to maximize the variance reduction of the predicted travel time. Similar monitoring benefit can also be seen in the study of Zhu et al. (2014), in which travel time estimation variance was considered as a function of sampling size as well as a prior variance. Travel time estimation or prediction variance reduction as an alternative

objective gives a more understandable way to quantify the monitoring benefit, which can be viewed as the information collection reliability.

Travel time estimation or prediction error is a direct index describing the monitoring benefit of paired sensors. Though paired sensors can directly collect and measure travel time information for a particular roadway segment, there might be noise from different sources, such as measurement errors and detection outliers. Therefore, specific estimation techniques are usually required to process and filter those raw data points to give an accurate estimate of the mean travel time value. Moreover, considering data collection delay, prediction techniques are sometimes required for real-time information provision. Xing, Zhou, and Taylor (2013) proposed an information-theoretic approach to evaluate hybrid traffic sensors deployment strategies. As a by-product of the proposed Kalman filtering travel time estimation framework, the corresponding posterior error matrices were used to quantify the travel time estimation uncertainty reduction with respect to a particular deployment configuration. An analytical determinant maximization model, as well as a beam-search heuristic, were given to search for the optimal sensor deployment configuration iteratively. Considering the travel time prediction inaccuracy from existing inductive loops, Park and Haghani (2015) developed a two-stage integer programming model to determine the number and installation locations of Bluetooth sensors with the objective of minimizing corridor level travel time prediction error. The Empirical analysis was first conducted to evaluate the prediction accuracy of existing loops with real-world probe data. Then Bluetooth sensors were deployed to overcome the drawback of loop-based travel time prediction, with the assumption that Bluetooth based travel time estimation is perfect and can be directly used for real-time operations. However, there are two issues remaining to be further discussed and resolved: (1) what is the deviation of Bluetooth based travel time prediction error in addition to the measurement error, since only measurement error is considered and the prediction is assumed to be perfect; (2) what is the acceptable error for real-time travel time information provision in such applications. The second question is tricky and has not been answered by any

existing study. For example, 3% and 4% prediction errors nearly have the same influence in a real-world application. But if they are weighted with the traffic flow, the objective value can change to an extremely large extent. Consequently, the deployment strategy will largely be affected.

Paired sensor based traffic monitoring system requires synchronization of vehicle detection information from multiple locations. For most existing studies regarding to AVI based sensor location problem, either sensor failure issue is not considered, or the consequences of sensor failure are assumed to be trivial matters. There are also a few studies working on AVI sensor location problem considering the impact of sensor failure. Li and Ouyang (2011) proposed a reliable facility location model to optimize the paired traffic sensor deployment strategy considering probabilistic sensor failures. According to the proposed valid sensor assignment (i.e., pairing) rule, the surveillance benefit with an exponential number of sensor failure scenarios was consolidated into a single compact expression. The optimization problem was then formulated as a mixed integer programming model. For the same problem, alternative formulations including a continuum approximation model and reliable fixed-charge sensor location models were also given in a later study by Li and Ouyang (2012). Critical parameter settings, such as failure probability and spatial heterogeneity, were discussed based on the results from bunches of numerical case studies. Danczyk, Di, and Liu (2016) developed a probabilistic optimization model with the objective to minimize expected travel time estimation error. Sensor failure scenarios were completely and uniquely examined with a customized binary based enumeration scheme. Numerical experiments indicated optimal sensor placement strategy with probability concern was significantly different from that without sensor failure considerations. In other words, sensor failure has non-trivial consequences on the surveillance benefits.

### 2.2.2.3 Probe Sensor Based SLP

Estimating urban traffic conditions through probe sensing techniques has attracted increasing attention in recent years. As is introduced in the previous subsection about the probe

sensor techniques, the primary goals of most studies are how to accurately estimate the traffic condition based on the available data reported by probe vehicles. However, few research studies are focusing on the problem that how and where to dispatch probe vehicles to improve urban traffic surveillance. For convenience, we name this problem as the probe sensor location problem (PSLP). Extremely high operation cost for dispatching and controlling probe vehicles is the main reason for the rarity of this type of research. For example, to obtain a satisfactory level of traffic condition estimation accuracy for a particular network, the number of dispatched floating cars should be large enough according to the sampling theory. As a consequence, the expense of operating and controlling those floating cars will be fairly large. Instead, purchasing and deploying some static traffic sensors along the highways would be less expensive and more reliable. Therefore, studying the PSLP theoretically makes sense, but is of little practical application.

To the best of our knowledge, there is only one recognized study in the literature dealing with the PSLP. (R. Du et al. 2015) proposed two patrol algorithms to plan the paths of controllable floating cars to proactively participate in urban traffic monitoring system considering the unevenness of taxi traces. In other words, the controllable floating cars are viewed as dynamic sensors as a supplement of other probe data sources to provide traffic information in a particular urban network. By applying the proposed floating car patrolling algorithms, the network-level traffic state estimation error decreases from 35% to 10%, compared against the random sampling approach. However, the patrolling cost and the operation issue were not discussed in detail. In all, improving the urban network traffic monitoring reliability with controllable floating cars might be a doable method. But the trade-off between the operation cost and surveillance improvement should be well coordinated from a long-term perspective.

### 2.2.3 Summary

To summarize this chapter, this subsection further presents and discusses several key milestone studies in terms of the traffic sensor location optimization problem among the abundant literature. To provide a more clear view, Table 2 is provided to demonstrate and compare those milestone SLP studies from different application perspectives, i.e., sensor type, optimization objective, application type and benefit quantification.

As is shown in the table, the sensor type of SLP in the literature are generally divided into two main categories based on sensor functionality. Point sensor is used to measure spot-based traffic states, such as traffic speed and volume, while paired sensors are used to mainly provide path travel time and flow information by tracking individual vehicle trajectories. Accordingly, the optimization objective can be defined based on the type of sensor used. In early years, the objectives of deploying point sensors are mainly collecting and providing travel time information (Chan and Lam 2002; Danczyk and Liu 2011; Kianfar and Edara 2010). Travel time estimation approaches based on those sensors are indirect, and empirical data analysis is always required to generate the sensor coverage benefit. However, travel time estimated from point sensors is less accurate than that estimated directly by AVI sensors. With the rapid development of various low-cost AVI traffic detection technologies, paired sensors have become more popular than traditional point sensors in terms of travel time estimation (Asudegi and Haghani 2013; H. Park and Haghani 2015). But, the traditional point sensors are still within researchers' scopes, especially in the field of traffic flow observation and O-D demand estimation (Bianco, Confessore, and Reverberi 2001; Fei, Mahmassani, and Murray-Tuite 2013).

From some other perspectives, Li and Ouyang (2012) first considered the sensor failure effect in AVI traffic sensor location problem by proposing a valid sensor pair-up rule as well as a set of heuristic algorithms. Xing, Zhou, and Taylor (2013) developed two sensor location determination algorithms for probe sensors, point sensors and paired sensors based on Kalman

data fusion technique. Also, to the best of the author's knowledge, the optimization model developed by Zhu et al. (2014) is the first among the literature considering the problem from a dynamic routing perspective given the mobility of particular traffic sensors. However, the operational cost issue has not been explicitly discussed. In other words, infinite movement of traffic monitoring sensors is not practically possible. Park and Haghani (2015) considered the sensor relocation budget (i.e., number of movements) as a constraint in their optimization model to determine the time-dependent locations of portable Bluetooth sensors. But the benefit loss during relocation process was not incorporated. This dissertation will overcome this issue when modeling the monitoring benefit during a given time horizon. In other words, a more realistic operational model will be provided when dynamically dealing with portable sensor location determination.

Moreover, in terms of the travel time provision based sensor location optimization problem, the spatial correlation of the monitoring benefit was ignored or considered as trivial by all the existing studies. However, this is unreasonable and may largely affect the optimal solution since travel time information of the segments are always highly correlated site to site. This means measuring the travel time of a particular segment cannot only benefit this segment itself but can also improve the knowledge of the travel time information of its adjacent or overlapped segments. Thus, in terms of real-time travel time collection and provision, the optimization benefit of a sensor location problem will be comprehensively studied and properly defined in this dissertation by considering both temporal and spatial traffic information (e.g., travel time, travel speed, and level of service) characteristics.

In brief, the optimization models and the monitoring benefit quantification schemes developed in this dissertation will supplement the existing works in two ways.

(1) A realistic operational framework in addition to traditional static planning models in dealing with location-and-relocation problems with respect to portable AVI sensors is provided.

(2) Spatial correlation characteristics of traffic state information are explicitly incorporated into the benefit calculation.

Table 2: Summary of Selected Milestone Studies

<b>Author (Year)</b>	<b>Sensor Type</b>	<b>Objective</b>	<b>Application Type</b>	<b>Other Notes</b>
<b>Chan (2002)</b>	Fixed Point Detector	Travel Time Error	Two-stage Planning	Regression-based Benefit (w.r.t. sensor density)
<b>Sherali (2006)</b>	Fixed Path Sensor (AVI)	Coverage Benefit	Static Planning	Variance-based Benefit
<b>Mirchandani (2008)</b>	Fixed Path Sensor (AVI)	Flow Coverage & Travel Time Reliability	Static Planning	Bayesian Update based Benefit
<b>Danczyk (2011)</b>	Fixed Point Sensor	Travel Time Error	Static Planning	Empirical Error
<b>Kianfar (2011)</b>	Fixed Point Sensor	Travel Time Error	Static Planning	Empirical Error Segment Clustering
<b>Li (2012)</b>	General Sensor (Pair up)	Assumed Coverage Benefit	Static Planning	Sensor Failure
<b>Gentili (2012)</b>	A Review: “Locating Sensors on Traffic Networks: Models, Challenges, and Research Opportunities”			
<b>Asudegi (2013)</b>	Portable Path Sensor (AVI)	Flow Coverage & Variance Reduction	Static Planning	Variance-based Benefit
<b>Xing (2013)</b>	Heterogeneous Sensors	Travel Time Error	Static Planning	Estimation by Kalman Filtering
<b>Fei (2013)</b>	Fixed Point Sensor	O-D Estimation	Two-Stage Planning	Flow Equilibrium w.r.t. Incident
<b>Zhu (2014)</b>	Mobile Sensor	Traffic Information Acquisition	Dynamic Operation	Sampling-based Benefit
<b>Bianco (2014)</b>	Fixed Point Sensor	Arc Flow Observability	Static Planning	(First) Mathematical Formulation
<b>Park (2015)</b>	Portable Path Sensor (AVI)	Travel Time Error	Two-Stage Planning	Regression-based Benefit

## Chapter 3 **Real-Time Traffic Surveillance Efficiency Modeling:** **A Stochastic System Perspective**

Real-time traffic-related information such as travel time, traffic speed and traffic volume, is of significant importance to advanced traveler information system (ATIS). Effective surveillance on such traffic state parameters can provide travelers with the accurate knowledge to plan their trips. To obtain such information of a given transportation network in a real-time manner, cost-effective deployment of traffic sensors plays a key role. Specifically, corridors or highway segments with high travel time fluctuations and traffic volumes should be given more emphasis when deciding to place traffic sensors. Deploying sensors on such locations can assist to improve travel time prediction accuracy. Consequently, travelers can make better decisions to improve their traveling experience, such as re-routing or re-scheduling their predefined trips. Since traffic state on any highway segment is time-dependent, the evolution process can be considered as a stochastic process. In fact, the stochastic process is highly facility and location specific. Therefore, appropriate quantification methods should be used to figure out locations, on which the deployment of traffic sensors can introduce significant monitoring benefit to the transportation system. This chapter is developed to answer the monitoring benefit-related questions posted in previous two chapters. Models in terms of quantifying surveillance benefit of a specific highway location are proposed and discussed in following sections.

### *3.1 Variance and Covariance Based Approach*

Travel time is a key measurement reflecting traffic conditions of a highway segment or corridor. For a given corridor, higher travel time usually indicates the occurrence of traffic congestion, while lower travel time means a relatively smooth traffic condition. Moreover, since travel time is the most direct metric to understand people's trip length, it is more commonly accepted and used

as a dynamic message to travelers. In addition to travel time, another important traffic condition metric is average traffic speed. From a system management perspective, both travel time and traffic speed data can be used to infer the traffic congestion level of a particular highway facility. However, average traffic speed is not commonly used as a real-time message within an Advanced Traveler Information System (ATIS), although it is convertible with travel time in most scenarios. In other words, from users' perspective, travel time information is preferable in planning and understanding their trips.

Quantifying site-specific sensor deployment benefit by historical travel time variance is a conventional approach. In existing literature, the objective functions of many traffic sensor location optimization models were formulated with travel time variance as the benefit coefficients (Sherali, Desai, and Rakha 2006; Asudegi and Haghani 2013). An appropriate interpretation for choosing travel time variance as deployment benefit is that placing sensors on locations with high travel time fluctuations can timely collect and capture the traffic dynamics both for offline usage (e.g., traffic dynamics analysis) and online purpose (e.g., notifying users their real-time travel time). Instead, there is no need to place sensors on segments with small travel time variances, since the traffic states at those locations are more likely to be stable. As a consequence, traveling through such segments are reliable.

Travel time variance of a specific highway segment  $l$  during a given time horizon  $H$  can be calculated by Equation (3-1).

$$Var(T_l^H) = E \left[ (T_l^H - \bar{T}_l)^2 \right] = \frac{1}{N} \sum_{i=1}^N [T_l^H(i) - \bar{T}_l] \quad (3-1)$$

where,  $T_l^H$  is a random variable representing the travel time of segment  $l$  within time horizon  $H$ , and  $\bar{T}_l$  denotes the average travel time. The variance of the travel time variable can be empirically estimated with a given number of historical travel time samples. In the above formula,  $T_l^H(i)$  represents the  $i^{th}$  travel time sample, and  $N$  denotes the total sample size.

Travel time variance and standard deviation are two measurements indicating the magnitude of traffic state fluctuation with a unit of second<sup>2</sup> and second, respectively. The magnitude level of travel time is highly location specific. For example, the travel time variance of a shorter segment might be higher than that of a longer segment. But this does not necessarily mean the traffic in the shorter segment suffers more fluctuations than that of, the longer one since the average travel time of the shorter segment is much smaller. The coefficient of variation (CV) is an effective indicator to quantify travel time variability based on the detected samples. CV is defined as the ratio of the standard deviation to the mean (as is given by Equation (3-2)) and is considered a normalized measure of dispersion of a probability distribution.

$$CV_i^H = \frac{\sqrt{\text{Var}(T_i^H)}}{\bar{T}_i} \quad (3-2)$$

where,  $\bar{T}_i$  denotes the mean travel time of segment  $i$  during time horizon  $H$ , and  $\text{Var}(T_i^H)$  is the travel time variance given by Equation (3-1). Since CV is a normalized measure of the dispersion magnitude of a random variable, segments with higher travel time CV are likely to suffer more travel time fluctuations than segments with lower travel time CV. Therefore, CV can be used as a reference to rank the travel time monitoring priority of the segments within a given network, in order to capture the most travel time variations across the entire network.

As is mentioned in previous chapters, placing a sensor at a specific location cannot only bring monitoring benefits for this location itself but can also bring monitoring benefits for other locations to some extent. This can be explained by the existence of traffic pattern correlations between any two geographically close facilities. For example, considering a freeway segment is monitored by a pair of AVI sensors and the travel time information of this segment is obtained in a real-time manner, then the travel time information of its upstream segment can be inferred or predicted with a high confidence level given there exists a highly correlated travel time pattern between these two adjacent segments. This type of additional monitoring benefit should also be considered when deciding the sensor placement locations, given the objective is to capture the

traffic state variance of the entire network maximally. Covariance of historical travel time data is the most direct estimator reflecting the spatial traffic state linear relationship between any two highway segments. Pearson correlation coefficient is a normalized indicator and describes the linear dependence of two random variables, taking values between -1 and +1. For travel time correlation between segment  $l$  and  $k$ , this coefficient can be empirically calculated by Equation (3-3).

$$Cor(T_l^H, T_k^H) = \frac{\sum_{i=1}^N [T_l^H(i) - \overline{T_l^H}] [T_k^H(i) - \overline{T_k^H}]}{N \cdot \sigma_l^H \cdot \sigma_k^H} \quad (3-3)$$

where,  $T_l^H$  and  $T_k^H$  are random variables representing the true travel time value of segment  $l$  and segment  $k$  within time horizon  $H$ , respectively.  $T_l^H(i)$  and  $T_k^H(i)$  denote historical travel time samples of time interval  $i$  from segment  $l$  and  $k$ , respectively.  $N$  is the size of the travel time sample pairs collected to estimate the correlation coefficient.  $\overline{T_l^H}$  and  $\overline{T_k^H}$  are the sample means, and  $\hat{\sigma}_l$  and  $\hat{\sigma}_k$  are the sample standard deviations of  $T_l^H$  and  $T_k^H$ . As is noted, the Pearson correlation coefficient can be obtained only if both of the sample standard deviations are finite and nonzero. Considering travel time on any highway segment is finite and none constant, the Pearson correlation coefficient is valid for travel time values between any two segments. Further, the absolute magnitude of this coefficient (i.e.  $|Cor(T_l^H, T_k^H)|$ ), can be used to evaluate the spatial monitoring benefit to location  $k$  resulted from a sensor placement at location  $l$  given that location  $k$  is out of travel time surveillance, and vice versa.

There is a disadvantage of using Pearson correlation coefficient to quantify the spatial benefit resulted from sensor deployment. The above correlation coefficient can only indicate the degree of linear dependence between travel time values of two segments. Specifically, this is a good metric to evaluate the aforementioned spatial benefit given that a linear formula, i.e.  $T_k^H = a \cdot T_l^H + b + \epsilon$  can represent the travel time relationship of two segments. If the underlying

relationship of spatial travel time follows this formula, the Pearson correlation coefficient can directly give the magnitude of spatial dependency. Consequently, the higher this coefficient is, the higher the additional monitoring benefit will be if a sensor is deployed at either of those two locations. Because, with a higher correlation coefficient, the travel time on one segment can be more accurately predicted by the travel time on the other segment by using the above linear formula with estimated parameters  $a$  and  $b$ . However, the linearity relationship cannot generally describe the spatial travel time patterns existing in a real-world transportation network. Instead, in most cases, there exists various nonlinear relationships between travel times values of two different segments correlated with each other. Therefore, using Pearson correlation coefficient as the spatial sensor deployment benefit might not always be valid.

### 3.2 Travel Time Prediction Accuracy Improvement

#### 3.2.1 Travel Time Prediction Necessity

The general definition of segment travel time at a particular time point  $t_i$  is the duration that a vehicle spends to get through the segment given the entrance time equals to  $t_i$ . Based on this definition, the travel time of a vehicle detected by AVI traffic sensors can be calculated as the time difference between the second detection time point and the first detection time point (as is illustrated by Figure 4). In this example, travel time of the vehicle entering at time point  $t_i$  is  $t_{i+n} - t_i$ . As is noted, no matter how timely the collected travel time data is transmitted to the central data processing server, there is always a time lag reporting the real-time segment travel time. In other words, with the deployed AVI sensors, the segmental travel time at the time  $t_i$  is known at least  $(t_{i+n} - t_i)$  time units later. Therefore, to make the travelers entering at time point  $t_i$  understand their expected travel time through the segment, prediction techniques must be used to predict the travel time at time  $t_i$  based on the historical data collected by  $t_i$ .

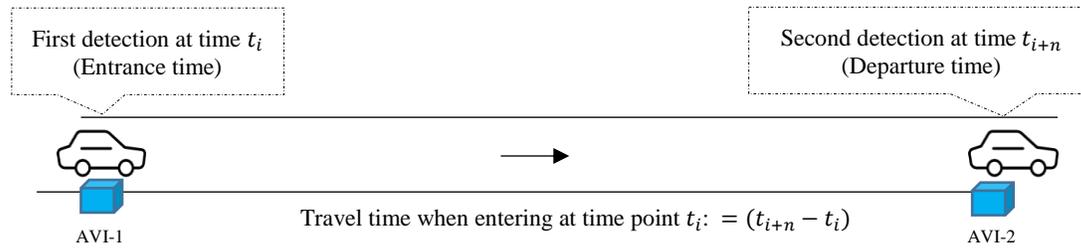


Figure 4: Illustration of Segment Travel Time Measurement by AVI Sensors

Reliable travel time prediction technique is of fundamental importance to any Advanced Traveler Information System (ATIS) or online navigation system. In those systems, it is necessary to calculate the future travel times of a particular path at different time points to recommend the shortest paths to the travelers and assist them in planning their departure times. Thus, travel time prediction models are required not only for short-term prediction purpose (e.g., 1-minute and 5-minute intervals) but are also useful for mid-term (e.g., 30-minute interval) prediction applications. Therefore, improving the network-level travel time prediction accuracy should be considered as an important objective to deploy travel time collection sensors.

### 3.2.2 Travel Time Prediction Error without Real-time Surveillance

The travel time prediction accuracy or error should be investigated following a before-and-after analysis to figure out the traffic surveillance benefit in terms of travel time prediction. Specifically, travel time prediction errors with and without traffic surveillance should be both known and used to estimate the prediction accuracy improvement.

For a specific roadway segment  $l$  without any traffic surveillance, the travel time during time period  $H$  on this segment can only be predicted based on its historical travel time distribution of the same period (e.g., time of the day, the day of the week, and peak or nonpeak). This approach can be deemed as a random guess process, but with probabilistic inferences from the history. If there are no historical travel time observations, then this prediction approach turns to be a completely random guess process. In this case, the most rational way is to predict the travel time by dividing the segment length with the posted speed limit of this segment.

Since travel time is affected by many random factors, such as recurrent congestion, nonrecurring traffic incidents, and weather conditions, it may vary a lot even for the same location at the same time of the day. If we consider the mean travel time of a particular segment during a given time period as a random variable, then it can be assumed to follow a statistical distribution. Without any real-time surveillance, the segmental travel time can be predicted by using the mean value of the underlying population of the travel time random variable. This is a rational and conservative method since the expected prediction error is minimized. In practice, since the mean value and the standard deviation of the population of interest (i.e., the segmental travel time of a given period) are both unknown, the population mean can be assumed to follow a Student's  $t$ -distribution. Considering there are  $N$  historical travel time observations  $T_l^H(i)$  from segment  $l$  during time period  $H$ , a  $C$  level confidence band of the mean travel time can be calculated as  $\widehat{T}_l^H \pm t_{(1-C)/2}^{N-1} \cdot \frac{\widehat{\sigma}_l^H}{\sqrt{N}}$ , where  $\widehat{T}_l^H$  is the estimated sample mean,  $\widehat{\sigma}_l^H$  denotes the estimated sample standard deviation given by Equation (3-4), and  $t_{(1-C)/2}^{N-1}$  is the upper  $\frac{1-C}{2}$  critical value for the  $t$  distribution with  $(N-1)$  degrees of freedom.

$$\sigma_l^H{}^2 = \frac{1}{N-1} \sum_{i=1}^N \left[ T_l^H(i) - \widehat{T}_l^H \right]^2 \quad (3-4)$$

Figure 5 graphically demonstrates the  $C$ -level mean travel time confidence band generation with a Student's  $t$ -distribution based on the estimated sample mean and sample standard deviation. Statistically speaking, the mean travel time value of segment  $l$  at period  $H$ , falls in the dashed area with a probability equaling to  $C$ . Correspondingly, at confidence level  $C$ , the upper bound  $T_{l,U}^H$  and lower bound  $T_{l,L}^H$  of the true travel time can be given by Equation (3-5) and (3-6), respectively.

$$T_{l,U}^H = \widehat{T}_l^H + t_{(1-C)/2}^{N-1} \cdot \frac{\widehat{\sigma}_l^H}{\sqrt{N}} \quad (3-5)$$

$$T_{i,L}^H = T_i^H - t_{(1-C)/2}^{N-1} \cdot \frac{\sigma_i^H}{\sqrt{N}} \quad (3-6)$$

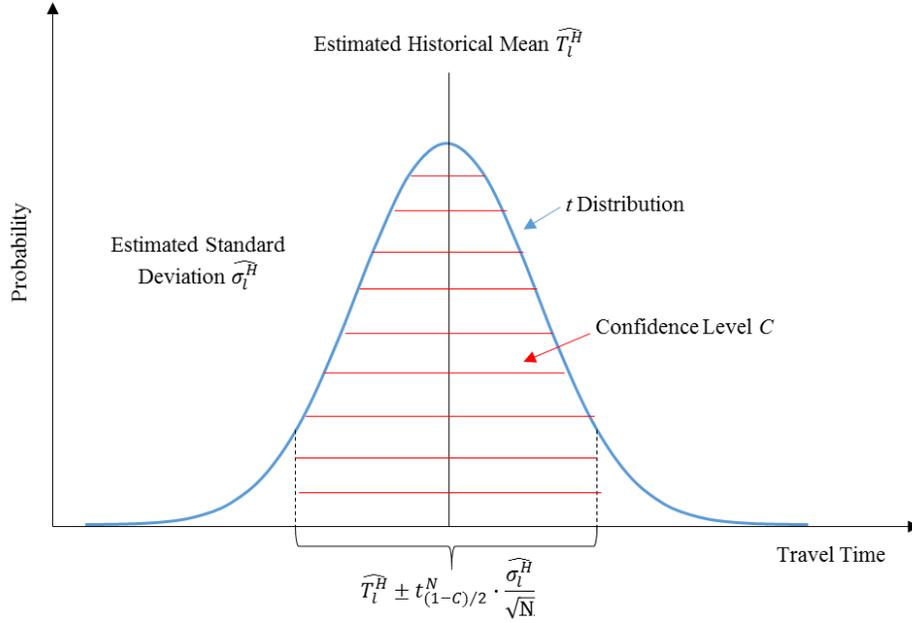


Figure 5: Mean Travel Time Estimation with Historical Observations Based on t-Student Distribution.

Further, the maximum error of a predicted travel time value  $T_i^H(x)$  can be obtained by comparing the deviations against its lower and upper bound by Equation (3-7). Considering the symmetric property given by the sample mean value, as well as the lower and upper bounds, the maximum prediction error is minimized with  $T_i^H(x) = \widehat{T}_i^H$ .

$$\mathcal{E}_{i,\max}^H = \max \left\{ \left| T_i^H(x) - T_{i,L}^H \right|, \left| T_i^H(x) - T_{i,U}^H \right| \right\} \quad (3-7)$$

Therefore, in case that the travel time of a specific segment is out of real-time surveillance at a given time period, the historical sample means can be referred to as the predicted travel time with the minimal expected prediction error equaling to  $T_{i,U}^H - \widehat{T}_i^H$  (or  $\widehat{T}_i^H - T_{i,U}^H$ ).

### 3.2.3 Travel Time Prediction Error with Real-time Surveillance

The main advantage of travel time prediction with real-time surveillance is the introduction of the travel time information of the most recent time periods. Compared against the time-dependent

historical mean value based prediction method discussed in the last section, travel time prediction accuracy for a particular roadway with the knowledge of most recent travel time information can be significantly improved. In other words, travel time prediction with most recent traffic information always outperforms the prediction method that is only based on historical observations far away from the current time point. This has been proved by existing literature in which various advanced data-driven based prediction models, and traffic flow based prediction models were developed. The success of those models in travel time prediction relies on one basic assumption that traffic state evolves following some underlying patterns which can be figured out by historical data. For a given roadway, if there exists a significant travel time evolution pattern and this evolution process is known in advance, then the short-term and mid-term future travel time on this roadway can be accurately predicted with the knowledge of the travel time in past time intervals. The core target of thousands of existing studies in travel time prediction field was to estimate and mine the underlying traffic state or travel time evolution patterns to predict the future travel time more accurately. Successful travel time prediction methods in existing literature include parametric based methods (e.g., ARIMA model), nonparametric based methods (e.g., Neural Network), hybrid methods (e.g., Traffic Dynamic Models), and ensemble methods (e.g., Random Forest Model). For a comprehensive view and comparison of those advanced travel time prediction methodologies, interested readers are referred to the latest dissertations in terms of travel time prediction by Sadabadi (2014), and Zhang (2015), and the state-of-art literature review by Vlahogianni, Karlaftis, and Golias (2014).

There is no definite conclusion which mathematical model is the best for travel time prediction. Each prediction model has its unique strengths and weakness since the stochastic process of travel time evolution is highly facility and environment specific. Thus, for a particular roadway, the best prediction model can only be selected among all the optional models after a comprehensive evaluation by using the data belonging to this facility. Here, we use  $f_t^H(\cdot)$  to generally denote a prediction model. Given the real-time travel time information is available (i.e.

the traffic is in real-time surveillance), the travel time of a specific roadway  $l$  at a future time interval  $h$  can be calculated by the preselected prediction model with two types of input data. They are, historical travel time information and real-time travel time information. Historical travel time data is indispensable for the future prediction, since it is used to estimate the parameters and structure of the prediction model. After the prediction model is all set, real-time travel time information is input to the model and the future value of interest can be calculated. This procedure applies to all kinds of travel time prediction models, and can be represented by the following Equation (3-8).

$$T_l(h) = f_l^H(\overrightarrow{T}_l^R, \overrightarrow{T}_l^H) \quad (3-8)$$

where,  $\overrightarrow{T}_l(h)$  is the predicted travel time on segment  $l$  at time interval  $h$ .  $\overrightarrow{T}_l^R$  and  $\overrightarrow{T}_l^H$  are the real-time and historical travel time information as inputs of the prediction model. Here, the superscript  $H$  is a time or environment classifier, representing category information such as time of the day, day of week and a specific weather condition. This classifier might be omitted, if the prediction model is uniquely selected and estimated with the overall historical data without classification. However, even for the same location, the travel time evolution pattern might change in different conditions. This is intrinsically determined by the change of travelers' behaviors in different traffic or weather conditions. Therefore, developing time-dependent or environment-dependent prediction models is desirable since it can capture the traffic patterns in a higher resolution.

Similarly, future travel time on a segment  $l$  can also be predicted by using the real-time travel time information from another segment  $k$ , given there exist travel time correlations between these two segments. Equation (3-9) can describe this spatial based prediction process.

$$T_{k,l}(h) = f_{k,l}^H(\overrightarrow{T}_k^R, \overrightarrow{T}_l^H, \overrightarrow{T}_k^H) \quad (3-10)$$

where,  $\overrightarrow{T}_l^H$  and  $\overrightarrow{T}_k^H$  denote the historical travel time observed on segment  $l$  and  $k$ , respectively.

These two sets of historical data are used to estimate the spatial travel time prediction model

$f_{k,l}^H(\cdot)$  with the real-time data stream  $\overrightarrow{T_k^R}$  from segment  $k$  as input and the predicted travel time value  $\widehat{T_{k,l}}(h)$  on segment  $l$  as output. It is noted that this type of spatial traffic prediction approach is useful only if there exists an underlying pattern between the traffic information of two different segments (e.g. upstream segment vs downstream segment).

Consequently, the absolute prediction errors by those two types of real-time surveillance (i.e., direct surveillance and indirect surveillance) based prediction approaches can be calculated by Equations (3-11) and (3-12).

$$\varepsilon_l(h) = |T_l(h) - \widehat{T_l}(h)| \quad (3-11)$$

$$\varepsilon_{k,l}(h) = |T_{k,l}(h) - T_l(h)| \quad (3-12)$$

where,  $T_l(h)$  is the true value of the travel time on segment  $l$  at time index  $h$ .

It should be noted that true value of the travel time can only be known after the data of that time index has been collected. Therefore, the future forecast error of a particular model can only be evaluated and statistically given based on the historical prediction performance. Root-mean-square error (RMSE) and mean-absolute-percentage error (MAPE) are two statistical measures widely used to evaluate the prediction accuracy of a forecasting model. The formulas for calculating RMSE and MAPE are given by Equation (3-13) and (3-14).

$$RMSE = \sqrt{\frac{\sum_{h=1}^N \varepsilon^2(h)}{N}} \quad (3-13)$$

$$MAPE = \frac{1}{N} \sum_{h=1}^N \frac{\varepsilon(h)}{T(h)} \quad (3-14)$$

where,  $\varepsilon(h)$  is the absolute prediction error at time interval  $h$ , which is given by Equation (3-11) or (3-12), and  $N$  is the total number of individual predictions used to test the overall prediction performance.  $T(h)$  denotes the ground truth value of the travel time at time index  $h$ , as can be collected after the time index passed by. RMSE is also known as root-mean-square deviation

(RMSD), which measures the average prediction deviation from the ground truth with unit equaling to the target prediction variable. MAPE is also known as mean-absolute-percentage deviation (MAPE), which expresses the prediction accuracy as a percentage value. Statistically speaking, for a specific prediction model, the higher the RMSE and MAPE values are, the lower prediction accuracy the model has. Thus, RMSE and MAPE calculated based on historical travel time dataset from a particular location with a specific forecast model can be used as estimators for the expected prediction performance of future scenarios by applying this model at this location.

### 3.2.4 Surveillance Benefit Based on Travel Time Prediction Error Reduction

Traffic surveillance benefit in terms of travel time prediction can be calculated as the difference of travel time prediction error with surveillance and the prediction error without surveillance. Mathematically, the prediction error based temporal and spatial traffic surveillance benefits can be expressed by the following Equations (3-15) and (3-16), respectively.

$$B_l^H = \varepsilon_l^{H,+} - \varepsilon_l^{H,-} \quad (3-15)$$

$$B_{k,l}^H = \varepsilon_{k,l}^{H,+} - \varepsilon_{k,l}^{H,-} \quad (3-16)$$

where,  $\varepsilon_l^{H,+}$  and  $\varepsilon_l^{H,-}$  represent the expected travel time prediction error for segment  $l$  in time horizon  $H$  on condition that  $l$  is with real-time surveillance and without real-time surveillance, respectively.  $\varepsilon_{k,l}^{H,+}$  denotes the expected spatial travel time prediction error for segment  $l$  by using the real-time information from segment  $k$ , and  $\varepsilon_{k,l}^{H,-}$  denotes the expected travel time prediction error on condition that its travel time can only be predicted by the random guess approach. The surveillance benefit derived above can be explained as the travel time prediction error reduction. The concept is similar with the before-and-after analysis. The decision of whether deploying a real-time surveillance system of a particular roadway highly depends on the impact of the deployment. Here, the deployment impact is calculated as the prediction error reduction. The higher the prediction error is reduced, the more preferable the deployment is.

In terms of traffic information prediction, it is irrational to make the network-level traffic sensor deployment decisions without the consideration of the prediction errors for those segments which are without surveillance. In existing traffic sensor location optimization studies with the objective of minimizing total travel time estimation or prediction error, many studies ignored this consideration. Instead, they simply set their goals as minimizing the total prediction errors for locations with sensor deployed (Kianfar and Edara 2010; X. Li and Ouyang 2012). This can be further illustrated by the example demonstrated in Figure 6. For two highway segments, labeled as Seg-1 and Seg-2, the travel time prediction errors with real-time data stream available are 7% and 10% respectively expressed as MAPE, and 30 seconds and 50 seconds respectively expressed as RMSE. If we consider to deploy a real-time surveillance system on either of these two segment and set the objective as minimizing the total real-time prediction error, the most suitable location is Seg-1, since the real-time prediction error on Seg-2 is larger. In fact, if we set the objective as maximizing the total prediction error reduction, the preferable location is Seg-2 because the network-level MAPE and RMSE can be reduced by 50% and 100 seconds if real-time surveillance is provided for Seg-2. This can also be understood from another perspective. In most scenarios, the real-time travel time prediction errors in different segments are all relatively small. In other words, the prediction errors are not significantly different (e.g., 7% vs. 10%). However, the prediction errors (by random guess) varies a lot for segments that are all without real-time surveillance (e.g., 60% vs. 40%). This large variation is determined by the significant difference in the traffic fluctuation levels across different segments. For example, the travel time of a bottleneck segment is not easily predicted simply by using its historical mean value, since the historical variance is high. While, for a segment with small traffic demand all times, the future travel time can be easily predicted even without real-time surveillance. Therefore, considering the deployment benefit as maximizing the prediction error reduction instead of minimizing real-time error is saying to place the surveillance system at locations where it is truly needed.



Figure 6: Illustrative Examples of Travel Time Prediction Error Reduction: (a) MAPE Based Error Reduction; (b) RMSE Based Error Reduction

Moreover, different from the existing studies with respect to traffic sensor location optimization, the spatial travel time prediction error is additionally considered in this research. This is because the traffic on highway segments belonging to a regional network are not completely independent of each other. Hence, compared against the traffic prediction performance by simply referring to one's historical data, the prediction performance by taking advantage of other's real-time traffic information might be better. In other words, the benefit improvement of a particular segment without sensor deployment decision does not necessarily stay zero, considering its spatially-correlated segments are under surveillance.

### 3.3 Entropy-Based Model to Quantify Traffic Condition Prediction Uncertainty

#### 3.3.1 Background and Introduction

A stochastic system is the one in which the value of parameters (i.e., system states), measurements, or disturbances are evolving with uncertainty. In common, the state or measurement of a particular stochastic system is evolving with time and realized based on some random probabilities or patterns that can be analyzed statistically but might not be predicted precisely. Based on the fundamental study in stochastic mechanics given by Nelson (1985), any system or process is claimed at least partially stochastic if this system or process must be analyzed using probability theory.

Studying and modeling the uncertainty of a specific system is an active research area in multiple disciplines, including science, economics, and engineering. For example, Hung and Ma (2009) systematically analyzed the uncertainties involved in the life cycle impact assessment procedures and investigated the uncertainty of environmental performance for individual impact categories, such as global warming, human health, and ecotoxicity. Variations of human behavioral principles, such as regret and cognitive dissonance, can largely determine the efficiency of a particular financial system. Consequently, the financial system might evolve and perform with various efficiency uncertainty (Shiller 1999). Specifically, many detailed factors affecting the uncertain measurement of a stock market were investigated by the work of Veronesi (1999), and Connolly, Stivers, and Sun (2005). Examples in the engineering field are such as, system identification with parameter estimation or model updating (Friswell and Mottershead 1995), measurement system configuration considering system uncertainty (Papadimitriou, Beck, and Au 2000; Robert-Nicoud, Raphael, and Smith 2005), and prediction of infrastructure system parameter uncertainty (Zhang 2015).

In the field of transportation system information prediction, great efforts have been made in past decades. Researchers have designed and applied a lot of effective models for traffic information prediction, such as traditional statistical methods (Rice and Van Zwet 2004; Fei, Lu, and Liu 2011), simulation-based methods (Ben-Akiva et al. 1998; Hu et al. 2012), support vector machines (Wu, Ho, and Lee 2004; Castro-Neto et al. 2009), and various neural network models (Smith and Demetsky 1994; Zheng, Lee, and Shi 2006; Lv et al. 2015). The common objective of such advanced models is to accurately predict a point value for a particular traffic state variable such as travel time and traffic flow. As is pointed out by a comprehensive literature review by Vlahogianni, Karlaftis, and Golias (2014), traffic condition is a complex phenomenon and is often affected by many exogenous factors such as unexpected traffic incidents and non-recurrent abnormal weather conditions. Therefore, traffic condition prediction is associated with different

levels of uncertainties due to the stochastic nature of the transportation system. There are still limited studies focusing on the uncertainty of such predictions in terms of traffic state prediction.

In existing studies of traffic prediction uncertainty modelling, prediction interval (PI) is adopted as an uncertainty measurement and aims to provide more information describing the uncertainty associated with point predictions. For example, Khosravi et al. (2011) proposed several approaches to estimate the travel time prediction intervals based on neural network model. In terms of neural network based travel time prediction, Zeng and Zhang (2013) developed an ensemble-based method to estimate the prediction bands under different initial conditions. Bayesian based models are applied as well to provide such prediction intervals due to the probabilistic properties of Bayes inference (van Hinsbergen, Van Lint, and Van Zuylen 2009; Fei, Lu, and Liu 2011). From the perspective of statistical volatility, Zhang, Haghani, and Zeng (2015) developed a component GARCH model to provide travel time series uncertainty dynamics accounting for seasonal patterns. All the above efforts aim to provide confidence bands for traffic state predictions based on a specific prediction model. The common assumption is that prediction model should be specified in advance. In other words, the aforementioned prediction interval is very model dependent. From the systematic perspective, traffic state prediction uncertainty of a given highway system is intrinsically determined by two factors, i.e. system stochasticity level and effectiveness of the measurement space. To better understand the system state prediction uncertainty, one should jointly consider these two points and more systematically evaluate the underlying prediction uncertainty.

In fact, most of the systems with multiple states evolution process, in reality, can be classified as a stochastic system or at least a partial stochastic system. This is intrinsic because the operation of a system may be affected by many internal and external random factors. Specifically, the system state may not be evolving deterministically due to nondeterministic impacts, such as random noise and external forces. From the measurement aspect, the measurable variables of a particular system cannot be guaranteed to be error-free. Moreover, nonrecurring

and random events might introduce an unexpected disturbance to the system. As a consequence, the system evolution process is prone to be uncertain, and cannot be estimated and predicted precisely. An obvious example is the traffic condition evolution process of a specific transportation system, given random and nonrecurring traffic incidents (e.g., accidents and road maintenance operations) and recurring O-D demand pattern but with random fluctuations. Those random factors make the traffic condition evolve with uncertainty. More specifically, both estimation and prediction of the traffic condition (e.g., travel time and congestion level) cannot be guaranteed error free. In general, the more uncertain the system evolution process is, the more difficult the system status can be predicted (accurately).

### 3.3.2 System State Uncertainty

A better approach to measure and quantify the uncertainty of a random variable is the entropy as defined by Shannon and Weaver (1949). The entropy used to quantify the occurrence uncertainty (or predictability), of a specific event  $\omega$ , is defined as Equation (3-17).

$$H(\omega) = -\left[ p(\omega) \cdot \log_2(p(\omega)) + (1 - p(\omega)) \cdot \log_2((1 - p(\omega))) \right] \quad (3-17)$$

where  $p(\omega)$  denotes the happening probability of state (or event)  $\omega_i$ , consequently  $(1 - p(\omega))$  represents the probability that the system is not in state  $\omega$ . The basic goal of Shannon entropy is to quantitatively evaluate the occurrence uncertainty of a single event or a list of events. As is demonstrated in Figure 7, for a specific event, when the occurrence probability is equal to the nonoccurrence probability (i.e. 0.5), the entropy has the maximum value. Instead, if the probability is equal to one or zero, the entropy is zero. In other words, if the occurrence probability of an event is 0.5, it indicates people have zero knowledge on whether this event will happen. Consequently, the uncertainty is largest. On the other hand, if the probability is one or zero, people can conclude on the occurrence of the event without any uncertainty. As a result, the uncertainty is smallest (i.e., zero).

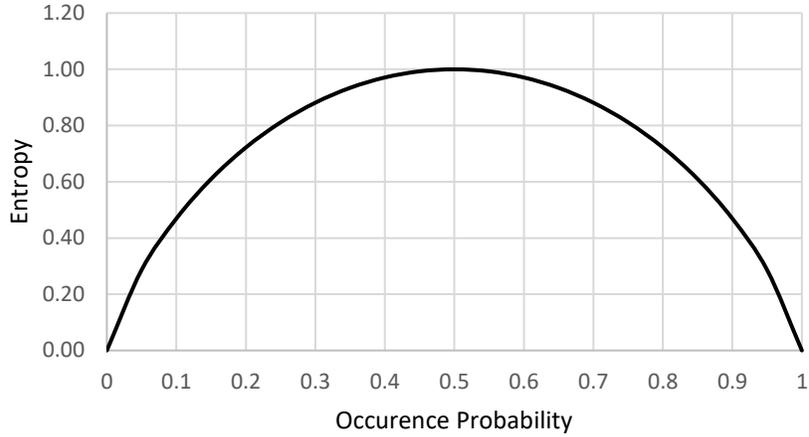


Figure 7: Entropy of A Single Event with respect to the Occurrence Probability

In general, for a particular system with discrete state variable  $\omega_i$  evolving in a stochastic process, the entropy of the system state is defined as Equation (3-18).

$$H(\omega) = -\sum_{i=1}^N p(\omega_i) \cdot \log_2 [p(\omega_i)] \quad (3-18)$$

where  $p(\omega_i)$  is the probability that the system state is  $\omega_i$ , and  $N$  denotes the total number of possible statuses in which  $\omega_i$  could be. To calculate the state entropy of a given system, all the possible states should be taken into account (i.e.  $\sum_{i=1}^N p(\omega_i) = 1$ ). As is calculated by the above formula, the entropy takes the maximum value when the occurrence probability for each state is equal (i.e.  $p(\omega_i) = p(\omega_j)$  for any  $i$  and  $j$ ). In this situation, people have zero knowledge on the estimation of the current system status without any measurement, or prediction of the system status in next interval. Consequently, the estimation or prediction uncertainty is largest in such scenario. In other cases, the higher the heterogeneity of the system state status distribution is, the smaller the entropy is. Because higher heterogeneity of the state probability distribution indicates a more significant likelihood across the state space.

Moreover, entropy is also applicable to quantify the system state uncertainty with continuous state space. For a given random variable taking values from a continuous space, the entropy can also be derived by using the same concept as in Equation (3-19).

$$H(\omega) = -\int_{\omega} p(\omega) \cdot \log_2 [p(\omega)] d\omega \quad (3-19)$$

where  $p(\omega)$  represents the probability density of  $\omega$ . As is shown, Equation (3-19) is the continuous form of Equation (3-18). However, there might not be a closed-form formula to calculate the above integration with respect to the given probability density function. Lots of studies had been done in terms of calculating the entropy of continuous random variable, but this is out of the scope of this research. Interested readers can refer the work by Rrnyi (1961), Ahmad and Lin (1976), and Beirlant et al. (1997).

### 3.3.3 Temporal System State Prediction Uncertainty

For a stochastic system, it is possible to predict (or estimate) its future (or current) status based on a set of measurements from the system. Using recently collected travel time series to predict future travel time on a given roadway is an example of the temporal transportation system state prediction. However, since both internal and external factors may randomly introduce a disturbance to the system, it is practically impossible to predict the state variables of a system precisely. On the one hand, developing an advanced model to improve prediction accuracy is important. On the other hand, realizing and studying the prediction uncertainty from a general perspective is also of fundamental significance, because the state evolution uncertainty is system specific. For the system with very small state evolution uncertainty, it is usually not necessary and cost-effective to investigate and develop more advanced techniques in terms of the system status monitoring and prediction. Instead, for those systems with high state evolution uncertainty, people should consider devoting more efforts in finding and developing effective surveillance approaches. Hence, we developed an entropy-based model to estimate and quantify the system state prediction uncertainty.

We define the temporal system state prediction as the process in which temporal measurements from the system itself are collected and used to predict the future system status.

Without loss of generality, assume the state of a target monitoring system takes value from set  $\Omega = [\omega_1, \omega_2, \omega_3, \dots, \omega_N]^T$  with  $N$  types of statuses, and the measurement from this system through a specific surveillance approach has possible values belonging to set  $\Phi = [X_1, X_2, X_3, \dots, X_M]^T$ . In reality, the system state will be uniquely estimated or predicted given a specific mathematical model with a particular measurement. However, the predicted value might not be always equal to the true value of the internal state due to some random factors mentioned before. Statistically speaking, the more likely the prediction is precise, the higher predictability the system status is by applying the prediction model. Based on probability theory, the probability of the system state will be  $\omega_i$  given the measurement is  $X_j$  can be expressed by the conditional probability  $p(\omega_i|X_j)$ . We propose the following conditional probability matrix to map and investigate the predictability on state status of system  $l$  given measurements from the surveillance system.

$$P_l = \begin{pmatrix} P_{11} & P_{12} & P_{13} & \cdots & P_{1N} \\ P_{21} & P_{22} & P_{23} & & P_{2N} \\ P_{31} & P_{32} & P_{33} & & P_{3N} \\ \vdots & & & \ddots & \vdots \\ P_{M1} & P_{M2} & P_{M3} & \cdots & P_{MN} \end{pmatrix} \quad (3-20)$$

where,  $p_{ji} = p(\omega_i|X_j)$ , and represents the probability that the true system state is  $\omega_i$  given the measurement from the surveillance system is  $X_j$ . By applying Shannon entropy introduced in previous section, the uncertainty (or unpredictability) of the system state status based on a particular measurement  $X_j$  can be obtained as the entropy conditioning on  $X_j$ , which is expressed by Equation (3-21).

$$H_l[\omega | X_j] = \sum_i p_{ji} \log_2\left(\frac{1}{p_{ji}}\right) \quad (3-21)$$

This conditional entropy quantifies the unpredictability of the system state based on a unique measurement, which is used to estimate the true system state. Zero conditional entropy indicates

the system state can always be precisely predicted under the condition the measurement is  $X_j$ . The larger the conditional entropy is, the higher uncertainty the prediction will be considering measurement  $X_j$ .

Further, given the system is under measurement, the expected system state prediction entropy can be derived by incorporating the probability (i.e., occurrence frequency) of each possible measurement, as shown in Equation (3-22).

$$H_l(\omega|X) = \sum_j \pi_j \sum_i p_{ji} \log_2\left(\frac{1}{p_{ji}}\right) \quad (3-22)$$

where,  $\pi_j$  denotes the probability that measurement is  $X_j$ . This can also be interpreted as the long-term occurrence frequency of  $X_j$  in the surveillance system. Integrating all possible measurement states and system states, the entropy given by Equation (3-22) yields the expected uncertainty on system state prediction given the system is under real-time surveillance and the measurements from the surveillance system is used to conduct the prediction.

In contrast, prediction of the system state without any real-time measurement can only be made based on the historical distribution of the state status. In practice, the system of interest without real-time measurement can be interpreted as there is no real-time surveillance system measuring and providing useful information regarding the system's real-time operation. Consequently, people have zero knowledge of the current status of the system. Prediction uncertainty on the system state without real-time measurement should be derived based on the historical state distribution  $\Pi(\omega) = \{p(\omega_1), p(\omega_2), p(\omega_3), \dots, p(\omega_N)\}$ , considering this is the only knowledge regarding the system of interest. Hence, the system state prediction uncertainty without real-time system surveillance is calculated by Equation (3-18). Further, the state prediction uncertainty reduction of a specific system  $l$  can be obtained as the difference of the prediction entropies with and without real-time measurements, as is expressed in Equation (3-23).

$$\Delta H_l = H_l(\omega|X) - H_l(\omega) \quad (3-23)$$

This entropy-based index can be used to quantify and evaluate the surveillance benefit of any stochastic system to estimate or predict the real-time state of the system. Specifically, the more the entropy is reduced, the higher benefit the surveillance system can bring. Instead, lower entropy reduction indicates there will not be a significant improvement on the state prediction given a surveillance system is installed. There are two possibilities for insignificant entropy reduction even after the introduction of a real-time surveillance system. One situation is that the system state is extremely stable, then there is no need to have an additional surveillance system. The other situation is that the system state is extremely stochastic, as a result, even real-time surveillance is not capable of improving the prediction performance.

#### 3.3.4 Spatial System State Prediction Uncertainty

For a specific stochastic system  $l$ , its system state might be inferred based on the measurements or states of other associated systems. We define the system state prediction process based on external inference from other parallel or correlated systems as the spatial state prediction.

Different from the temporal system state prediction, in which the prediction is made only based on the measurements from the system itself, spatial state prediction depends on the measurements from other systems associated with the target system. Spatial system state prediction is useful in some real-world applications. For example, the travel time of the upstream freeway segment can be estimated based on the travel time data collected in the downstream segment. Another example is the wind speed information at a specific location can be predicted based on the wind parameters measured at a set of weather stations around the target location.

Similarly, spatial measurement-state probability mapping matrix  $P_{k,l}$  can be estimated and generated as below.

$$P_{k,l} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1N} \\ p_{21} & p_{22} & p_{23} & & p_{2N} \\ p_{31} & p_{32} & p_{33} & & p_{3N} \\ \vdots & & & \ddots & \vdots \\ p_{K1} & p_{K2} & p_{K3} & \cdots & p_{KN} \end{pmatrix} \quad (3-24)$$

where,  $l$  represents the target system, the state of which is being predicted, and  $k$  represents another spatial-correlated system, the measurements or states from which are used to predict the state of  $l$ .  $N$  and  $K$  denote the total number of different states  $l$  has, and the total number of different measurements can be obtained from  $k$ , respectively. In the spatial measurement-state probability mapping matrix,  $p_{ji}$  is the probability that the system state in  $l$  is  $\omega_i$  on condition that the measurement from system  $k$  is  $X_j$ . Mathematically,  $p_{ji} = p(\omega = \omega_i | X = X_j)$ .

Therefore, the state prediction uncertainty of system  $l$  given an associated system  $k$  is being measured (or monitored) can be derived from Equation (3-25).

$$H_{k,l}(\omega|X) = \sum_{j=1}^K \pi_j \sum_{i=1}^N p_{ji} \log_2 \left( \frac{1}{p_{ji}} \right) \quad (3-25)$$

where,  $\omega$  denotes the state variable of system  $l$ , and  $X$  represents the measurement variable from system  $k$ .  $p_{ji}$  is the measurement-state transition probability in  $P_{k,l}$ , as shown in Equation (3-24).  $\pi_j$  is the long-term probability (i.e. occurrence frequency) of the  $j^{\text{th}}$  measurement from  $k$ . The entropy calculated by Equation (3-25) quantitatively describes the state prediction (or estimation) uncertainty based on the measurement from an external system as prediction input.

Hence, the prediction uncertainty reduction of  $\omega$  by taking into account of external measurements that are spatially correlated with the target prediction variable can be obtained as Equation (3-26).

$$\Delta H_{k,l} = H_{k,l}(\omega|X) - H_l(\omega) \quad (3-26)$$

In the above notations, we use index  $k$  to denote the external system with associated measurements available for predicting the system state of  $l$ . However, this does not strictly mean

the measurement must be taken from a single external system. In other words,  $k$  may also represent a set of multiple systems that are correlated with  $l$ . In this case, the measurement vector  $X_j$  is the combination multiple measurements from all the spatial-correlated systems. This is highly application specific.

### 3.4 Application to Real-time Travel Time Surveillance and Prediction

#### 3.4.1 Introduction

In this section, we applied the proposed uncertainty estimation model to the transportation system to study and demonstrate the impact of real-time surveillance on travel time prediction performance. A transportation network can be considered as a stochastic system with stochastically evolving traffic states distributed at different locations of the network. Further, each link or segment of the network can be viewed as a subsystem with dynamic and stochastic traffic parameters. Here, we refer to traffic parameter as a random variable representing travel time of a particular highway corridor. Given the existence of various factors (e.g., adverse weather, traffic accident, and stochastic traffic demand), which stochastically happen and affect the transportation system, travel time usually evolves in a stochastic process and cannot always be predicted precisely. Due to the existence of temporal-spatial traffic state patterns within the network, monitoring and collecting real-time travel time information is of fundamental importance to predicting both short-term and mid-term route travel time. Therefore, installing real-time surveillance system (e.g., traffic sensor) for the highway system is necessary for the development of a real-time traffic information provision system (e.g., Advanced Traveler Information System).

#### 3.4.2 State Space Definition

For a highway segment, the evolution of travel time can be modeled as a dynamic process. For a given time window  $M$ , travel time series is represented by a discrete state vector  $S_i = [x_i, x_{i+1}, \dots, x_{i+M}]^T$ . The temporal distance of two state vector  $S_i$  and  $S_j$  is defined as  $d_{ij} = |i -$

$j]$ . The one temporally measured later than the other one is called as the future state vector in terms of the former state vector with step distance equal to  $d_{ij}$ .

Further, to comprehensively describe the travel time pattern of a segment, we define a three-component  $C_i = [Mag_i, Trend_i, Var_i]$  vector indicating the pattern of a given state vector  $S_i$ . The characteristic vector  $C_i$  consists of three different components that quantitatively describes the travel time pattern of a given state vector  $S_i$ . They are, the magnitude level, trend, and variation of the travel time in a specific time window. Each characteristic indicator is calculated based on the state vector  $S_i$  as below.

$$Mag_i = E S_i = \frac{1}{|S_i|} \cdot \sum_{k=i}^{i+M} x_k \quad (3-1-27)$$

$$Trend_i = \sum_{k=i+1}^{i+M} (x_k - x_{k-1}) \quad (3-1-28)$$

$$Var_i = Variance(S_i) = \frac{1}{N} \cdot \sum_{k=i}^{i+M} x_k^2 - Mag_i^2 \quad (3-1-29)$$

We proposed the above characteristic vector consisting of three statistic estimators to comprehensively describe the travel time pattern on a segment for a given period from time stamp  $i$  to  $i + M$ . The first index  $Mag_i$  indicates the average travel time on the segment for the period. The second index  $Trend_i$  describes whether the travel time exhibits a describing trend, an increasing trend or a stable trend. The trend index greater than 0 indicates the travel time exhibits an increasing pattern, and less than 0 indicates the travel time exhibits a decreasing pattern. In some cases with trend index equal to 0, the last parameter  $Var_i$  can be used to evaluate the travel time stability. Specifically, when the travel time is fairly stable (e.g., at night),  $Var_i$  tends to be zero. Otherwise,  $Var_i$  tends to be larger and indicates the travel time of the segment in period  $i$  has high variation. Empirically speaking, in the cases when  $Var_i$  is large, the future travel time is more difficult to predict. On the contrary, when  $Var_i$  is small for travel time series measured recently, future travel time can be easier to predict based on  $Mag_i$  and  $Trend_i$ .

Therefore, in the application of this study, we use the above three-component vector to represent the travel time state of a particular highway segment within the  $L$ -length time window. For a highway segment with real-time surveillance by travel time detection sensors, the state vector is known for any past time periods. In such a case, uncertainty on future travel time pattern can be estimated based on the proposed entropy-based uncertainty model. Similarly, with real-time travel time series that is measured and serves as input, the data-driven based prediction model can be used to forecast the future travel time series.

### 3.4.3 Numerical Experiments

#### 3.4.3.1 Preliminaries and Objective Description

This section utilizes real-world travel time data from a highway network to demonstrate the proposed uncertainty estimation model. As mentioned before, the performance of travel time prediction is highly location specific. To provide a comprehensive view on travel time prediction performance at different locations, we adopted a large-scale highway network for the study. The case study network is Washington D.C.-Baltimore commuting network consisting both major arterial and freeway corridors. As is shown in Figure 8, the network contains 88 directional highway segments.

Three-month travel time data (i.e., May 2016 to July 2016) provided by INRIX is used to conduct the case study. INRIX makes use of a common industry convention known as Traffic Message Channel location code (TMC) to report travel time and travel speed data on the freeway and major arterial roads (INRIX n.d.). In the study network, each network segment consists of multiple TMC segments. We firstly used a backtracking algorithm to concatenate the TMCs travel times to obtain the true experienced travel time for each corridor segment (X. Zhang, Hamedi, and Haghani 2015).

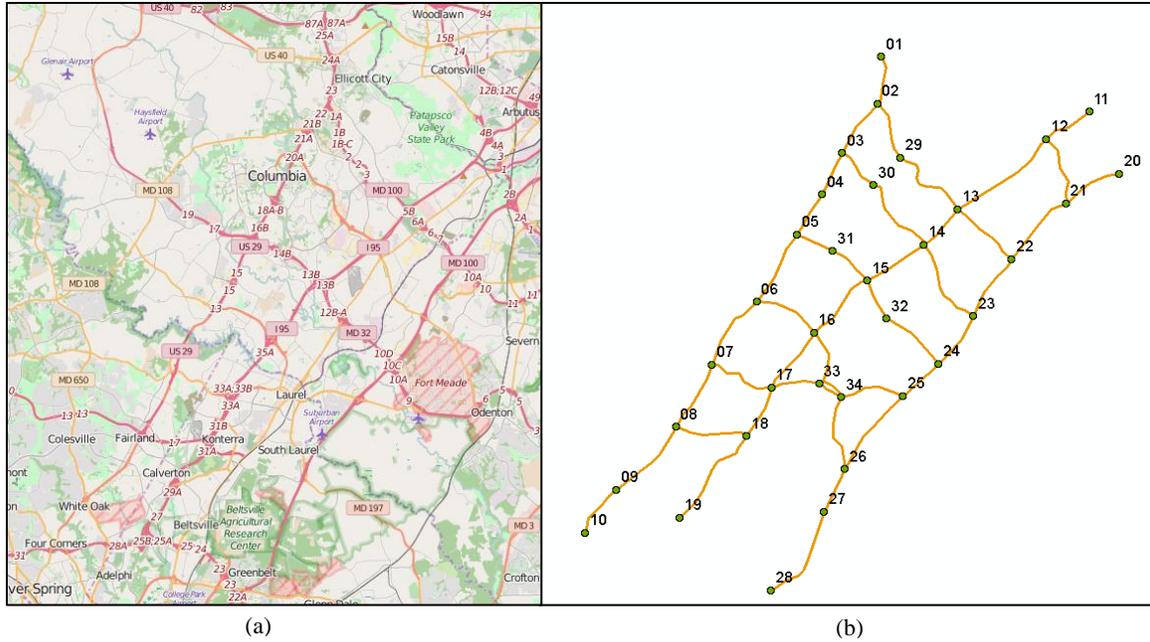


Figure 8: Case Study Network: (a) Map View of the Target Network (i.e., Commuting Network between Washington D.C. and Baltimore); (b) Extracted and Abstracted Network.

In this study, we do not focus on seeking a perfect data-driven prediction model that works for a single corridor segment. There are quite a number of advance prediction models that have been applied to travel time prediction, such as ARIMA models, Neural Network Models and deep learning models (e.g., random forest model). In existing studies, people have paid lots of attention to compare and evaluate those models with data from the same location. It has been shown that for the same dataset the prediction performances by different models are similar if the model parameters are carefully calibrated. However, the question of how datasets can affect the prediction performance has not been answered too well.

In the following numerical experiments, we will focus on investigating the dispersion of travel time prediction performance by a particular prediction model at various locations, and will empirically demonstrate that the prediction performance is very location dependent. Specifically, an advanced data-driven prediction model cannot always be perfect for every highway corridor and sometimes might fail.

Further, we will demonstrate how to evaluate the travel time prediction uncertainty with the proposed stochasticity estimation model. And we will illustrate the prediction uncertainty under real-time measurements and without real-time measurements. Finally, by comparatively evaluating the prediction performance and uncertainty index on each highway segment, we will provide insights into the relationship between the prediction error reduction and the uncertainty reduction given the availability of real-time surveillance.

#### 3.4.3.2 Travel Time Prediction Error Evaluation

For a highway segment with some historical travel time data, one can develop a prediction model for real-time prediction purpose. Specifically, parameters of the model are trained with the historical data and future travel time will be estimated with real-time data feeds as input. Ideally, a perfect prediction model is the one that has zero prediction error. However, perfect travel time prediction cannot happen in a real-world transportation system due to many stochastic factors. On the other hand, prediction performance is supposed to be different at different locations even when the same prediction model is used. The argument is that travel time variation of a particular highway is intrinsically determined by many factors, such as uncertainty level of non-recurrent traffic events (e.g., traffic accident) and predictability of recurrent traffic events (e.g., time-dependent commuting traffic volume).

Based on a preliminary statistical analysis, travel time variation in a particular segment differs in different time periods. This can be illustrated by the CV analysis of the entire network displayed in Figure 9. Moreover, different highway locations exhibit different travel time variations for the same time of the day. Since different locations exhibit different variation patterns, it is necessary to fit a specific prediction model for each segment with the historical travel time data from itself.

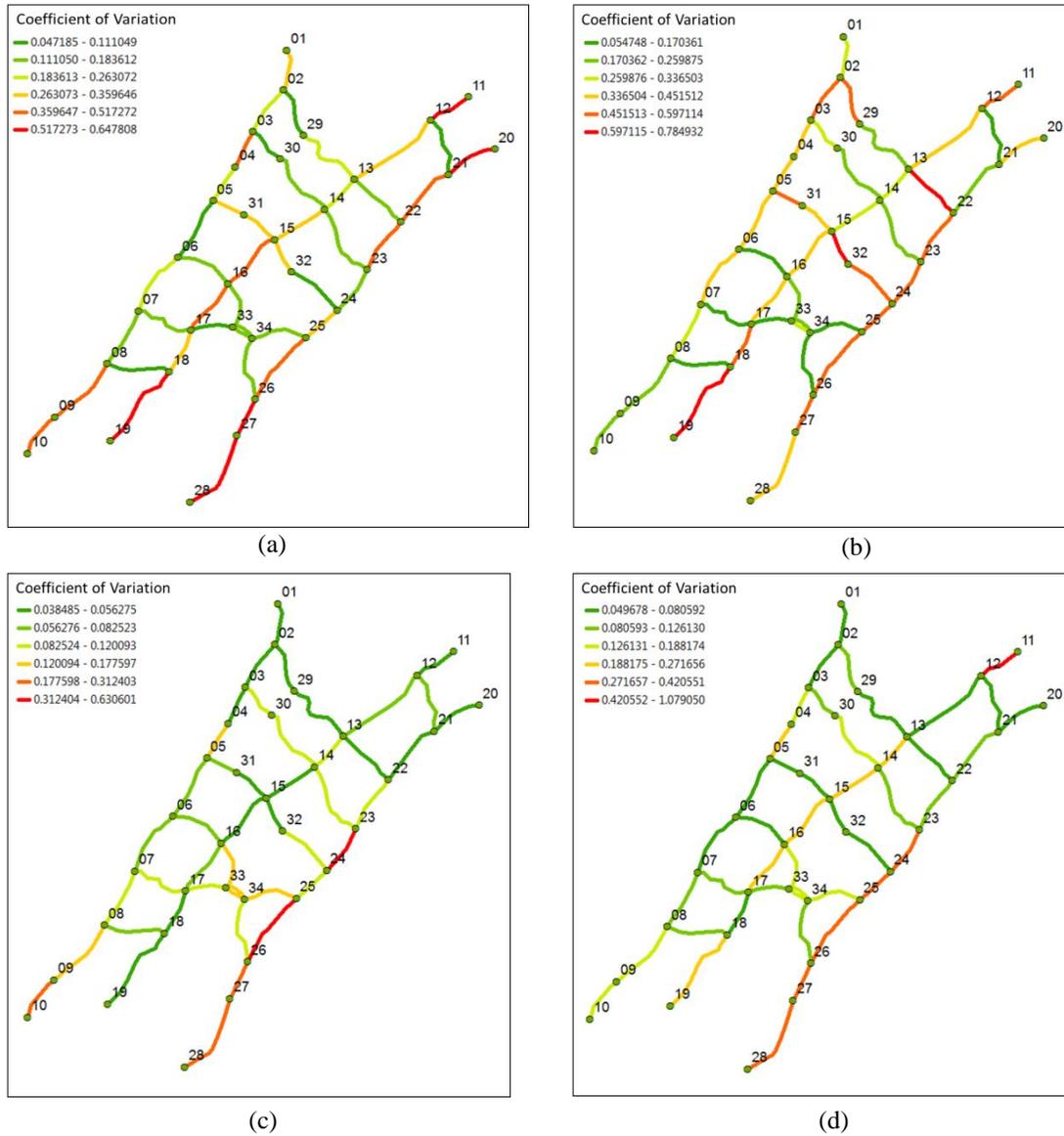


Figure 9: Spatial Distributions of Travel Time Coefficient of Variations in the Study Network: (a) Weekday Morning Peak Period; (b) Weekday Afternoon Peak Period; (c) Weekend Morning Peak Period; (d) Weekend Afternoon Peak Period.

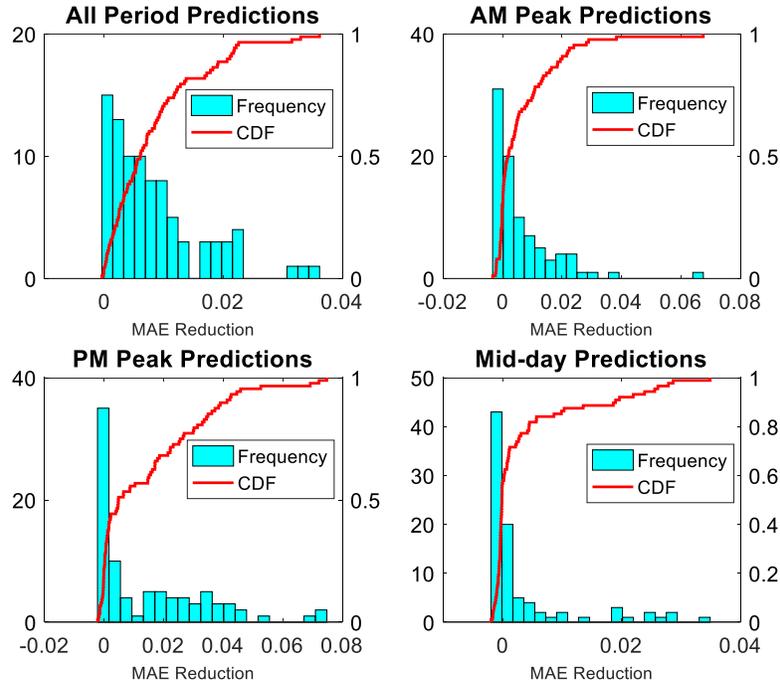
For data-driven based prediction, random forest (RF) has been empirically proved to have satisfactory prediction robustness and accuracy (Y. Zhang and Haghani 2015). To evaluate and investigate travel time prediction performance for each network segment, we adopted RF as the prediction model. The reason RF is selected for benchmark study is that RF has high robustness with respect to different travel time datasets. More concretely, the boosting and bagging schemes enable one to build up quickly a suitable RF with a satisfactory prediction based on the training

dataset. Hence, for each network segment, a specific RF can be estimated and built up with data collected from the same location for temporal prediction test. Moreover, for spatial information based prediction, associated travel time data from different locations can be jointly used to estimate a suitable RF model.

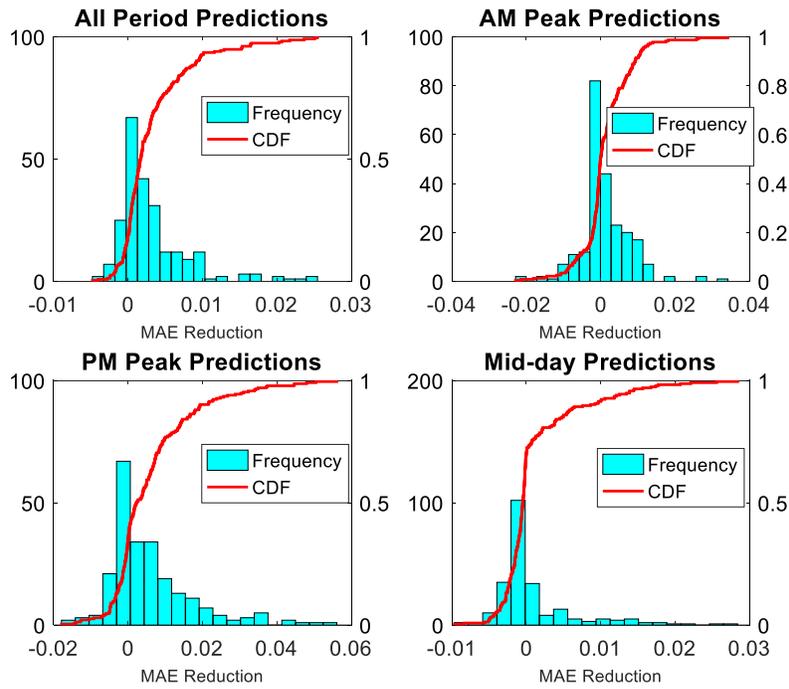
As we have mentioned above, the main focus here is to evaluate the travel time prediction performance at different locations comparatively. Even though we adopted RF as the basic prediction framework in the following analysis, one can pick another type of statistical prediction. Comparison of the prediction performance by different models is out of the scope of this research.

Figure 10, Figure 11 and Figure 12 display the distribution of travel time prediction accuracy improvement for overall segments with real-time information that is measured and input to the RF prediction model. Here we assume the memory storing the real-time data is 30 minutes, i.e., the model uses the recent 30-minute data as prediction input, and aims to predict travel time in future 20 minutes. In each figure, the upper plot shows the prediction error reduction with real-time temporal information served as forecast input, and the lower plot shows the prediction error reduction with real-time spatial information served as forecast input. Here, spatial information-based real-time prediction of a specific segment means using the real-time travel time data from either downstream or upstream segment for prediction.

With the focus on the impact of real-time travel time information on prediction performance improvement, we numerically compared the real-time prediction errors against the historical average based prediction errors and obtained the expected error reduction for each segment during different periods with a different type of real-time information (i.e., temporal or spatial information). As is noted in the distribution plots, there exist cases that the error reduction is less than 0. This implies the real-time information based prediction does not improve the prediction performance compared to using historical average data.

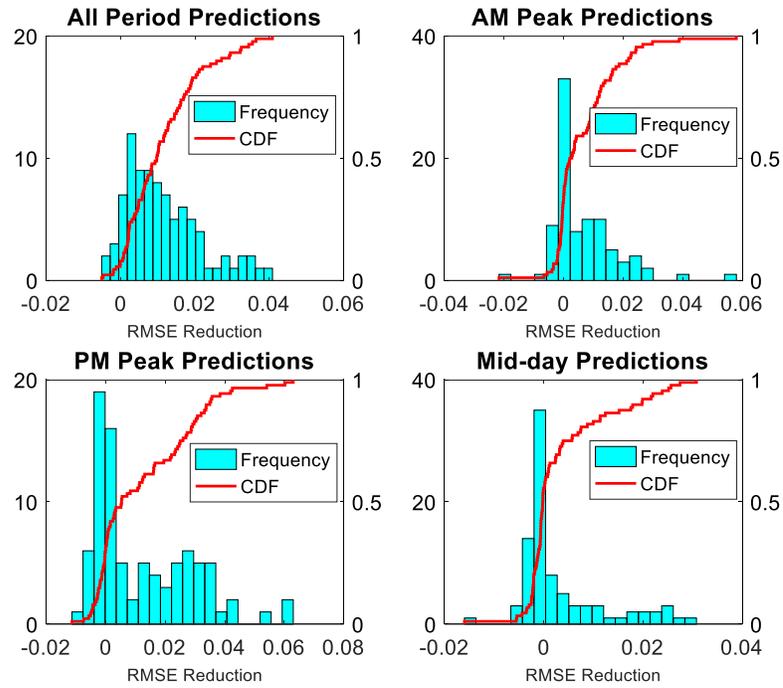


(a): Temporal Information-Based Prediction

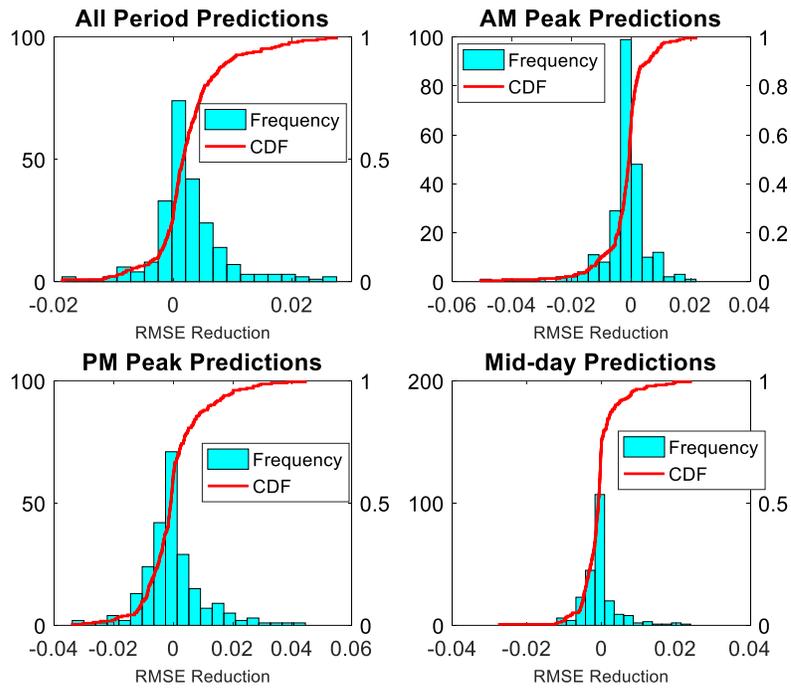


(b): Spatial Information-Based Prediction

Figure 10: Real-time Information Based Travel Time Prediction MAE Reduction Distributions Across Different Locations at Various Time Periods.

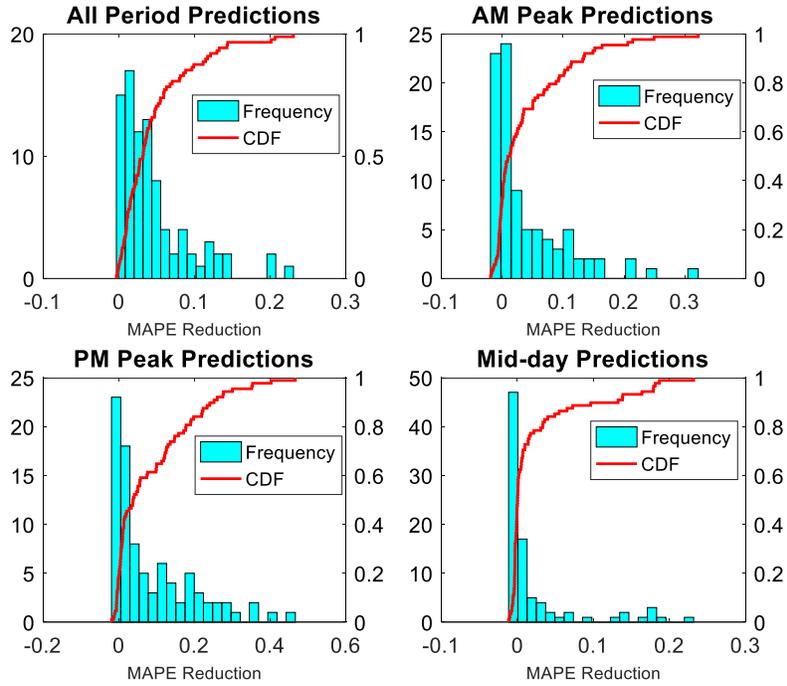


(a): Temporal Information-Based Prediction

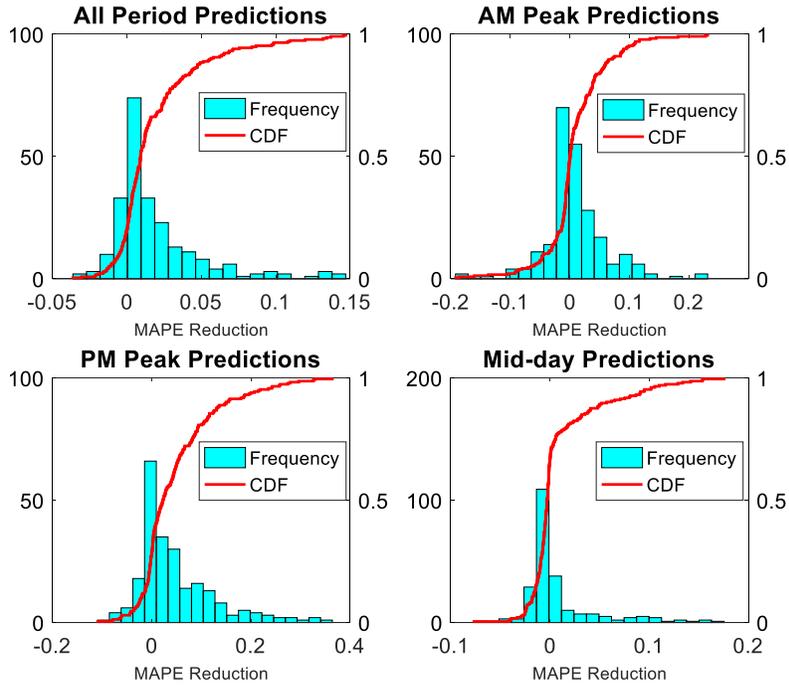


(b): Spatial Information-Based Prediction

Figure 11: Real-time Information Based Travel Time Prediction RMSE Reduction Distributions Across Different Locations at Various Time Periods.



(a): Temporal Information-Based Prediction



(b): Spatial Information-Based Prediction

Figure 12: Real-time Information Based Travel Time Prediction MAPE Reduction Distributions Across Different Locations at Various Time Periods.

Comparison of prediction performance at different locations with real-time data as input provides several implications:

- Temporal information-based predictions are more effective than spatial information-based predictions. This can be introduced by the cumulative distribution functions (CDF) describing error reductions with respect to temporal data based predictions and spatial data based predictions. More specifically, over 90% of segments' travel time predictions can be further improved if temporal real-time data is provided, and that percentage is only around 60%-70% when spatial real-time data is served as input.
- Travel time prediction error reductions are more significant during the AM peak and PM peak periods than during the mid-day period. In other words, the travel time error based on historical average prediction is not significantly larger than that based on real-time prediction since the travel time fluctuation is not high in the mid-day periods.

Comparisons of the three error measurements (i.e., MAE, RMSE, and MAPE) based on the travel time predictions for all the study segments are displayed in Figure 13 to Figure 15. All these three error measurements can be used to quantify the prediction performance. But they provide different perspectives. Specifically, MAE provides an intuitive view on the average dispersion of the prediction from the true value, and MAPE describes this average dispersion from a relative perspective. RMSE evaluates the prediction robustness.

As is shown in Figure 13, RMSE and MAPE yielded from travel time prediction at the same location approximately have a linear relationship. But a lower MAPE does not necessarily indicate a lower RMSE. A similar relationship can be observed between MAE and RMSE as well. The monotonous positive relationship between MAE and MAPE is more significant, especially for predictions based on temporal information. In summary, the three error measures sometimes are in high agreement for predictions at some locations. But there are still lots of cases in which they do not agree with each other to describe the prediction performance. Therefore, to

comprehensively evaluate the performance of a particular prediction case, it is risky to ignore one of the three error measurements.

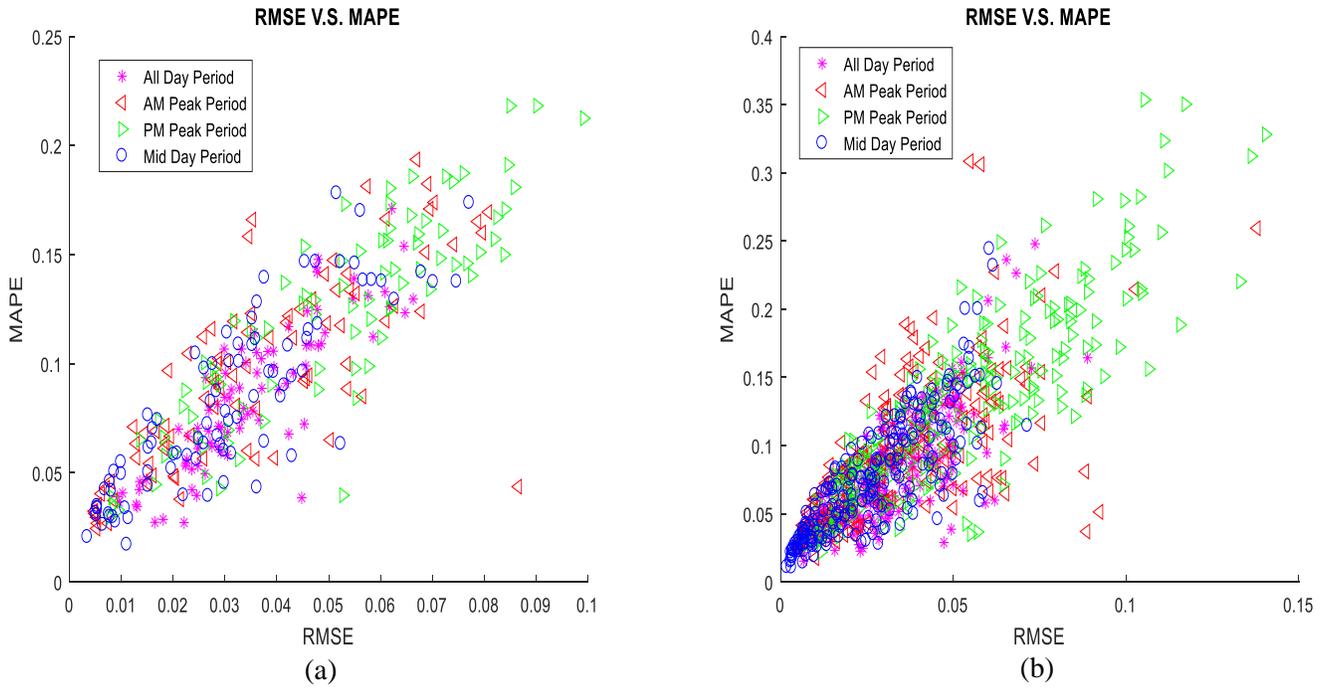


Figure 13: Empirical Relationship Between Travel Time Prediction RMSE and MAPE: (a) Real-time Temporal Information Based Prediction, (b) Real-time Spatial Information Based Prediction.

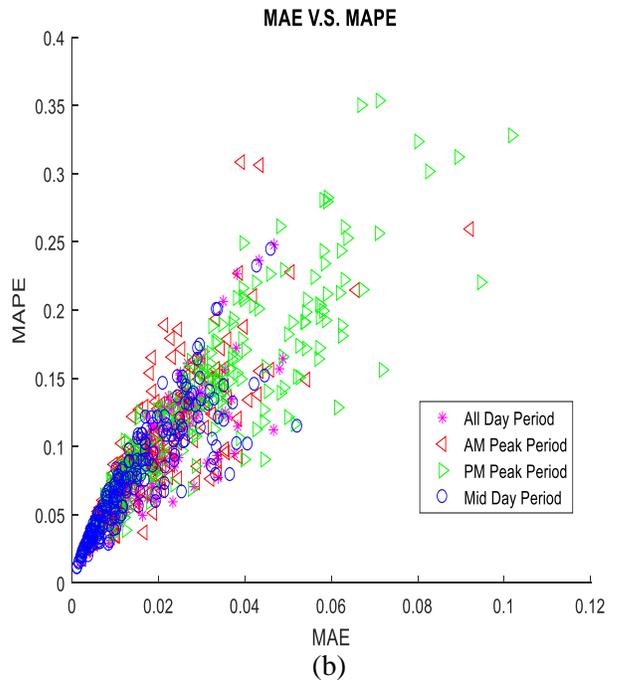
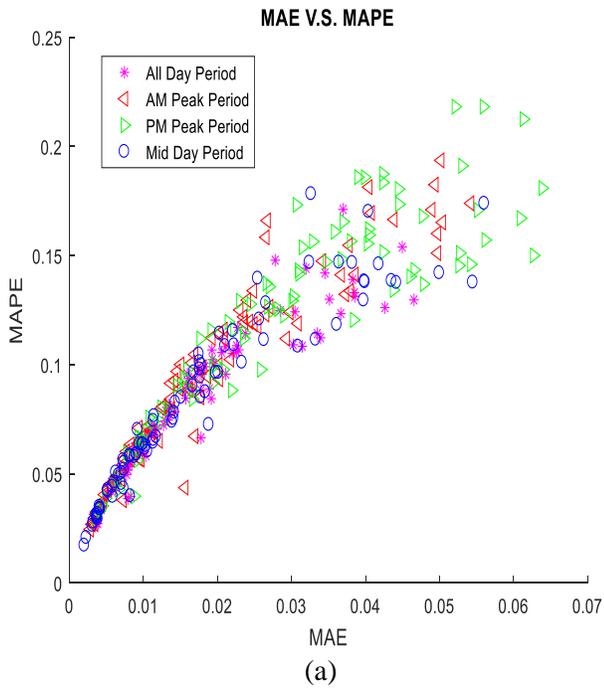


Figure 14: Empirical Relationship Between Travel Time Prediction MAE and MAPE: (a) Real-time Temporal Information Based Prediction, (b) Real-time Spatial Information Based Prediction.

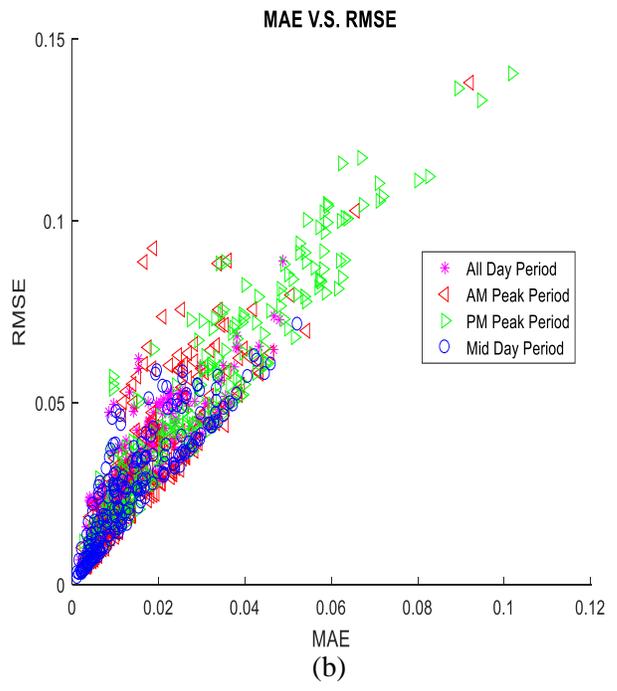
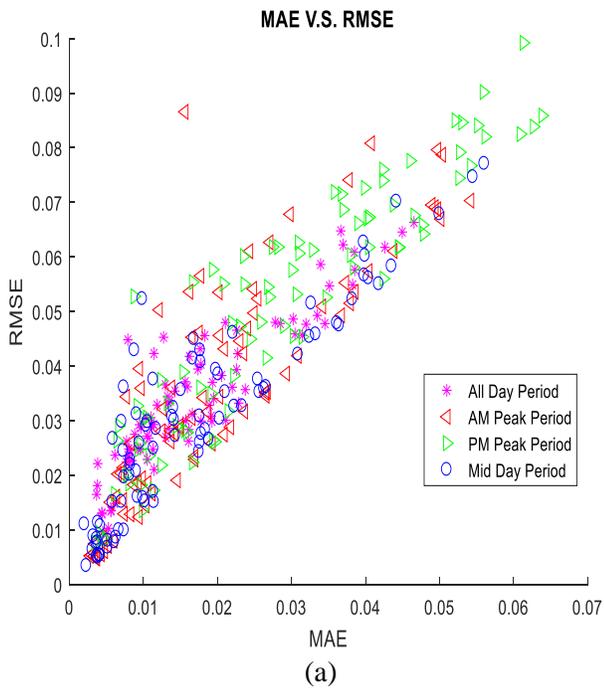


Figure 15: Empirical Relationship Between Travel Time Prediction MAE and RMSE: (a) Real-time Temporal Information Based Prediction, (b) Real-time Spatial Information Based Prediction

### 3.4.3.3 Travel Time Prediction Uncertainty Evaluation

We applied the above-defined travel time characteristic vector to evaluate the travel time prediction uncertainty with/without real-time input data. The characteristic vector consists of three components, i.e., magnitude parameter, trend parameter, and variation parameter. Here we name the three-component characteristic vector as the state vector. For a given time window (e.g., 30 minutes), travel time series of a specific segment can be calculated by the formulas given in section 3.4.2. Different highway segments have different lengths, which determine the average travel time from different locations are different. Therefore, we linearly normalize the travel time data of each segment by using the standard score based method:

$$X_i^{norm} = \frac{X_i - \hat{\mu}}{\hat{\sigma}}$$

where,  $\hat{\mu}$  and  $\hat{\sigma}$  are the sample mean and sample standard deviation of the travel time data for a particular segment.  $X_i$  is the measured travel time value and  $X_i^{norm}$  is the normalized travel time value.

Each piece of travel time series corresponds to three real values determining a specific state vector. Travel time is a continuous variable. Thus there are an infinite amount of different realizations of the state vector defined in this study. Hence one needs to classify the state vector into several representative classes to use the proposed state uncertainty estimation model. There is a tradeoff between the class number and the model vitality. State classification with very few classes may not fully represent the travel time pattern. On the one hand, classifying travel time states into too many classes will decrease the sample size for each particular state given a fixed dataset. On the other hand, empirical travel time data always contains noises. Higher classification resolution may cause the classification to be prone to classification error. By preliminary analysis with travel time data used in this study, we classify the travel time magnitude level into five classes, the trend parameter into five classes and the variation parameter into three classes (Table 3-Table 5).

Table 3: Travel Time Magnitude Classification.

Classification Interval	Implication
(-inf, 0)	Normal Traffic
(0,1)	Level-1 Congestion
(1,2)	Level-2 Congestion
(2,3)	Level-3 Congestion
(3,inf)	Level-4 Congestion

Table 4: Travel Time Trend Classification.

Classification Interval	Implication
(-inf, -1.5)	Significant Decreasing
(-1.5, 0.5)	Moderate Decreasing
(0.5, 0.5)	Unchanged
(0.5, 1.5)	Moderate Increasing
(1.5, inf)	Significant Increasing

Table 5: Travel Time Reliability Classification.

Classification Interval	Implication
(0, 0.25 <sup>2</sup> )	Stable Travel Time
(0.25 <sup>2</sup> , 0.5 <sup>2</sup> )	Moderate Unreliable Travel Time
(0.5 <sup>2</sup> , inf)	Significant Unreliable Travel Time

Based on the data normalization method,  $mag_i = 1$  indicates the magnitude of the travel time series in the given time window is equal to the entire sample mean (i.e. historical average travel time). As one may note in the above classification parameter table, we consider the magnitude levels of travel time series less than the historical mean in the same class. This is because travel time distributions on most freeway or arterial segments in the study area are commonly highly left-skewed (i.e. historical mean travel time is close to free-flow travel time). In such a case, travel times less than the historical mean value are all considered in one class and imply the traffic condition is good. For abnormal travel time values that are significantly higher than the mean travel time, one should pay more attention and distinguish them with different classes, since they imply severe traffic congestions with different congestion levels.

With the proposed travel time state classification, the entropy estimation model was applied to evaluate real-time information-based prediction uncertainty for each segment in the

study network. Figure 16 and Figure 17 display two evaluation results obtained from the temporal information-based uncertainty estimation model for two arterial segments, i.e., segment 0706 and segment 0605. Heat maps in each figure visualize the state transition probability matrixes at a different time of the day. The first number above the colored matrix denotes the travel time state uncertainty without specifying the real-time measurement (unconditional based state distribution entropy), and the second number above the matrix denotes the travel time state uncertainty given the real-time measurement is specified (conditional based state distribution entropy).

As is shown in the results, the conditional state uncertainty representing the real-time prediction uncertainty is always lower than the unconditional state uncertainty representing the historical distribution based inference. For both segments, the uncertainty reduction during the AM peak periods with knowledge of the real-time temporal travel time information is very significant (i.e., from 1.99 to 1.04 for segment 0706, and from 2.22 to 1.13 for segment 0605). Another interesting finding is that the state prediction uncertainties for both segments during the AM peak period are commonly larger than that during the mid-day and the PM peak periods. This implies the travel time on these two segments is much harder to predict in the AM peak periods than that in the other time periods. For example, the state prediction uncertainty on segment 0706 without real-time information during the mid-day period is only 0.71, while this value is 1.04 for prediction based on the real-time information.

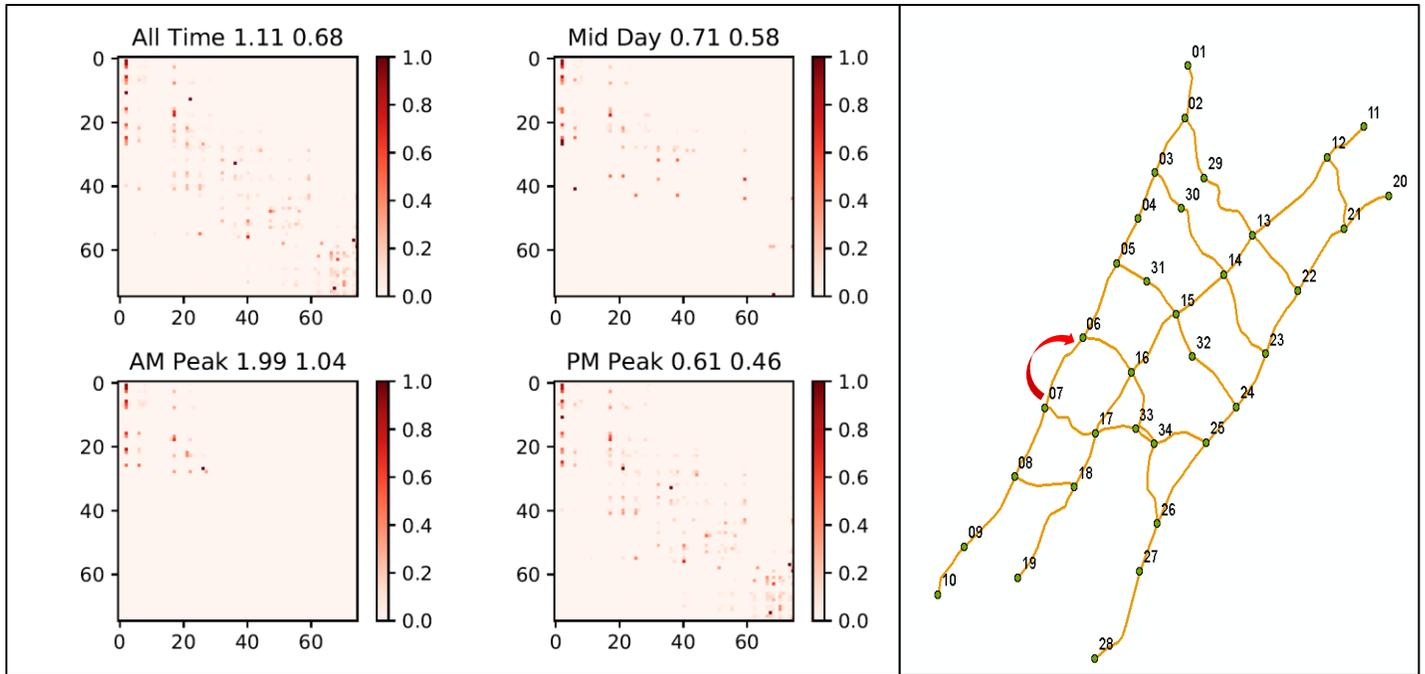


Figure 16: Travel Time Prediction Uncertainty Comparison Without/With Real-time Temporal Information for Segment 0706 at Different Periods.

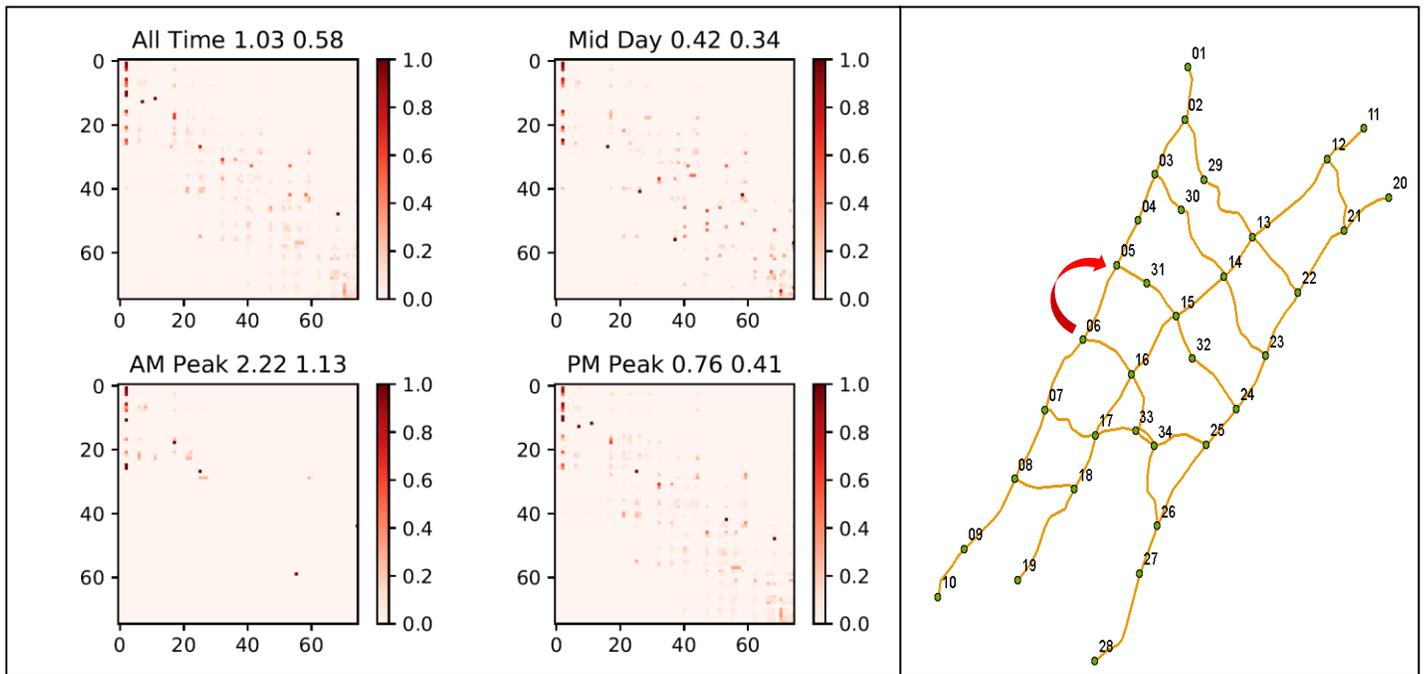


Figure 17: Travel Time Prediction Uncertainty Comparison Without/With Real-time Temporal Information for Segment 0605 at Different Periods.

Further, Figure 18, Figure 19, Figure 20, and Figure 21 demonstrate the travel time state prediction uncertainties on segment 0706 without/with real-time spatial data as input.

Specifically, Figure 18 and Figure 19 display the calculated state probability mapping matrixes between segment 0706 and its immediate downstream segment, and between segment 0706 and its second immediate downstream segment. Figure 20 and Figure 21 display the calculated state probability mapping matrixes between segment 0706 and its immediate upstream segment, and between segment 0706 and its second immediate upstream segment.

The spatial information based plots indicate that travel time uncertainty on segment 0706 can be reduced as well if real-time information from its downstream or upstream segment is provided. Further, by comparing state uncertainty reductions under different sources of real-time information, one can find the immediate upstream segment's state information is more valuable than that of the second immediate upstream segment's information when predicting the travel time for segment 0706. This is still true for the downstream information based prediction.

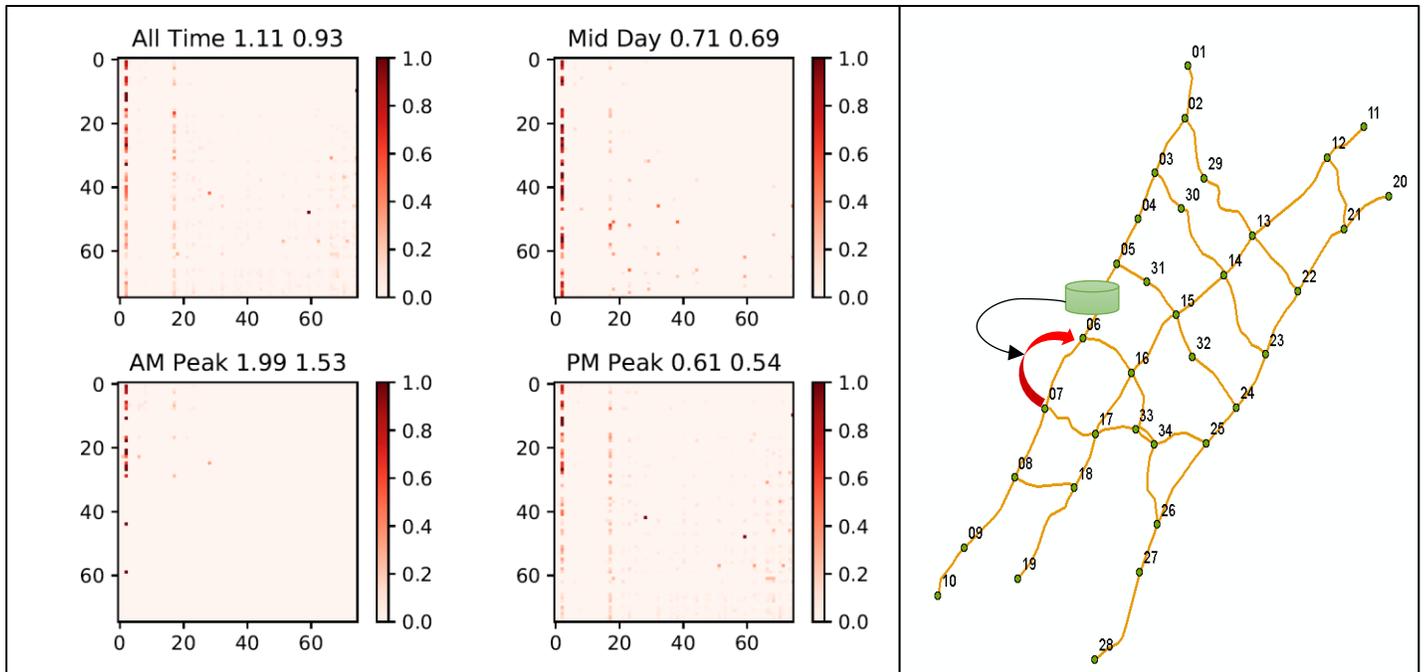


Figure 18: Travel Time Prediction Uncertainty Comparison Without/With Real-time Immediate Downstream Segment (i.e., 0605) Information for Segment 0706 at Different Periods.

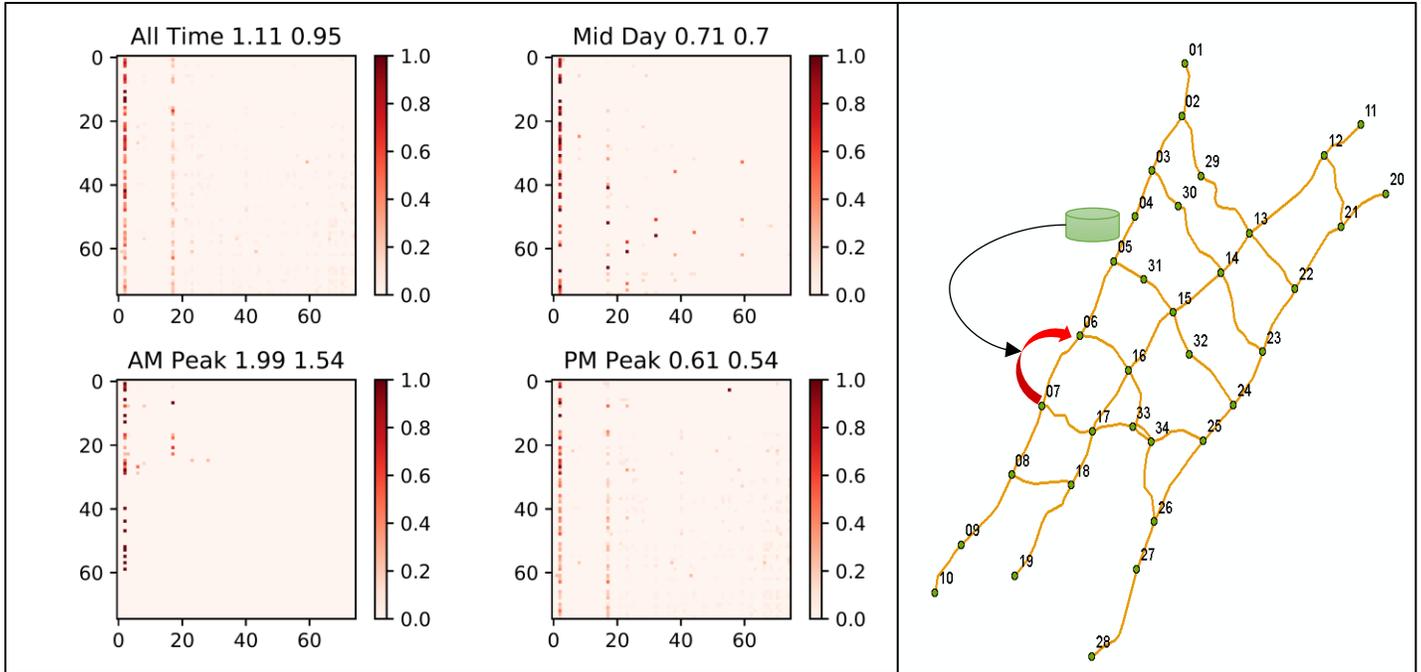


Figure 19: Travel Time Prediction Uncertainty Comparison Without/With Real-time Second Immediate Downstream Segment (i.e., 0504) Information for Segment 0706 at Different Periods.

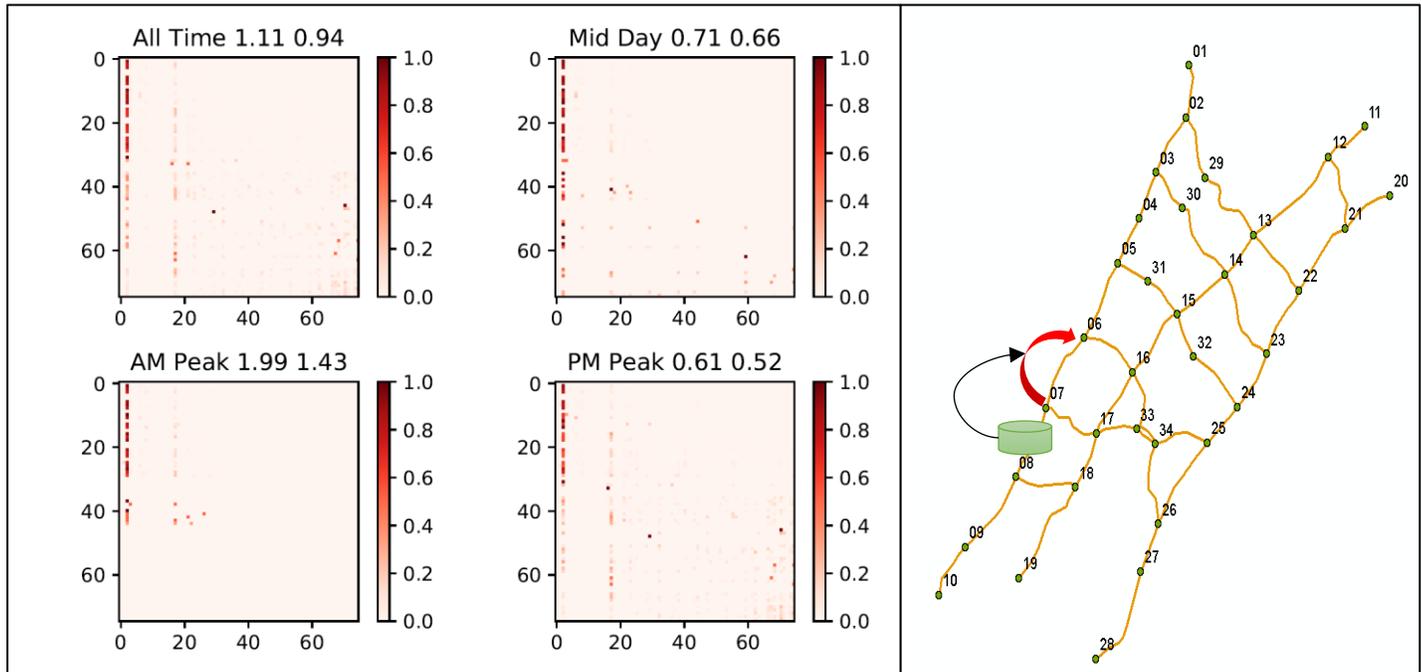


Figure 20: Travel Time Prediction Uncertainty Comparison Without/With Real-time Immediate Upstream Segment (i.e., 0807) Information for Segment 0706 at Different Periods.

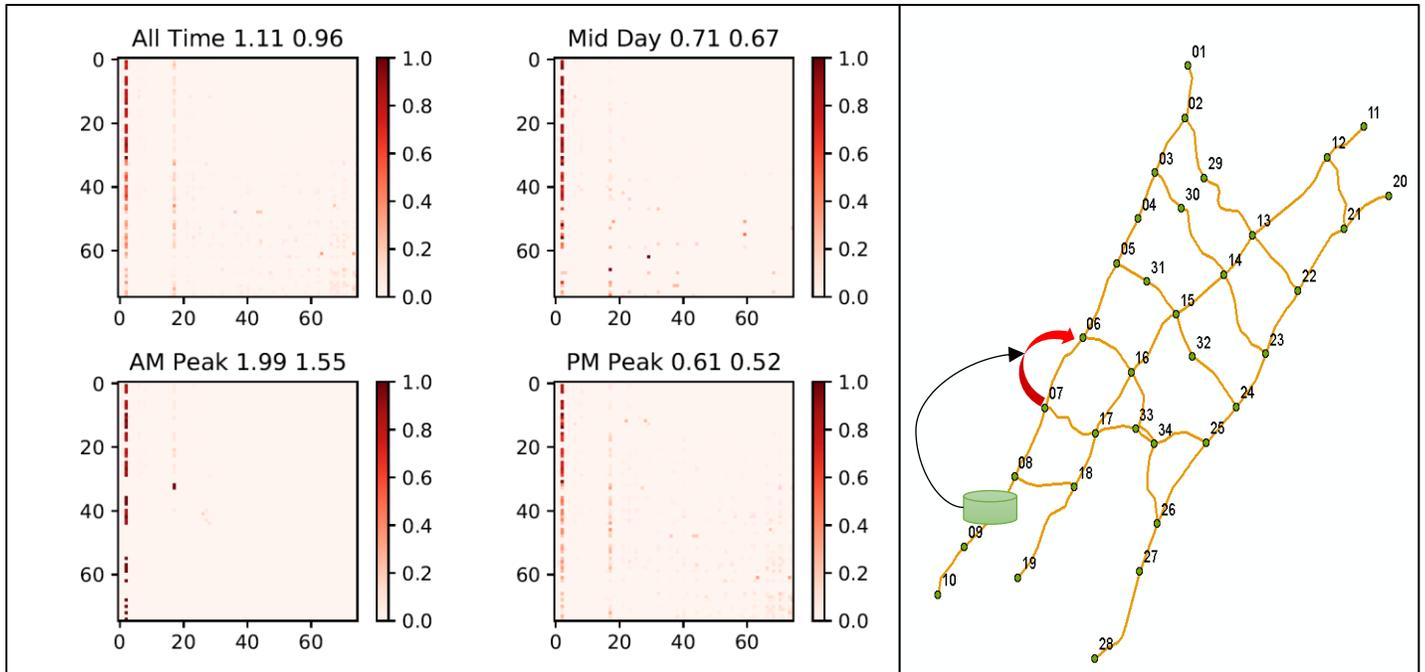


Figure 21: Travel Time Prediction Uncertainty Comparison Without/With Real-time Second Immediate Upstream Segment (i.e., 0908) Information for Segment 0706 at Different Periods.

The model was applied to all highway segments in the study network to evaluate and compare the travel time prediction uncertainty reductions at various locations based real-time travel time information. Figure 22 plots the comparison of historical information-based travel time prediction uncertainty and real-time temporal information-based travel time prediction uncertainty on different segments of different time of the day. In the plot, the Y-axis denotes the prediction uncertainty value calculated based on historical travel time distribution, and the X-axis denotes the prediction uncertainty value calculated on the condition that the real-time measurements are given. Specifically, the real-time measurements for each segment are referred to the travel time series collected from the segment in the past 30 minutes with 5-minute aggregation interval. As is shown, the prediction uncertainty for each segment of each period with the real-time information given is lower than that without the real-time information (i.e., all data points are below the diagonal line  $y = x$ ). This indicates that real-time temporal information can reduce the travel time prediction uncertainty for every segment in the network. Moreover, one can see from the figure that the travel time prediction uncertainty reductions during PM peak period

are more significant. This implies knowledge of real-time information is more beneficial for reducing the prediction uncertainty with respect to the study highway segments.

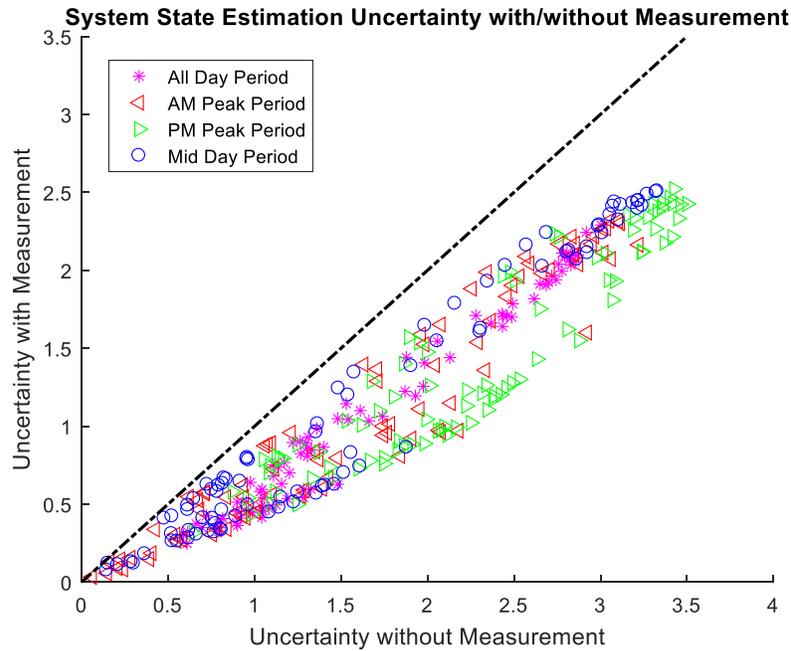


Figure 22: Historical Inference based Travel Time Prediction Uncertainty V.S. Real-time Temporal Information based Travel Time Prediction Uncertainty.

Figure 23 shows the travel time prediction uncertainties for all segments without/with the real-time spatial information as conditional inputs. Here, spatial real-time information, which is used to predict travel time for a specific segment, refers to the real-time travel time data collected from the downstream or upstream segment instead of directly measured from the segment itself. Overall speaking, taking advantage of real-time information from downstream/upstream segments can reduce the travel time prediction uncertainty as well, especially for predictions made during the PM peak periods. Comparing the results shown in Figure 22 and Figure 23, one can note prediction uncertainty reductions under the temporal real-time information are more significant than that under the spatial real-time information. These empirical results reveal that for travel time predictions on a specific segment in the study network, knowing the temporal information is more useful than the spatial information.

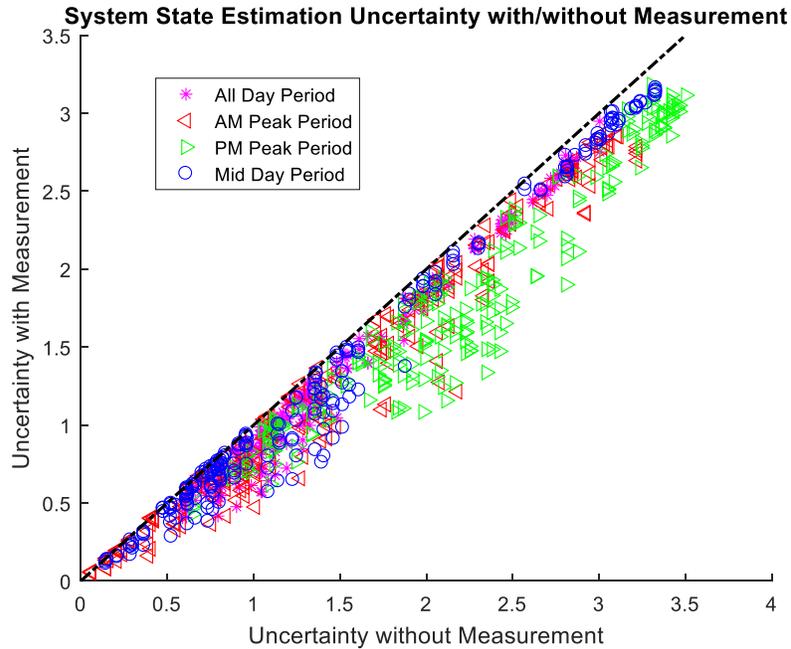


Figure 23: Historical Inference based Travel Time Prediction Uncertainty V.S. Real-time Spatial Information based Travel Time Prediction Uncertainty.

#### 3.4.3.4 Empirical Analysis of Prediction Error Reduction and Uncertainty Reduction

Travel time uncertainty was modeled and estimated by considering the travel time pattern evolving in a stochastic process for a specific highway system. Travel time uncertainty without any real-time information is estimated based on its historical distribution. When a particular piece of real-time information is given, travel time uncertainty can be estimated based on the conditional distribution. As is demonstrated by the above empirical analysis, travel time uncertainty can be reduced to some extent when either spatial or temporal real-time travel time information is provided for each segment in the study network.

System state uncertainty can be interpreted and understood as one quantitative measure of the state predictability. For a stochastic system, one can never predict the system state with perfect accuracy due to the system stochasticity. But one can evaluate and anticipate the prediction accuracy based on the uncertainty level. By referring to the state uncertainty defined in this research, one can expect the prediction performance based on a particular vector of

information. In practice, people care more about the effectiveness of a given real-time measurement system. For example, if the measurements provided by the real-time surveillance system can assist to improve the system state prediction accuracy to a large extent, one can conclude the surveillance system is effective. Otherwise, the measurements are not useful for predictions, and one should consider another surveillance system or define another measurement space.

Prediction errors (e.g., MAE, MAPE, and RMSE) have different units from the uncertainty index proposed in this study. Conventionally, the error can provide a more intuitive way for people to understand the prediction performance. However, empirically evaluated prediction errors may be affected by many factors such as model selection, model robustness, and model fitting. In other words, to have insights into the state predictability based on prediction errors, one should thoroughly specify and evaluate the prediction performance by many different prediction models. The proposed uncertainty measurement, which is in units of entropy, although can evaluate the system state predictability from a stochastic perspective, it does not provide an intuitive way to understand the system predictability. Here we numerically compare the prediction error reductions given by the RF prediction model and the prediction uncertainty reductions evaluated by the proposed model. The comparisons are displayed in Figure 24, Figure 25 and Figure 26.

As is pointed out in previous experiments, a real-time data-driven based prediction model may fail for predictions on some segments compared to the historical inference-based prediction approach. These cases, in which prediction errors with real-time information is even larger than that of historical inference-based predictions, are marked in the red shaded area in each figure. One finding already pointed out in previous subsections can be confirmed as well based on the following figures. For temporal information-based real-time travel time prediction, it is easier to train an effective prediction model than that by using spatial information as input (i.e., downstream or upstream travel time data). Further, one can note from the comparison plots, all

three error measurements (i.e., MAE, RMSE, and MAPE) obtained by the RF models have relationships with the uncertainty measurements. For temporal information-based predictions, the travel time prediction error reduction of a particular segment has a nonlinear and monotonous increasing relationship with the uncertainty reduction estimated by the entropy model. For spatial information-based predictions, the error reduction has a linear and monotonous increasing relationship with the uncertainty reduction. This empirical analysis provides an alternative to anticipate the prediction errors based on the uncertainty model.

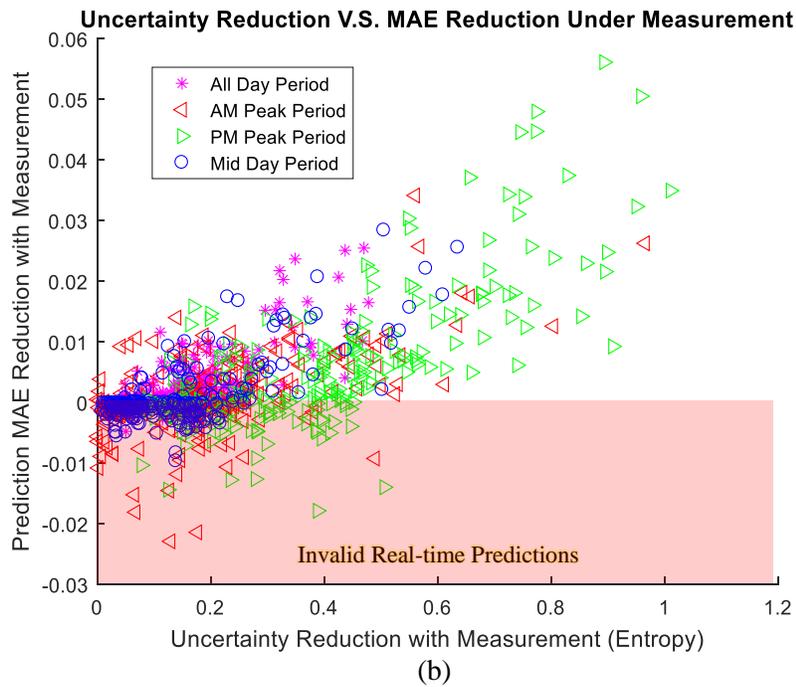
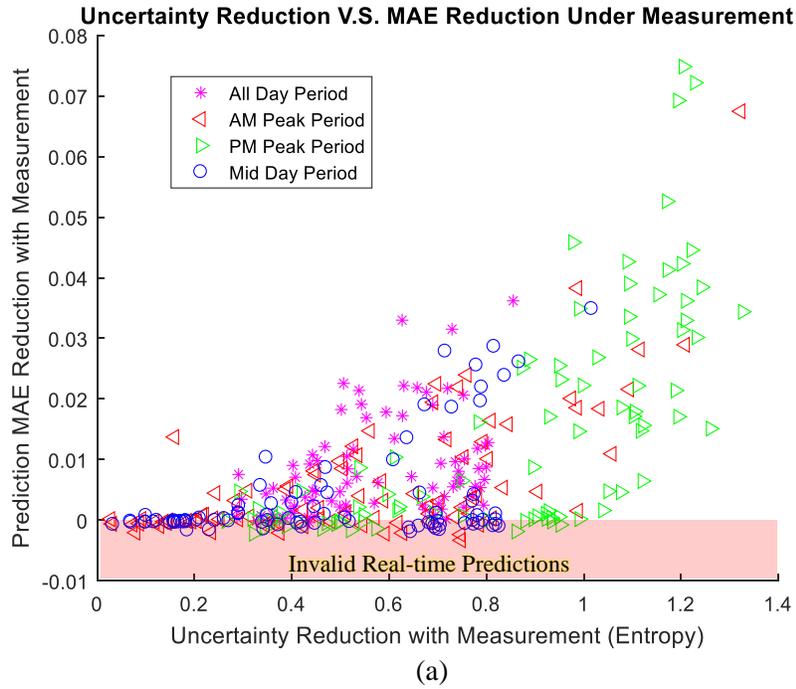


Figure 24: Travel Time MAE Reductions by RF Predictions V.S. Travel Time Uncertainty Reductions Estimated by the Proposed Model: (a) Conditioning on Real-time Temporal Information; (b) Conditioning on Real-time Spatial Information.

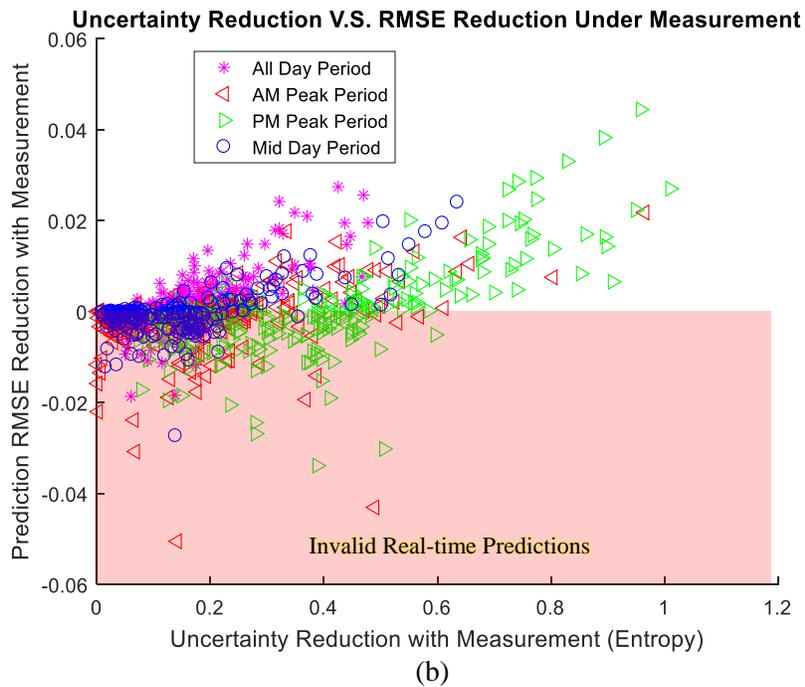
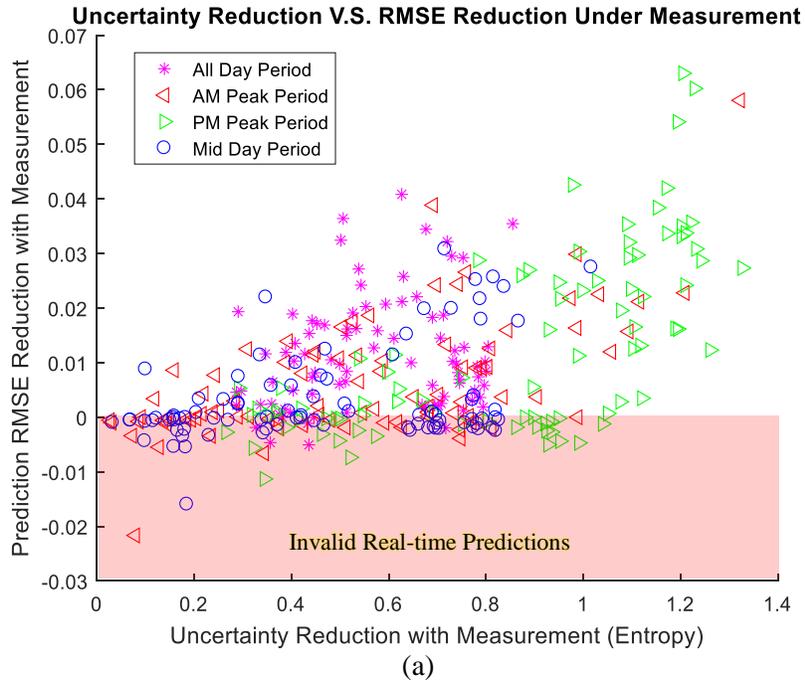


Figure 25: Travel Time RMSE Reductions by RF Predictions V.S. Travel Time Uncertainty Reductions Estimated by the Proposed Model: (a) Conditioning on Real-time Temporal Information; (b) Conditioning on Real-time Spatial Information.

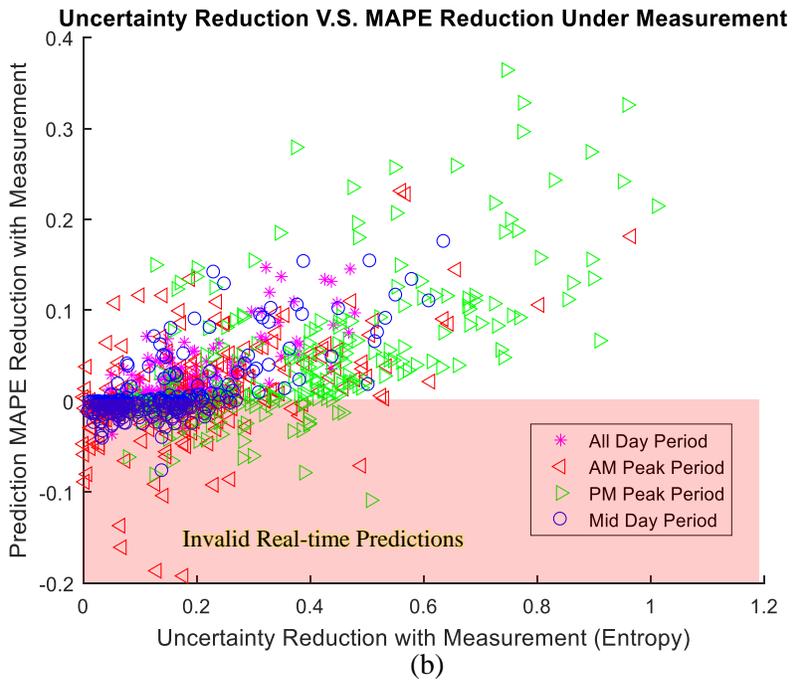
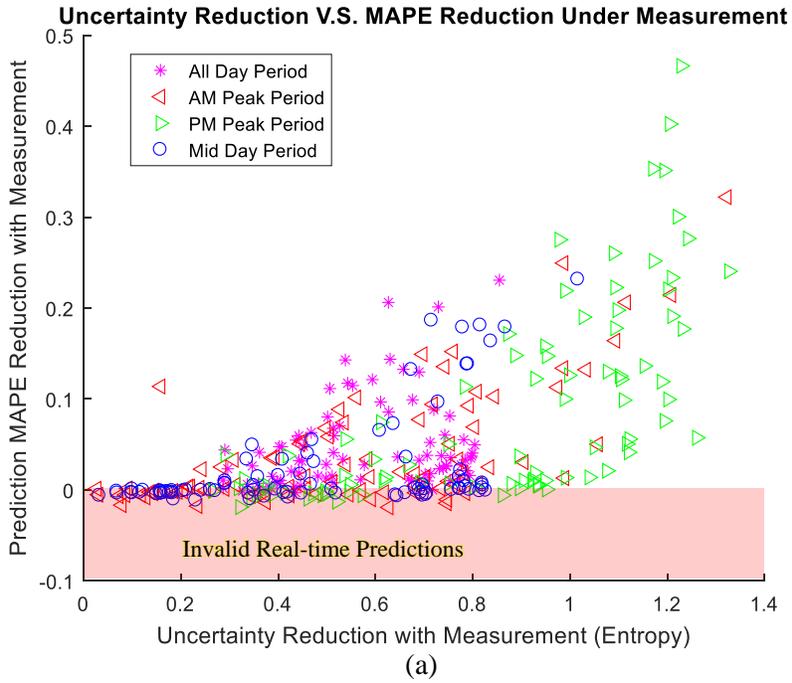


Figure 26: Travel Time MAPE Reductions by RF Predictions V.S. Travel Time Uncertainty Reductions Estimated by the Proposed Model: (a) Conditioning on Real-time Temporal Information; (b) Conditioning on Real-time Spatial Information.

### 3.5 *Summary and Conclusions*

In this chapter, we studied the benefit of introducing real-time surveillance system from two perspectives. The first one is the system state prediction error reduction and the second one is system prediction uncertainty reduction. In terms of evaluating state prediction error reduction of a specific system, we introduced the procedures to statistically estimate the prediction error without real-time measurements and the prediction error given specific real-time measurements. For this type of prediction performance evaluation, a specific prediction model needs to be specified. In addition to the prediction error reduction-based surveillance effectiveness evaluation, we defined a new framework to evaluate the system state uncertainty based on entropy from a stochastic perspective. The proposed uncertainty estimation model can be generalized and applied to different stochastic systems. State prediction uncertainty with real-time surveillance can thus be estimated as the conditional state distribution entropy conditioning on such real-time measurements.

Both the error reduction approach and the proposed uncertainty model are carefully designed and applied to a real-world transportation system to evaluate the surveillance effectiveness on temporal and spatial travel time prediction. Specifically, random forest model is selected as the real-time prediction error evaluation method due to its generality and robustness. The effectiveness of real-time temporal information and real-time spatial information are obtained and analyzed for 88 highway segments. The error reduction analysis based on real-world travel time data show that the RF model with real-time input data cannot improve the prediction performance everywhere in the study network. For most highway segments, the RF model can take advantage of the real-time travel time to bring prediction benefits compared to the historical inference-based prediction.

Numerical experiments using the uncertainty evaluation model indicate both real-time temporal and spatial travel time data can improve travel time predictability. Specifically, real-

time temporal information is more valuable for real-time travel time prediction. This empirically confirms the statement that data-driven based travel time prediction usually outperforms the historical average prediction. At least, knowing most recent travel time series on a road does not bring a negative effect on real-time predictions for the road.

Based on the empirical results given by both error reduction analysis and uncertainty reduction analysis, real-time temporal travel time data is shown to be more useful than real-time spatial travel time data for travel time predictions for a specific segment. Moreover, the relationship between travel time error reductions and travel time uncertainty reductions are analyzed and given. We empirically prove and demonstrate the monotonously increasing relationship between travel time error reduction and travel time uncertainty reduction given by two completely independent approaches. This provides a new possibility to evaluate and anticipate prediction error reductions based on the proposed uncertainty estimation framework.

# Chapter 4 Traffic Sensor Network Planning Based on A Prior Information

## 4.1 Static Planning Model for Sensor Location Optimization

### 4.1.1 Introduction

In this section, a static optimization model is proposed to determine the installation locations of traffic sensors with the objective of collecting and providing travel time information within a given highway network. As is previously discussed, Automatic Vehicle Identification (AVI) technology is the most direct and efficient approach for large-scale travel time information collection. Thus, the proposed optimization model is specifically in terms of AVI-based traffic sensors locations determination. Although the developed optimization model offers the flexibility to additionally incorporate the point sensors locations determination (by introducing new decision variables), sensors belonging to this type are excluded in this study due to their trivialness in travel time measurement.

### 4.1.2 Preliminaries

The physical layout is represented by a network  $G = (N, A)$ , where  $N$  is a set of nodes representing intersections, interchanges, or intermediate locations of a highway network, and  $A$  denotes the set of arcs connecting those nodes. In a highway network, we use  $L$  to denote a particular set of target surveillance segments. An element  $l_i$  of  $L$  is usually considered with homogeneous physical layout, and it can either be a single arc  $a_i \in A$ , or can be a multi-arc corridor consisting of several consecutive arcs. In a travel time provision based sensor location problem, it can exactly be the arc set  $A$ , or a pre-determined set consisting both network arcs and paths. This mainly depends on the highway network structure and the corridor of interest for collecting and providing travel time information. Let  $Z \subseteq N$  represents the set of feasible sensor

installation locations. Here the potential installation locations are considered as a subset of  $N$ , which is associated with the installation feasibilities determined by the real-world environmental conditions. In general, feasible sensor installation locations  $Z$  should always be evaluated in advance. Excluding infeasible locations cannot only guarantee the problem to be resolved effectively but can also enhance the solution efficiency.

Each segment  $l \in L$ , directionally connects a pair of end nodes within the network. The start node and end node connected by  $l$  are here represented by  $s(l)$  and  $e(l)$ , respectively. In this dissertation, to avoid confusion, arc or segment index are always denoted with alphabets  $l$  and  $k$ , while the node or location index are represented by alphabets  $i$  and  $j$ . Each segment of interest  $l$  has an associated traffic flow rate  $f_l$ , which can be viewed as the average traffic volume or the peak volume in the static model.

At this stage, the cost parameters in the objective function are referred to as the monitoring benefit gains, which are represented by  $B_l$  and  $B_{kl}$ . Specifically,  $B_l$  denotes the benefit gains by directly using sensors to monitor the travel time information for segment  $l$ . This type of monitoring benefit is nearly considered in each existing studies, and we name it as direct monitoring benefit or temporal monitoring benefit. Since there always exist correlations between the traffic states of some relevant segments, such as upstream and downstream segments or corridors sharing similar OD demand, the term “indirect monitoring benefit”  $B_{kl}$  is proposed and defined in this research.  $B_{kl}$  arises from the traffic state correlations between segments  $k$  and  $l$ . It describes the benefit gains for segment  $l$  when it is not directly monitored, but its associated segment  $k$  is monitored. Intuitively, the higher traffic state correlation between those two segments, the more likely this type of benefit gains will be larger. Due to the correlation consideration from a spatial perspective, we also call  $B_{kl}$  as spatial monitoring benefit gain.

Accordingly, we use  $I(l)$  to denote the set of associated segments of  $l$ . In a real-world application, this could be the set consisting  $l$ 's adjacent segments, and the segments overlapped with  $l$ . Also, preliminary studies can be conducted with empirical datasets to figure out the appropriate  $I(l)$ .

The popularity of GPS-enabled mobile devices, such as smartphones and portable navigators, enriches the traffic information datasets in a real-time manner. Those data are usually called as probe data, which can be used to estimate and provide travel time information by aggregating a set of probe vehicle trajectories. For any traffic sensor location problem, the existing sensors layout, as well as those real-time probe data, should not be ignored. Because if data from those sources are stable and reliable, consideration of those sources as a supplement can significantly reduce the new deployment cost and necessary operation cost. In the proposed model, we explicitly consider the existence possibility of probe data sources by introducing a binary parameter  $s_l$ . This binary indicator equals to one if there exists reliable probe data source for segment  $l$ , and zero otherwise.

### 4.1.3 Mathematical Formulation

#### 4.1.3.1 Decision Variables

In the proposed static sensor location optimization model, the key decision variable is represented by  $w_i$ , which equals to 1 if location  $i$  is chosen to install a sensor, and 0 otherwise. For the ease of model formulation, two other auxiliary decision variables are defined as well. The first auxiliary variable is  $y_l$ , which is a binary variable and equals to 1 if two AVI sensors directly monitor segment  $l$ , and 0 otherwise. In other words, a segment  $l$  is monitored if and only if its two endpoints have installed sensors. The second auxiliary variable is associated with the indirect segment monitoring and is represented by  $u_{k,l}$ . Given  $k \in I(l)$  and segment  $l$  is without

surveillance,  $u_{k,l}$  equals to 1 if segment  $k$  is chosen to estimate the traffic state of segment  $l$ , and 0 for all the other cases.

#### 4.1.3.2 Model Objective

The optimization objective of the static traffic sensor location problem is formulated in this subsection. As is illustrated above, the total benefits of a particular traffic monitoring configuration consist of two parts. The first one is benefit obtained by directly monitoring a specific segment, and the second type of benefit comes from the information gain of a particular segment indirectly inferred from other information-correlated segments.

##### (a) Direct Monitoring Benefits

When a pair of AVI sensors are deployed at the endpoints of a particular segment, the travel time information can be directly collected for this segment by detecting and matching the information regarding the vehicles that pass. Subsequently, the traffic state variations on this segment can be directly estimated or predicted based on the collected data. This is illustrated in Figure 27. Specifically, the data quality (e.g., sampling size) and the data delay (e.g., matching delay) which is largely dependent on the segment length, will affect the real-time information gains of this segment. This will be further discussed and modeled in later Chapters.

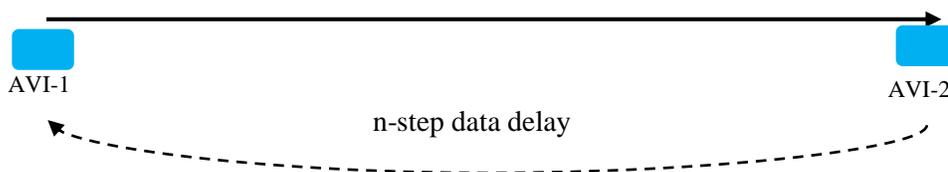


Figure 27: Segment with Direct Travel Time Measurement System

In this way, the total monitoring benefits of a particular network weighted by flow amount can be calculated as shown in Equation (4-1-1). The equation consists of three components. The first part represents the total monitoring benefits from those segments, which are monitored by sensors but without reliable probe data reports. The second part indicates the total monitoring benefits from the segments on which there are reliable probe data reports but

without sensors installed. The last part separately calculates the total information gains for those segments on which both sensors and reliable probe data reports are available. Here, we introduce a scaling parameter  $\beta$  which is greater or equal than 1, indicating the additional information gains of a particular segment when both physical sensors and reliable probe reports are available. A reasonable consideration is that multiple data sources might improve the estimation accuracy of the target system state through data fusion techniques. If this parameter is set to be 1, which means data fusion cannot improve information gains or the improvement is trivial, the last part of Equation (4-1-1) will degenerate to the first two parts in the equation.

$$B_{total}^{direct} = \sum_l f_l \cdot B_l \cdot (1 - s_l) y_l + \sum_l f_l \cdot B_l \cdot (1 - y_l) s_l + \sum_l f_l \cdot \beta \cdot B_l \cdot s_l y_l \quad (4-1-1)$$

(b) *Indirect Monitoring Benefits*

Indirect monitoring benefit arises from the concept that it is possible to predict the traffic state of one segment with the knowledge of the traffic state on another segment. Since traffic demands of different segments a particular corridor or region are usually of high correlation, obtaining one segment's measurements is helpful to deduce the traffic states of the other associated segments. The higher spatial correlation or prediction accuracy, the larger the information gain will be. In other words, the larger the indirect monitoring benefits will be.

Figure 28 gives two typical layout scenarios in which the indirect traffic monitoring benefits might arise. In those cases, the travel time information on segment  $k$  is directly measured by two AVI sensors at its two endpoints. The direct measurement of  $k$  can be used to deduce the real-time traffic state of segment  $l$ , which can either be overlapped with  $k$  or be non-overlapped but adjacent to  $k$ , based on some inner relationships between  $k$  and  $l$ . Therefore, although segment  $l$  is not directly monitored with sensors, we can still discern its real-time state evolution as long as we have reliable knowledge on another segment  $k$  with high state correlation with  $l$ . Consequently, additional monitoring benefits, which is named as indirect monitoring information gains, can be derived based on the indirect traffic monitoring process.

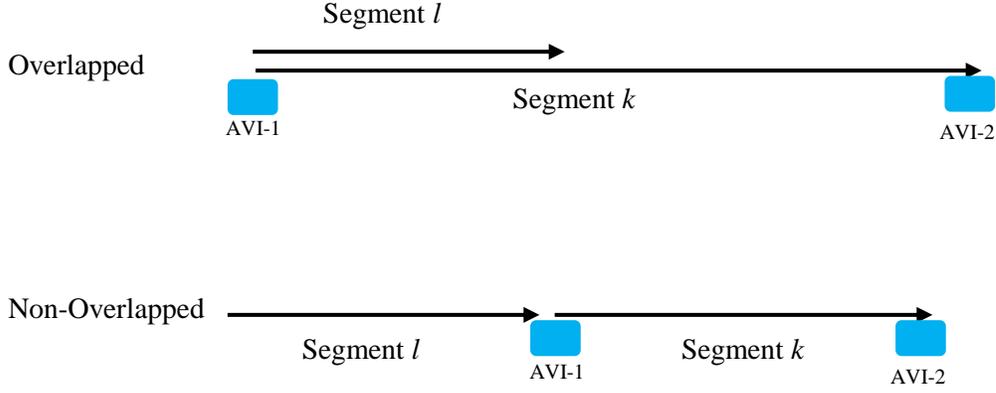


Figure 28: Two Typical Layout Configurations for Indirect Traffic State Estimation

Given the indirect monitoring based information gain  $B_{k,l}$  between any two correlated segments  $k$  and  $l$ , the total indirect monitoring based benefits can be calculated by the following linear Equation (4-1-2). Specifically, the total indirect monitoring benefit is the summation of the information gains of those segments that do not have direct measurement (i.e., neither sensor detection nor reliable probe report). The amplitudes of those information gains depend on the degree of the spatial information correlations. For example, if the traffic state on each segment is independent of other segments in the entire network,  $B_{total}^{indirect}$  will be zero. Further discussions and models to evaluate and determine the spatial information gain  $B_{k,l}$  will be provided in later chapters.

$$B_{total}^{indirect} = \sum_l \sum_{k \in I(l)} f_l \cdot B_{k,l} \cdot u_{k,l} \quad (4-1-2)$$

Therefore, the objective function can be obtained by maximizing the total sensor deployment benefits coming from both direct monitoring process and indirect statistical inference, which is given by Equation (4-1-3).

$$\text{Maximize : } B = \sum_l f_l \cdot B_l \cdot [(1 - s_l) y_l + (1 - y_l) s_l + \beta \cdot s_l y_l] + \sum_l \sum_{k \in I(l)} f_l \cdot B_{k,l} \cdot u_{k,l} \quad (4-1-3)$$

#### 4.1.3.3 Model Constraints

Specific constraints associated with the proposed optimization model are listed next. Those constraints are mainly divided into two categories. One is related to the budget limitation, and the other one is describing the relationships between decision variables and auxiliary variables.

$$\sum_i w_i \leq Q_{\text{number}} \quad (4-1-4)$$

$$\sum_i c_i w_i \leq B_{\text{money}} \quad (4-1-5)$$

Constraints (4-1-4) and (4-1-5) give the budget constraints on two levels. The first constraint indicates the maximum number of sensors that can be installed, and the second one describes the installation budget.  $Q_{\text{number}}$  is the maximum number of sensors available for placement.  $B_{\text{money}}$  represents the total monetary installation budget, and  $c_i$  denotes the installation cost at location  $i$ .

$$\sum_{k \in I(l)} u_{k,l} \leq (1 - a_l)(1 - y_l), \quad \forall l \quad (4-1-6)$$

$$u_{k,l} \leq 1 - (1 - a_k)(1 - y_k), \quad \forall l \text{ and } \forall k \in I(l) \quad (4-1-7)$$

Inequalities (4-1-6) and (4-1-7) together indicate when the decision variable  $u_{k,l}$  related to the indirect monitoring benefit, can be true. Specifically, constraint (4-1-6) means that the decision of inferring traffic states of any segment  $l$  based on other segments measurements can be made only if this segment is out of surveillance. Constraint (4-1-7) indicates a particular segment's measurement can be used to make the aforementioned spatial prediction only if this segment has surveillance.

$$2y_l \leq w_{s(l)} + w_{e(l)}, \quad \forall l \quad (4-1-8)$$

$$y_{l+1} \geq w_{s(l)} + w_{e(l)}, \quad \forall l \quad (4-1-9)$$

Constraints (4-1-8) and (4-1-9) indicate that for any segment of interest  $l$ , its traffic state information is directly measured by AVI sensing system if and only if two such sensors are installed at the endpoints of this segment.

$$y_l - y_k = 0, \quad \forall l, \forall k \in d(l) \text{ and } d(l) \neq \emptyset \quad (4-1-10)$$

In constraint (4-1-10),  $d(l)$  represents the dual segment of  $l$  (i.e.,  $l$  and  $d(l)$  have the same endpoints but reverse direction). The reason we provide this constraint is that for some AVI based sensing system, such as Bluetooth detectors, once a pair of detectors is deployed along a highway segment, they can detect and match the vehicles that pass for both directions. This property of such AVI monitoring system can save the deployment cost compared to the directional AVI monitoring system (e.g., toll station and license plate recognition).

#### 4.1.4 Model and Solution Method Discussion

The AVI sensor location problem is formulated as a linear binary integer programming model with the objective of maximizing the network-level traffic state monitoring benefits. There are three types of decision variables used to determine the sensor location. The first type of primary decision variable is at which location an AVI detector should be installed. The second class of primary decision variable is used to determine the additional monitoring benefits based on the spatial information correlations. Specifically, for segments without detectors, its spatial-correlated segments with detectors are evaluated and chosen to predict the traffic states of those out-of-surveillance segments. The third one is auxiliary variable for the ease of linearizing the model and indicates which segment is monitored. The total number of decision variables are

$|Z| + \sum_l |I(l)| + |L|$ , which is the summation of those three types of decision variables. The

total number of constraints is  $3|L| + \sum_l [|I(l)| + |d(l)|] + 2$ . For a fully connected and directed

network, if the target segments are exactly those arcs connecting the node sets, we could obtain

that  $|Z| = \frac{1}{2} |L| + 1$ . But for a real-world application scenario, the target monitoring segment can also be a path consisting of more than one consecutive arcs. Therefore, the problem size is largely determined by  $|L|$ , which is the number of target segments.

For any application with the proposed static location optimization model, as long as the number of target segments (i.e.  $|L|$ ) is not exponentially large (e.g., below million), the above binary programming model can be solved in minutes with commercial solvers, such as CPLEX and GUROBI. After all, the model is a static planning model and is not necessarily required to be solved in limited seconds for real-time usage. The solution time will be demonstrated and further discussed in later chapters with numerical experiments. Therefore, we omit the efforts to develop any customized algorithms or heuristics to solve this model.

## 4.2 Dynamic Planning Model for Sensor Location-Relocation Optimization

### 4.2.1 Background and Introduction

There are several factors potentially affecting the effectiveness of the optimal solution obtained from the static optimization model. Traditional sensor location optimization models in existing literature always consider the sensor deployment problem from a static perspective. In other words, all associated parameters affecting the placement effectiveness are assumed in advance and calculated in average. Consequently, the content of the solution is simply indicating at which location a specific sensor should be permanently installed once and forever. However, this might not guarantee the surveillance effectiveness since the limited resources (i.e., sensors) are permanently bundled at fixed locations. The highway system being studied is usually of high traffic fluctuations. Simply considering the system parameters as average values cannot comprehensively evaluate the network-level monitoring benefits. For example, traffic volume intensity and the probe data source reliability are both time-dependent. Suppose a highway

segment has high volumes in a specific period but extremely low volumes in other time periods, although the daily average volume on this segment is small, the variance of the volume is high. Consequently, the time-dependent travel time fluctuates to a large extent. Therefore, highway segments with higher volume averages are not necessarily more important than the ones with relatively smaller average volumes. As a consequence, the monitoring benefits for a given segment might fluctuate a lot among different time periods (e.g., time of the day, or day of the week). Therefore, given a limited surveillance budget, a dynamic traffic sensor monitoring network might bring more surveillance benefits than that of a static sensor network. Specifically, the functioning sensors can be located and relocated dynamically to maximize the total surveillance benefits for a given time horizon. In such a case, unnecessary waste of monitoring resources can be avoided. The emergence and rapid development of portable traffic detectors introduced in Chapter 2 make the sensor relocation issue much simpler than before. Therefore, developing a dynamic traffic sensor location optimization model for real-world applications becomes operationally possible.

In brief, this section reconsiders the sensor location problem from a dynamic perspective, and develops a more advanced optimization model to figure out the optimal traffic sensor locations in a time-dependent manner to maximize the surveillance benefit over a time horizon, given a limited number of portable traffic sensors (i.e. AVI configuration) and a limited probe-based real-time data stream. Also, the benefit loss during any relocation operation will be explicitly considered.

#### 4.2.2 Preliminaries

Associated preliminaries and necessary notations used to formulate the dynamic optimization model are provided in this section. First, a time-space network is introduced to depict the functioning process of the proposed dynamic sensor network. Next, the mathematical notations are defined and introduced.

#### 4.2.2.1 A Time-Space Network Representation

For the ease of depicting the dynamic process of traffic monitoring network, a time-space network structure is proposed. As is illustrated in Figure 29, the highway network is expanded across a given time horizon. Each node in the network represents an optional sensor installation location. The major difference from the static network is that a single location index no longer labels the location node in the time-space network. Instead, the node is uniquely labeled by a location index as well as a time index  $(n, t_m)$ . The first labeling component denotes the space location of a particular node, and the second component indicates the time stamp at which this node is being viewed. The reason for labeling a location with an additional time index is that the status of such a location (i.e., whether or not a sensor is placed at that location) might be time-dependent.

A dummy node  $(0, t^-)$  is introduced in the time-space network. This dummy node can be viewed as a “depot,” at which a set of sensors are stored. A dashed arc from this dummy node to any actual location indicates a sensor placement decision right before the start of the studied time horizon. During the studied time horizon, a solid arc connecting those time-space labeled nodes indicates an operational decision or action. Specifically, the arc connecting two nodes  $(x, t_{n-1})$  and  $(x, t_n)$  with the same location index  $x$  means the previously placed sensor at this location will continue to be placed at this location from time point  $t_{n-1}$  to  $t_n$ . In other words, this sensor will be staying and functioning at the same location for the next duration  $\tau_n = t_n - t_{n-1}$ . In the other case, an arc connecting two nodes  $(x, t_m)$  and  $(y, t_n)$  dictates a relocation operation of a sensor from location  $x$  to location  $y$  starting from time point  $t_m$  to  $t_n$  ( $n \geq m$ ). In the proposed time-space network, a sensor is installed at a location for a given period if and only if a path is routed through this period across the same location node.

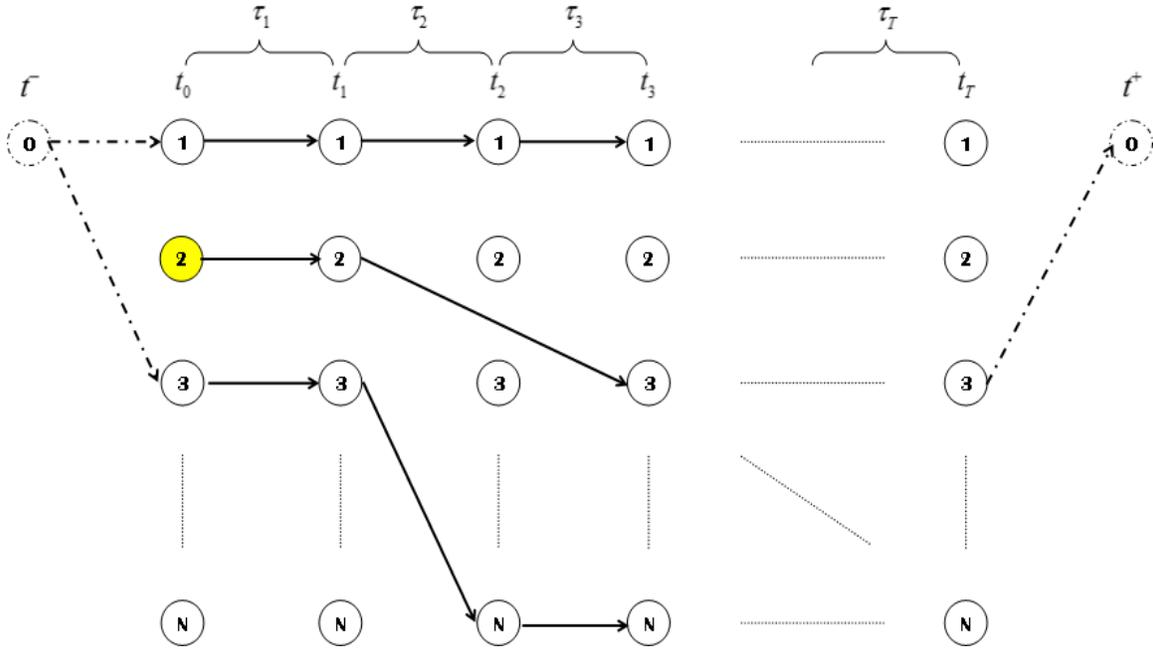


Figure 29: A Time-Space based Traffic Monitoring Network with Finite Sensors

Considering the possible existence of sensors installed in previous time horizons, an arc can be routed from a node without an inflow arc in the proposed time-space network, such as node  $(2, t_0)$ . In reality, a sensor might be uninstalled or under maintenance. To incorporate such cases into the optimization model, a second dummy node  $(0, t^+)$  is introduced and can be viewed as the sink node of the time-space network. For a dashed arc routed from time-space node  $(x, t_m)$  to the sink node, it means the sensor at location  $x$  will no longer be functioning in the network from time  $t_m$ .

#### 4.2.2.2 Notations

Table 6 lists the necessary notations used for the mathematical formulations provided in following sections.

Table 6: Notations for the Dynamic Sensor Location Model

<i>Sets and Subscripts</i>	
$G$	Highway network being studied.

---

$N = \{1, 2, 3, \dots\}$	Set of nodes representing intersections, interchanges, or intermediate locations of the highway network $G$ .
$A$	Set of arcs connecting those nodes of $N$ .
$l_i$	A highway segment, which can either be a single element of $A$ or be constituted by several consecutive arcs among $A$ .
$L = \{l_1, l_2, l_3, \dots, l_M\}$	Set of target traffic surveillance segments.
$l, k \in \mathcal{L}$	Segment index.
$i, j \in \mathcal{N}$	Node index.
$s(l)$	The starting point of segment $l$ . $s(l) \in \mathcal{N}, l \in \mathcal{L}$
$e(l)$	The ending point of segment $l$ . $e(l) \in \mathcal{N}, l \in \mathcal{L}$
$\Psi = \{t_0, t_1, \dots, t_T\}$	Set of time points discretizing the time-space network.
$\Gamma = \{\tau_1 = 1, \tau_2 = 2, \dots, \tau_T = T\}$	Set of discrete time slots constituting the time horizon.
$m, n \in \Psi$	Time point index.
$\tau \in \Gamma$	Time slot index with start and end time points equal to $t_{\tau-1}$ and $t_\tau$ , respectively.
$I(l)$	Set of segments, whose traffic states have potential to be used to predict the traffic state on segment $l$ .
$d(l)$	Set of dual segments of $l$ . A segment is defined as a dual segment of $l$ if and only if it shares the same endpoints with $l$ .
<u>Parameters</u>	
$\Delta t$	Duration of each discrete time slot. i.e. $t_\tau - t_{\tau-1} = \Delta t$ .
$f_l^\tau$	Traffic volume of segment $l$ during the time period $\tau$ .
$B_l^\tau$	Surveillance benefit if the traffic on segment $l$ is directly monitored during the time period $\tau$ .
$B_{k,l}^\tau$	Surveillance benefit gains (e.g., traffic state prediction uncertainty or error reduction) of segment $l$ during the time period $\tau$ , when it is not directly monitored, but the traffic

---

---

	information on segment $k$ is collected and used to infer $l$ 's information.
$a_l^\tau$	Binary parameter. Equal to 1 if segment $l$ has reliable probe data sources during the time period $\tau$ ; equal to 0, otherwise.
$r_i^{t_0}$	Binary parameter. Equal to 1 if a portable sensor is already installed at location $i$ at the beginning of the planning horizon; equal to 0, otherwise.
$r_i^{\tau}$	Binary parameter. Equal to 1 if a portable sensor should be installed at location $i$ at the end of the planning horizon (for future planning purpose); equal to 0, otherwise.
<u><i>Decision Variables</i></u>	
$w_i^\tau$	Binary variable equal to 1 if a portable sensor is installed at location $i$ during the period $\tau$ ; equal to 0, otherwise.
$y_l^\tau$	Binary variable equal to 1 if segment $l$ is directly monitored by two portable sensors during the period $\tau$ ; equal to 0, otherwise.
$u_{k,l}^\tau$	Binary variable equal to 1, if segment $k$ is chosen to predict the traffic state of $l$ given that $l$ is without surveillance during the period $\tau$ ; equal to 0, otherwise.
$x_{i,j}^{m,n}$	Binary variable equal to 1, if a portable sensor is relocated from location $i$ to $j$ , starting at time stamp $m$ and finishing at time stamp $n$ ( $n > m$ ); equal to 0, otherwise.

---

### 4.2.3 Mathematical Formulation

#### 4.2.3.1 Model Objective

The objective function of the proposed dynamic sensor location optimization model is given by Equation (4-2-1). The format of this objective function is the same as that of the static version of the optimization model. The only difference is that objective function of the dynamic version is taking into account time dimension. Specifically, the summation of the total monitoring benefits consists of the potential monitoring benefit of each segment during each discrete time period.

$$\begin{aligned}
\text{Maximize : } & \sum_{\tau} \sum_l f_l^{\tau} \cdot B_l^{\tau} \cdot \left[ \beta a_l^{\tau} y_l^{\tau} + (1 - a_l^{\tau}) y_l^{\tau} + a_l^{\tau} (1 - y_l^{\tau}) \right] \\
& + \sum_{\tau} \sum_l \sum_{k \in l(l)} f_l^{\tau} \cdot B_{k,l}^{\tau} \cdot u_{k,l}^{\tau}
\end{aligned} \tag{4-2-1}$$

It is noted that the weighting parameter in Equation (4-2-1), i.e., traffic volume, is considered time-dependent. This is important for the decision determination since traffic volume of a particular highway segment is usually of high variation. Considering the traffic flow fluctuation effect can more accurately evaluate the total surveillance benefit, and consequently, the sensor resources can be fully used. Also, the surveillance benefits  $B_l^{\tau}$  and  $B_{k,l}^{\tau}$  are considered as time-dependent as well. This is because the detection and estimation accuracy of the sensing system might differ across different time periods. For example, for a particular roadway monitored with two AVI-based travel time collection sensors, the prediction accuracy or uncertainty might differ a lot for peak and nonpeak period.

Constraints of the binary programming optimization model are classified and separately provided in the next three subsections. The first set of constraints describes the surveillance coverage issues and the relationships among major decision variables and auxiliary variables, which are used to linearize the formulation. The second type constraints talk about the feasibility issues in the sensor location-and-relocation process, which is depicted by the proposed time-space network. Finally, budget constraints are given in terms of the installation and operation costs.

#### 4.2.3.2 Surveillance Coverage Constraints

$$2y_l^{\tau} \leq w_{s(l)}^{\tau} + w_{e(l)}^{\tau}, \quad \forall \tau, \forall l \tag{4-2-2}$$

$$y_l^{\tau} + 1 \geq w_{s(l)}^{\tau} + w_{e(l)}^{\tau}, \quad \forall \tau, \forall l \tag{4-2-3}$$

Constraints (4-2-2) and (4-2-3) indicate that the traffic information of a segment is directly collected by an AVI sensing system during a particular period  $\tau$  if and only if two detectors are placed at both endpoints of this segment.

Considering the local correlations of traffic states between different highway segments across the network, the surveillance benefits arising from indirect traffic state prediction (i.e., spatial traffic state prediction) are incorporated into the objective function. Two set of inequalities (4-2-4) and (4-2-5) are defined to determine when the indirect traffic monitoring benefit of a particular segment can be considered. In detail, constraint (4-2-4) states that for a particular period, the traffic information of a specific segment can be used to predict other segments' information only if this segment is under surveillance at this time period. The surveillance either comes from reliable probe data sources or is based on the deployed sensors. On the other hand, constraint (4-2-5) indicates when the spatial prediction can be applied. Specifically, spatial traffic state prediction can be applied to a specific segment only if this segment is out of surveillance. In other words, if a segment is with reliable surveillance, the system will directly infer its real-time traffic states, which is assumed to be more reliable and accurate.

$$u_{k,l}^{\tau} \leq 1 - (1 - a_k^{\tau})(1 - y_k^{\tau}), \quad \forall \tau, \forall l \text{ and } \forall k \in I(l) \quad (4-2-4)$$

$$\sum_{k \in I(l)} u_{k,l}^{\tau} \leq (1 - a_l^{\tau})(1 - y_l^{\tau}), \quad \forall \tau, \forall l \quad (4-2-5)$$

For any zone-based AVI system (e.g., Bluetooth), once two detectors are placed at both endpoints of a particular highway segment, the traffic data of both directions can be collected. This consideration is illustrated by constraint(4-2-6), in which  $l$  and  $k$  are considered as dual segment of each other.

$$y_l^{\tau} - y_k^{\tau} = 0, \quad \forall \tau, \forall l, \forall k \in d(l) \text{ and } d(l) \neq \emptyset \quad (4-2-6)$$

#### 4.2.3.3 Facility Routing Feasibility Constraints

Constraints associated with the dynamic location-and-relocation operations on the sensor network are listed as below.

$$x_{0,i}^{\tau, t_0} + r_i^{t_0} \leq 1, \quad \forall i \in N \quad (4-2-7)$$

$$\sum_{j \in N} \sum_{\substack{n \in \Psi \cup \\ n > m}} x_{i,j}^{m,n} + x_{i,0}^{m,t^+} = \sum_{j \in N} \sum_{\substack{n \in \Psi \\ n < m}} x_{j,i}^{n,m} + x_{0,i}^{t^-,m}, \quad \forall i \in N, \forall m \in \Psi \setminus \{t_0, t_T\} \quad (4-2-8)$$

$$\sum_{j \in N} \sum_{\substack{n \in \Psi \\ n > t_0}} x_{i,j}^{t_0,n} + x_{i,0}^{t_0,t^+} = x_{0,i}^{t^-,t_0} + r_i^{t_0}, \quad \forall i \in N \quad (4-2-9)$$

$$\sum_{j \in N} \sum_{m \in \Psi \setminus \{t_T\}} x_{j,i}^{m,t_T} + x_{0,i}^{t^-,t_T} \geq r_i^{t_T}, \quad \forall i \in N \quad (4-2-10)$$

$$\sum_{j \in N} \sum_{m \in \Psi \setminus \{t_T\}} x_{j,i}^{m,t_T} + x_{0,i}^{t^-,t_T} \leq 1, \quad \forall i \in N \quad (4-2-11)$$

Constraints (4-2-7) to (4-2-11) guarantee facility (i.e., portable sensor) flow conservation in the time-space network, as is demonstrated in Figure 29. In detail, inequality constraint (4-2-7) indicates at the beginning of the planning horizon; a sensor can be considered to be deployed at location  $i$  only if this location has no sensor. In other words, at most one sensor can be installed at a location. Equality constraint (4-2-8) indicates that for any location at a particular intermediate time point  $m$ , the number of sensors routed out from this location after time  $m$  should be equal to the number of sensors routed into this location right before  $m$ . From the time-space network perspective, this constraint guarantees the temporal-spatial movement consistency of any sensor within the network. Constraints (4-2-12) give the sensor routing rules at the beginning of the time-space network. It indicates that at the beginning of the time-space network, for any location  $i$  covered with a sensor (either previously installed or newly installed), the sensor can be either continuously functioning at this location or relocated from this location to other locations in the following time period(s). Similarly, constraints (4-2-10) and (4-2-11) provide the boundary sensor routing rules at the end of the time-space network. Constraint (4-2-10) indicates that for any location  $i$  requiring a sensor installation at the end of the planning horizon, a sensor must be either routed to this location from other locations or continuously held at this location if this location has an installed sensor since the last time period. This constraint is proposed for practical purposes, such as for the ease of future planning horizon. Constraint (4-2-11) indicates the total

number of sensors routed to any location  $i$  at the end of the planning horizon should not exceed one given each location can only have at most one sensor deployed.

A location is monitored by a sensor during the period  $\tau$ , if and only if a sensor has been already placed at this location at the beginning of  $\tau$  and no relocation action is taken during this period. This can be stated by the following equality constraint (4-2-13), in which  $n - m = \tau$ .

From the time-space perspective, an arc being routed from node  $(i, t_{\tau-1})$  to  $(i, t_{\tau})$  means location  $i$  is with surveillance for the period  $\tau$ .

$$w_i^{\tau} - x_{i,i}^{m,n} = 0, \quad \forall i \in N, \forall \tau \in \Gamma, m = t_{\tau-1}, n = t_{\tau} \quad (4-2-13)$$

Moreover, practical concerns regarding relocation operations are also considered by directly manipulating the decision variables. In constraint (4-2-14), infeasible relocation operations are excluded by considering the minimal relocation time from location  $i$  to location  $j$ . If the relocation time  $D_{i,j}(m)$  is longer than  $(n-m)$ , the relocation operation cannot take place. In some cases, the relocation time can be simply assumed as the sensor setup time plus travel time between those two locations.

$$x_{i,j}^{m,n} = 0, \quad \forall i, \forall j \neq i, \forall (m, n) \text{ and } (n - m) < D_{i,j}(m) \quad (4-2-14)$$

#### 4.2.3.4 Budget Constraints

In total, there are three types of budget issues that should be considered when dealing with the proposed dynamic sensor network optimization problem. They are, capital cost, operation cost for relocation, and practical concerns in terms of relocation frequency, respectively.

$$\sum_{j \in N} \sum_{n \in \Psi} x_{0,j}^{\tau,n} \leq Q \quad (4-2-15)$$

Constraint (4-2-15) gives the capital cost budget in the form of restricting the number of sensors that are going to be deployed in the network for an upcoming time horizon. The number of deployed sensors cannot exceed the maximum number of sensors available in the depot (i.e.,  $Q$ ) at the beginning of the planning time horizon. This constraint is given with quantity format.

But it can be easily converted to monetary format by multiplying both sides of the inequality constraint by the unit price.

$$\sum_{\forall j \in N} \sum_{\forall n \in \Psi} c_{0,j}^{deploy} \cdot x_{0,j}^{-,n} + \sum_{\forall i \in N} \sum_{\substack{\forall j \in N \\ j \neq i}} \sum_{\forall m \in \Psi} \sum_{\substack{\forall n \in \Psi \\ n > m}} c_{i,j}^{reloc} \cdot x_{i,j}^{m,n} + \sum_{\forall i \in N} \sum_{\forall m \in \Psi} c_{i,0}^{back} \cdot x_{i,0}^{m,+} \leq B_{cost}$$

(4-2-16)

The budget constraint in terms of sensor installation and relocation operations for the entire planning time horizon is given by inequality constraint (4-2-16), where  $c_{i,j}$  denotes the installation-relocation cost from location  $i$  to  $j$ , and  $B_{cost}$  stands for the total budget. Constraint (4-2-16) consists of both the initial installation cost (from the depot to any actual location) and the cost of the subsequent relocation (across the actual locations). This is rational in real-world scenarios since both installation and relocation operations require human efforts to be completed. The cost heterogeneity among those site-specific operations can be reflected by the index-specific cost parameter  $c_{i,j}$ .

$$\sum_{\forall i \in N} \sum_{\substack{\forall j \in N \\ j \neq i}} \sum_{\forall m \in \Gamma^p} \sum_{\forall n > m, n \in \Gamma^p} x_{i,j}^{m,n} \leq B_{move}^{\Gamma^p}, \forall \Gamma^p \quad (4-2-17)$$

Constraint (4-2-17) provides the budget limitation with respect to the relocation frequency during a given time period  $\Gamma^p$ , which is the  $p^{th}$  element of the partition of the entire planning horizon  $\Gamma$ . This constraint indicates that for a particular period (e.g., daytime or nighttime), the total number of relocation operations cannot exceed the predetermined frequency  $B_{move}^{\Gamma^p}$  due to the practical operational issues. For example, during the night time,  $B_{move}^{\Gamma^p}$  might be set to zero since the fleet operators are off duty. Determination of the partition of  $\Gamma$ , as well as the corresponding relocation frequency limit of each period depends on the characteristics of the actual problem, such as time horizon being studied, network size and the patrol size of the field operation team.

#### 4.2.3.5 Decision Variables

Decision variables related to the above formulations are defined and listed as below. As is shown, all associated decision variables and auxiliary variables are binary. The decision variables declared in Equations (4-2-18), (4-2-19) and (4-2-20) are the same as those of the static optimization model except a second dimensional time index  $\tau$  is added. Statements (4-2-21) and (4-2-22) give the decision variables uniquely used in the proposed dynamic optimization model. Decision variables defined in (4-2-21) represent whether relocation operations take place, whereas variables defined in (4-2-22) indicate whether a particular sensor will continue to be held in a specific location.

$$w_i^\tau \in \{0,1\}, \forall i \in N, \forall \tau \in \Gamma \quad (4-2-18)$$

$$y_l^\tau \in \{0,1\}, \forall l \in L, \forall \tau \in \Gamma \quad (4-2-19)$$

$$u_{k,l}^\tau \in \{0,1\}, \forall l \in L, \forall k \in I(l), \forall \tau \in \Gamma \quad (4-2-20)$$

$$x_{i,j}^{m,n} \in \{0,1\}, \forall i, j \in \{0\} \cup N, j \neq i; \forall m \in \Psi \cup \{t^-\}, \forall n \in \Psi \cup \{t^+\}, n > m \quad (4-2-21)$$

$$x_{i,i}^{m,m+1} \in \{0,1\}, \forall i \in N, \forall m \in \Psi \setminus \{t_T\} \quad (4-2-22)$$

It is noted that the index domain of  $x_{i,j}^{m,n}$  are provided separately in (4-2-21) and (4-2-23), since only meaningful  $x_{i,j}^{m,n}$  decision variables are declared. If  $x_{i,j}^{m,n}$  are declared without considering the infeasible relationships among the four indexes, both the number of decision variables and the number of constraints will increase. For example, if  $x_{i,j}^{m,n}$  is declared in one single statement for any combination of  $(i,j,m,n)$ , the number of  $x_{i,j}^{m,n}$  decision variables in the model will be doubled. Moreover, additional infeasibility elimination constraints are required, such as Equation (4-2-24).

$$x_{i,i}^{m,n} = 0, \quad \forall i \in N, \forall m \in \Psi \setminus \{t_{T-1}, t_T\} \text{ and } \forall n \in \Psi \setminus \{m+1 \cdot \Delta t\} \quad (4-2-24)$$

#### 4.2.4 Model Complexity Analysis

For a given problem configuration, the total numbers of decision variables  $w_i^r$ ,  $y_i^r$ ,  $u_{k,l}^r$ , and  $x_{i,j}^{m,n}$  in the proposed formulation are,  $|\Gamma| \cdot |N|$ ,  $|\Gamma| \cdot |L|$ ,  $|\Gamma| \sum_{l \in L} |I(l)|$ , and  $|N+1| |N| \left\lceil \frac{1}{2} |\Psi+1| (|\Psi+1|-1) \right\rceil + |N| |\Psi|$ , respectively. As is discussed in subsection 4.2.3, we only declared the  $x_{i,j}^{m,n}$  variables with  $m < n$  to decrease the total number of decision variables and avoid additional constraints. Even in this case, the total number of decision variables of the dynamic model is extremely larger than that of the static model given in 4.1. This is mainly due to the introduction of the four-index decision variable  $x_{i,j}^{m,n}$ , the size of which is directly proportional to the square of the time-space network time scale. From another perspective, the problem formulated above can be viewed and converted as a vehicle routing problem (VRP) on a time-space network. This means the computation time to solve the problem will exponentially increase as the problem size linearly increases since VRP is an NP-hard problem. As a consequence, when the formulation is applied to a moderate size real-world scenario, it is likely impossible to obtain the optimal solution with any cutting-edge commercial MIP solver. Therefore, specific algorithms should be customized to resolve the problem to obtain a satisfactory solution or close-to-optimal solution.

#### 4.2.5 Valid Integer Cuts

The model complexity is mainly attributed to the four-index decision variable  $x_{i,j}^{m,n}$  representing relocations operations. The number of such variables substantially increases as the time-space network size increases. At the optimal condition, large portions of these variables are not active (i.e., they are equal to 0). We propose the following integer cuts to eliminate such redundant feasible solutions from the time-space solution pool.

**Valid Cut:** For sensor relocation operation between any pair of nodes  $i$  and  $j$  in the time-space network, if conducting relocation from time stamp  $m$  to time stamp  $n$  is feasible, i.e.  $x_{i,j}^{m,n} \in \{0, 1\}$ , then  $x_{i,j}^{m,t^*} = 0, \forall t^* > n$  is a valid cut to the optimal solution.

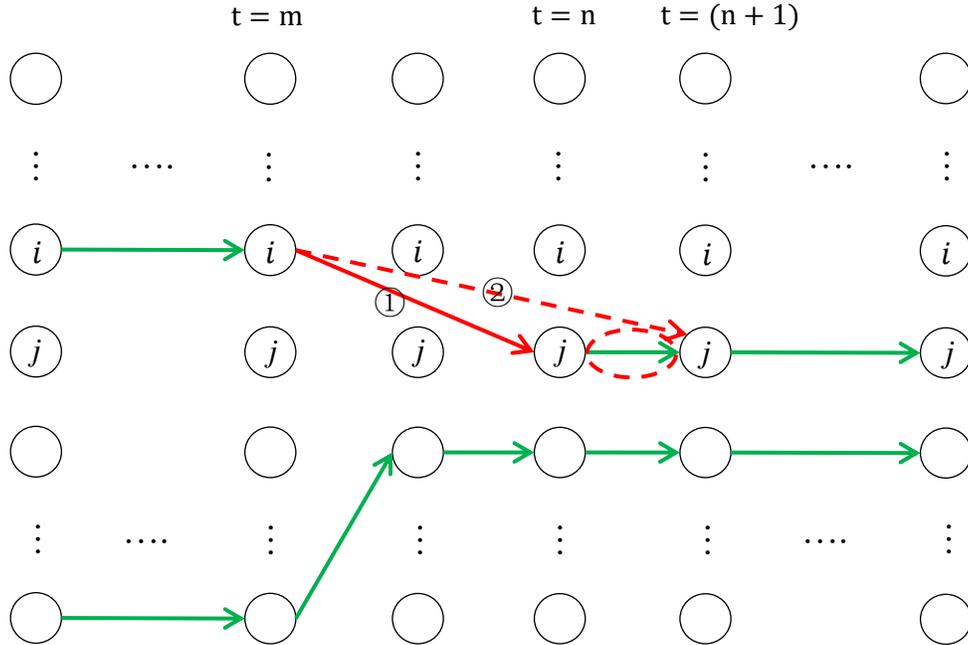


Figure 30: Illustration of Valid Cuts Regarding Relocation Operations.

Proof of the validity of the cuts is obvious due to the greedy property of the objective function and can be illustrated by the time-space network shown in Figure 30. Without loss of generality, suppose operation ① (i.e., the solid red arrow) and operation ② (i.e., the dashed red arrow) are two feasible relocations with the same starting time stamp (i.e.,  $t = m$  in the illustrative graph) between location  $i$  and  $j$ . The concept of the proposed cut indicates that if operation ① is feasible, then operation ② will not appear in the optimal solution. The validity of the cuts holds if and only if the contribution of operation ② to the objective function is less than or equal to the contribution of operation ① to the objective function. As is demonstrated in the time-space network, this is always valid if these two relocation operations have the same starting time stamp. In other words, if a relocation starting from time  $t = m$  can be completed at time  $t =$

n, there is no additional benefit to complete the relocation at any time  $t^* > n$  since the later one will result in surveillance loss comparing with the former one. Therefore, for any feasible relocation operation  $x_{i,j}^{m,n} \in \{0,1\}$ , forcing  $x_{i,j}^{m,t^*} = 0 \forall t^* > n$  will not affect the optimal solution.

### 4.3 Applications to DC-Baltimore Commuting Network

#### 4.3.1 Real-world Case Study Preliminaries

The Washington D.C.-Baltimore commuting network is selected for this case study. As is shown in Figure 31, the network consists of three major corridors connecting Washington D.C. region and Baltimore region. The three corridors are arterial US-29 (i.e., node 10 to node 01), freeway I-95 (i.e., node 19 to node 11), and freeway MD-270 (i.e., node 28 to node 20). Those three northern-southern corridors are connected by various arterials and freeways, such as MD-100 (i.e., link 0213) and MD-200 (i.e., link 0810). As is shown in Figure 31(b), each node is labeled by a two-digit numerical code and is the optional location for installing a traffic sensor. The nodes labeled in the network might be either an interchange or a major arterial intersection. The selection of those types of nodes as the potential sensor installation location is because those major interchanges and intersections mainly partition the traffic flow distribution heterogeneity. In other words, these interchanges and major arterial intersections cut the entire commuting network into various short corridors, the traffic information of which might be of interest.

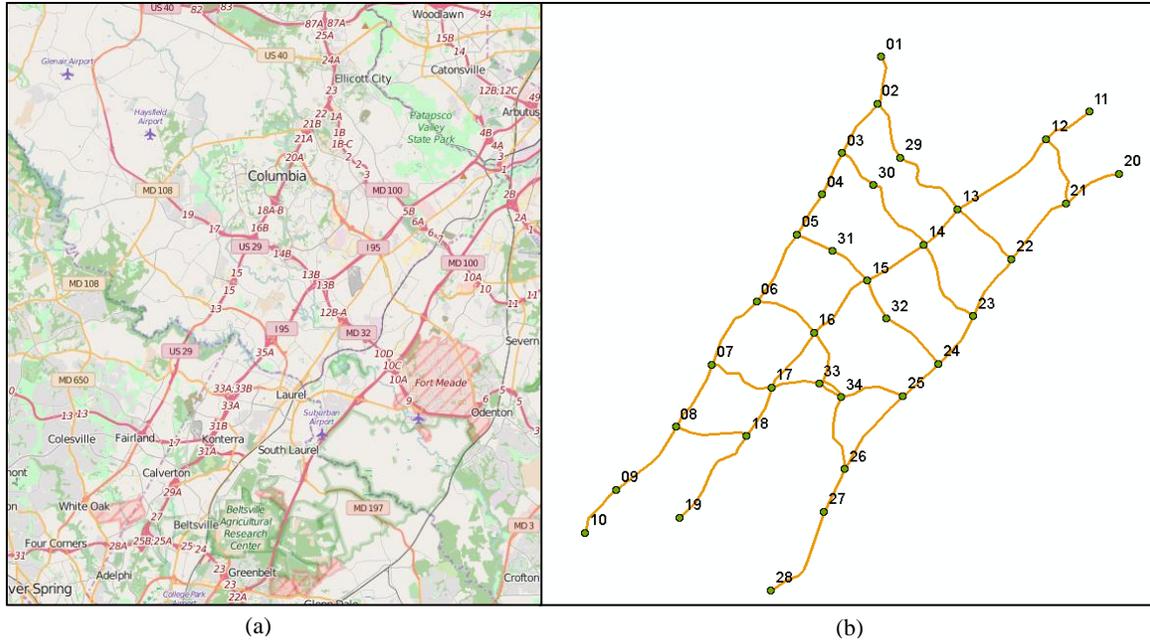


Figure 31: Case Study Network: (a) Map View of the Target Network (i.e., Commuting Network between Washington D.C. and Baltimore); (b) Extracted and Abstracted Network.

#### 4.3.2 Travel Time Data Description and Statistical Analysis

Temporal-spatial traffic state prediction performance (e.g., error improvement or uncertainty reduction) for all the network segments are key parameters required in the optimization model. These parameters can be either empirically assumed or inferred through some well-designed simulation processes. But they are still prone to large dispersion from the real-world scenarios and may not accurately reflect the real-world prediction performance. Therefore, it is more reliable to use the real-world data to estimate optimization parameters and generate the sensor deployment strategies.

In this study, real-world travel time and traffic speed data provided by INRIX Inc. are used to study the temporal-spatial traffic state prediction performance for each corridor segment in the case study network. A three-month period (i.e., May 2016 to July 2016) of travel time and traffic speed data were obtained for the mentioned estimation purpose.

INRIX makes use of a common industry convention known as Traffic Message Channel location code (TMC) to report travel time and travel speed data on freeways and major arterials.

Such codes are developed and maintained by electronic map database vendors to define specific roadway segments uniquely. The temporal resolution of the travel time report is 1 minute.

Considering each target segment may consist of multiple TMCs, a backtracking algorithm was applied to calculate the experienced travel time of each segment based on the travel time reports on each component TMC (Zhang, Hamedi, and Haghani 2015). INRIX provides a confidence score for the archived travel time and speed information for each time interval. Specifically, a confidence score of 30 indicates the reported data is very reliable, 20 means the data reliability is fair, and 10 implies the data is unreliable. For this study, only travel time data with a score of 30 was used, and the rest of the data reports were filtered out.

Figure 32 and Figure 33 display travel time variation on each segment at different time periods. Specifically, Figure 32 presents an overall view of the travel time variation calculated by the coefficient of variation (CV) for the entire study network at different times of the day (i.e., AM peak and PM peak) and different days of the week (i.e., weekday and weekend). As are displayed by the four heat maps, travel time variation on a particular segment exhibits different patterns at different time periods. For example, during the weekend, it indicates the traffic state on the south part of I-295 has significantly higher fluctuations than other regions. While on weekdays, south parts of I-95 and MD-100 have more unreliable travel time experience for drivers.

Figure 33 spatially gives the travel time CV distributions in a higher time-of-day resolution. As are displayed by the four subplots in Figure 33, one can see the CV for each segment differs to some notable extent by time within the morning period. Travel time uncertainty is much higher on I-95 (i.e., between interchange 15 and interchange 19) during the period 10:00-11:00 AM. The Laurel region (i.e., around nodes 33 and 34) tends to have higher unreliable travel time at noon (i.e., 11:00-12:00 AM).

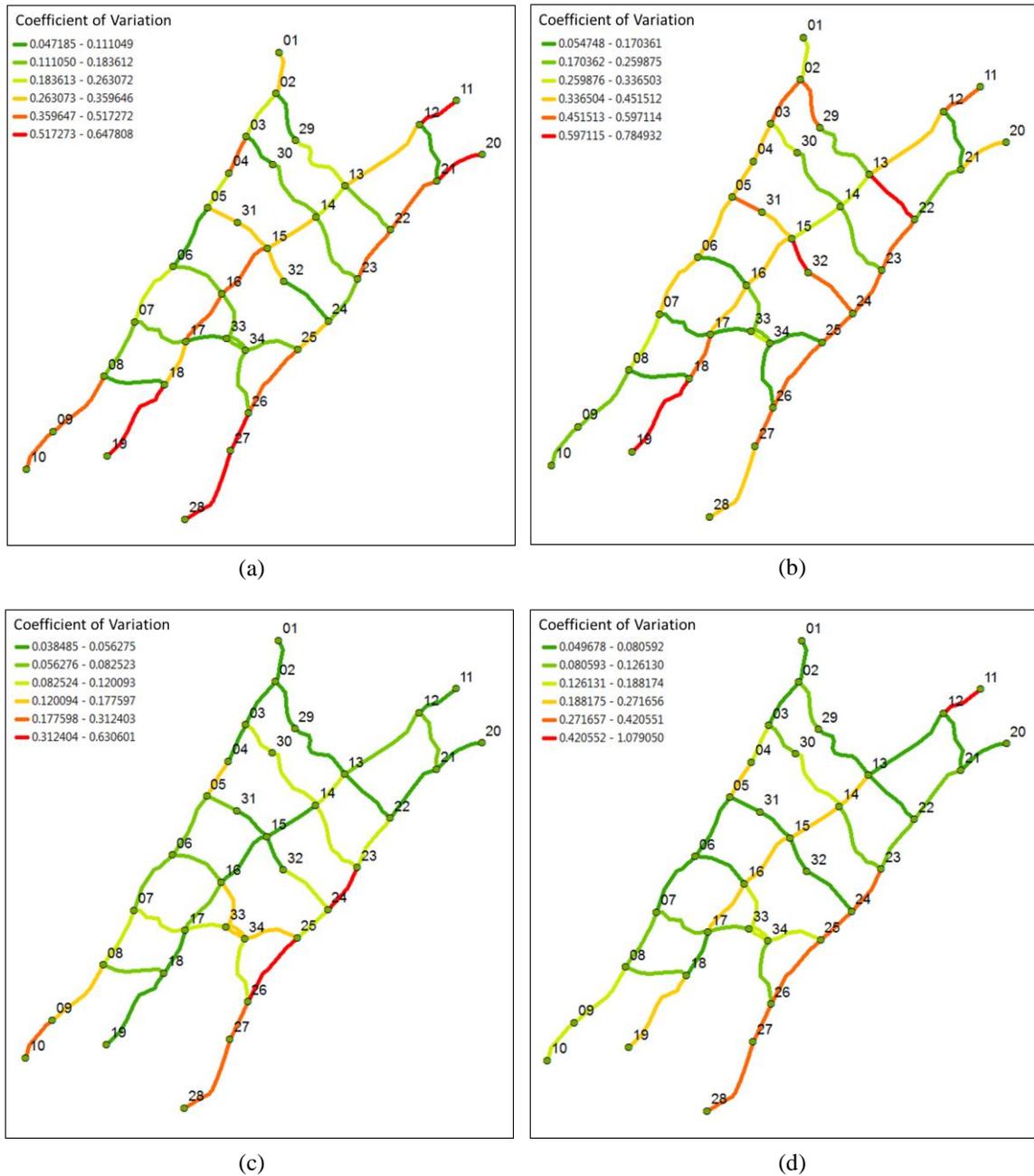


Figure 32: Spatial Distributions of Travel Time Coefficient of Variations in the Study Network: (a) Weekday Morning Peak Period; (b) Weekday Afternoon Peak Period; (c) Weekend Morning Peak Period; (d) Weekend Afternoon Peak Period.

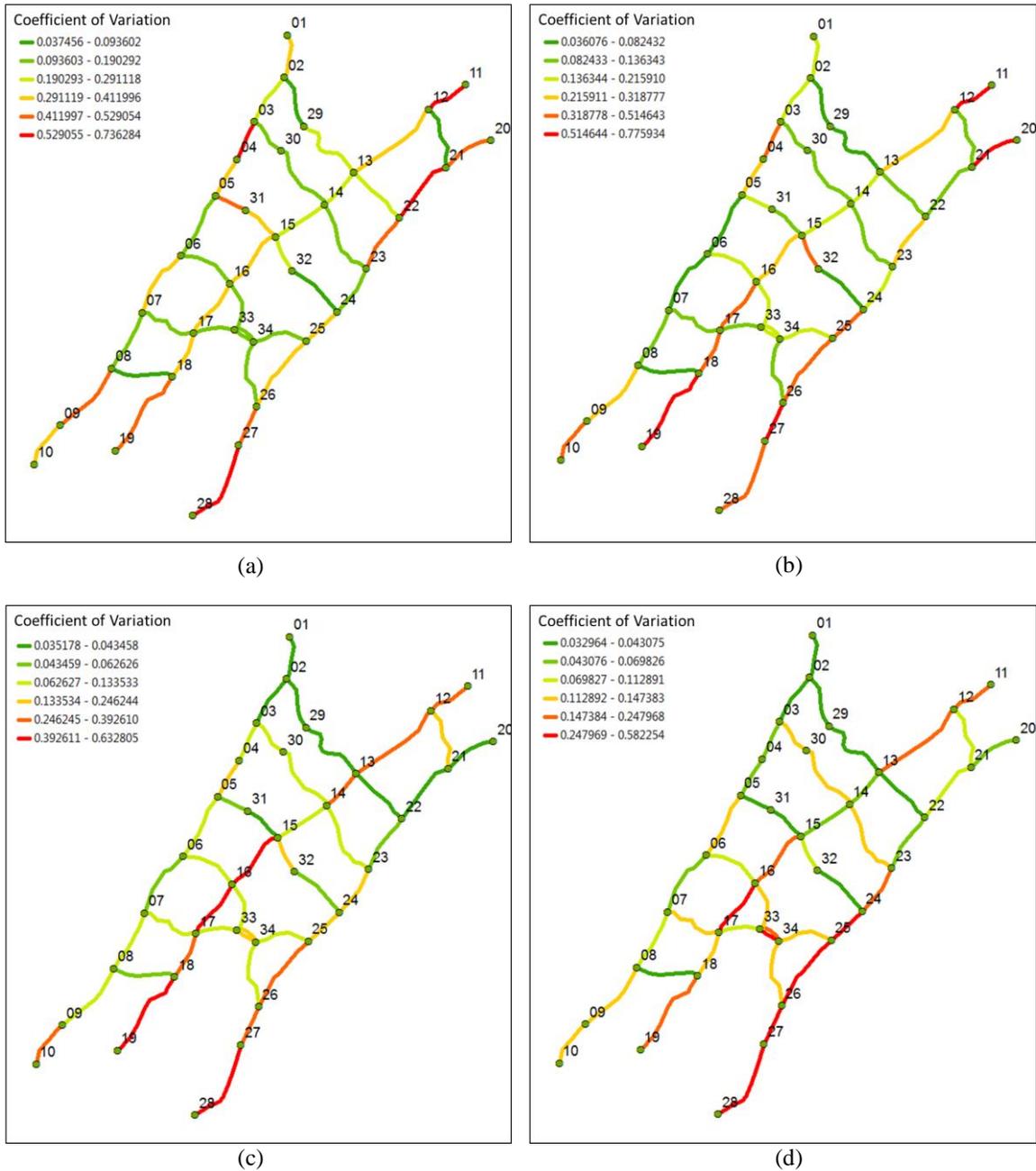


Figure 33: Time-dependent Spatial Distributions of Travel Time Coefficient of Variations in Weekday: (a) 8:00-9:00 AM; (b) 9:00-10:00 AM; (c) 10:00-11:00 AM; (d) 11:00-12:00 AM.

Travel time data visualizing the above traffic state variations are used to estimate the prediction performance parameters later for the optimization model. The above statistical analysis, as well as the empirical analysis conducted in Chapter 3, demonstrate the fact that traffic state prediction performance is not only location dependent but also time dependent. Therefore,

this fact should be realized and considered when designing real-time traffic surveillance and prediction system with a particular set of sensors.

#### 4.3.3 Traffic Volume Data Extraction and Imputation

Traffic volume information is important in the proposed sensor optimization model. It plays a key role in estimating total errors that can be reduced for the entire transportation network for a given time horizon. As is formulated in the objective function of the optimization model, traffic volume for each segment in each particular time period is a parameter used for weighing the total prediction errors reduction. In other words, traffic volume intensity and prediction error together determine the surveillance benefit and preference for a specific corridor segment.

Specifically, a highway segment with both higher traffic volume and reliable travel time notification (by inducing real-time surveillance system) indicates a higher preference for deploying such a real-time surveillance system. Because introducing such a surveillance system benefits travelers a huge amount. Instead, if the segment has very few traffic demand, even if deploying real-time surveillance system can result in travel time information accuracy improvement, the deployment preference may still be low since the benefit is relatively small.

In this research, annual average weekday daily traffic (AAWDT) counts and annual average weekend daily traffic (AAWNDT) counts of year 2016 reported by Maryland State Highway Administration (MDSHA) was used to approximate the traffic volume for each studied highway segments. The spatial distributions of the AAWDT and AAWNDT over the study network are displayed in Figure 34. As is well known, the traffic volume for a specific segment always varies a lot across a day. Knowing the volume information in a higher temporal resolution can be useful and practical for planning a both a static and dynamic sensor network. For example, for a static sensor network, usually the peak hour flow is taken into account, and for a dynamic sensor network, the optimization model requires the time-dependent volume information. For this study, we redistributed the average daily traffic volume into hourly basis based on the empirical

average traffic volume distributions published in the 2015 Urban Mobility Scorecard (Schrank et al. 2015). The temporal traffic volume distribution parameters for weekday and weekend are displayed in Figure 35 and Figure 36, respectively, and are further classified based on the facility type (i.e. freeway or arterial) and peak period (i.e. morning peak or afternoon peak). Further, detail parameters values are provided in Table 7 and

Table 8 with respect to the distributions displayed in Figure 35 and Figure 36.

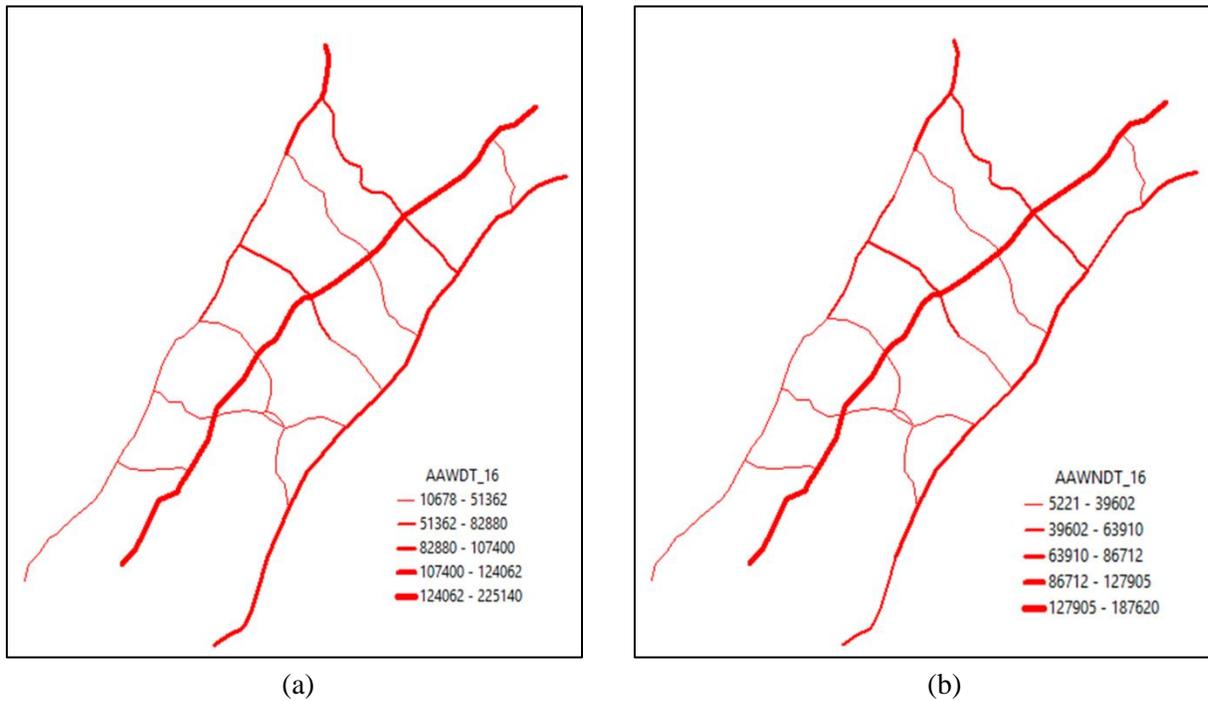


Figure 34: 2016 Annual Average Daily Traffic Counts Distribution in the Case Study Network: (a) Annual Average Weekday Daily Traffic (AAWDT); (b) Annual Average Weekend Daily Traffic (AAWNDT).

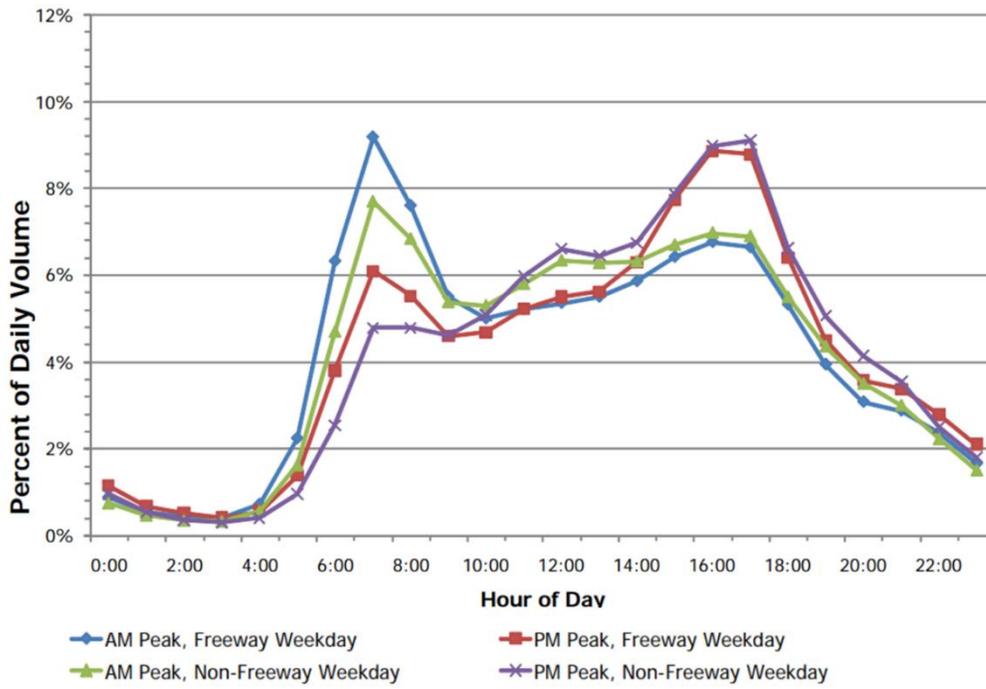


Figure 35: Time-dependent Weekday Average Traffic Volume Distribution w.r.t. Different Roadway Types (Schrank et al. 2015).

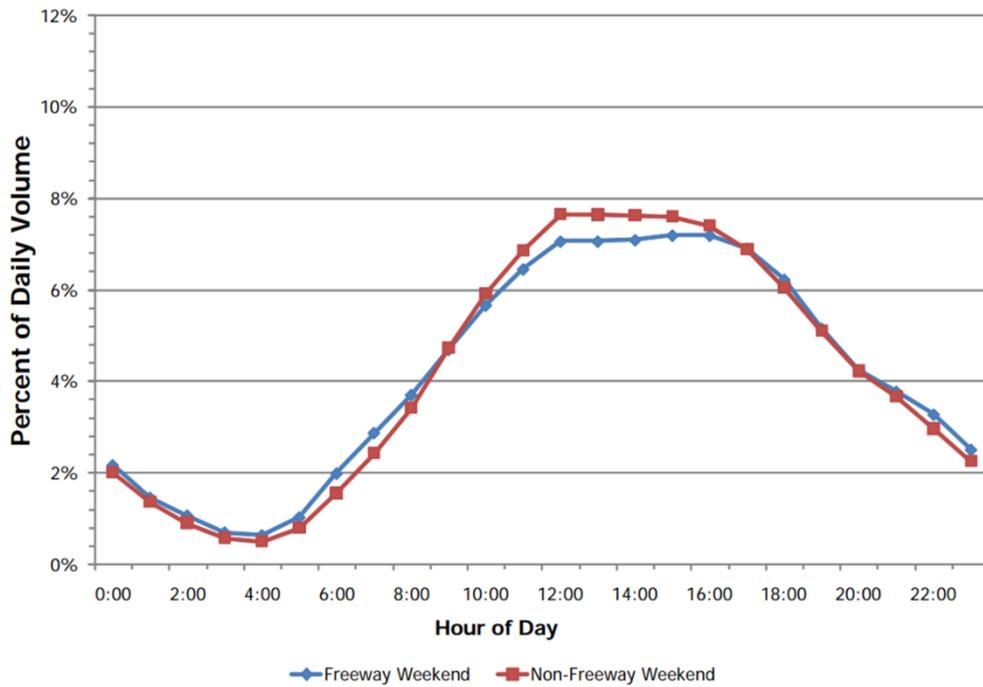


Figure 36: Time-dependent Weekend Average Traffic Volume Distribution w.r.t. Different Roadway Types (Schrank et al. 2015).

Table 7: Weekday Average Traffic Volume Distribution Probabilities at Different Time of Day under Different Roadway Types (Schrank et al. 2015).

Time of Day	Weekday				Weekend	
	Freeway AM Peak	Freeway PM Peak	Arterial AM Peak	Arterial PM Peak	Freeway	Arterial
0:00-1:00	1.6%	1.6%	0.8%	0.8%	2.1%	2.0%
1:00-2:00	0.8%	0.8%	0.5%	0.5%	1.5%	1.5%
2:00-3:00	0.5%	0.5%	0.3%	0.3%	1.1%	1.0%
3:00-4:00	0.5%	0.5%	0.2%	0.2%	0.9%	0.8%
4:00-5:00	0.8%	0.7%	0.5%	0.4%	0.8%	0.7%
5:00-6:00	2.4%	1.8%	1.5%	1.2%	1.2%	1.0%
6:00-7:00	6.4%	3.8%	4.8%	2.6%	2.0%	1.6%
7:00-8:00	9.2%	6.1%	7.6%	4.8%	3.0%	2.4%
8:00-9:00	7.6%	5.6%	6.9%	4.8%	3.5%	3.3%
9:00-10:00	5.2%	4.8%	5.2%	4.8%	4.6%	4.5%
10:00-11:00	5.0%	4.8%	5.2%	5.0%	5.6%	6.1%
11:00-12:00	5.2%	5.2%	5.8%	6.0%	6.4%	6.7%
12:00-13:00	5.5%	5.6%	6.4%	6.7%	7.0%	7.3%
13:00-14:00	5.6%	5.7%	6.3%	7.0%	7.0%	7.5%
14:00-15:00	6.5%	7.8%	6.8%	7.9%	7.0%	7.6%
15:00-16:00	6.8%	9.0%	7.4%	9.0%	7.2%	7.6%
16:00-17:00	6.8%	8.8%	7.0%	9.2%	7.2%	7.4%
17:00-18:00	5.4%	6.4%	6.0%	6.5%	6.8%	6.8%
18:00-19:00	4.0%	4.5%	4.4%	5.2%	6.4%	6.2%
19:00-20:00	4.0%	4.3%	4.6%	5.0%	5.0%	5.0%
20:00-21:00	3.0%	3.5%	4.0%	4.2%	4.3%	4.3%
21:00-22:00	3.0%	3.3%	3.5%	3.7%	3.6%	3.3%
22:00-23:00	2.4%	2.7%	2.5%	2.4%	3.2%	3.0%
23:00-0:00	1.8%	2.2%	1.8%	1.8%	2.6%	2.4%

Table 8: Weekend Average Traffic Volume Distribution Probabilities at Different Time of Day under Different Roadway Types (Schrank et al. 2015).

Day	Road Type	AM Peak	Mid Day	PM Peak	Night
Weekday	Freeway AM Peak	28.4%	34.6%	20.2%	16.8%
	Freeway PM Peak	20.3%	38.1%	24.0%	17.6%
	Arterial	24.5%	37.9%	22.0%	15.6%
	Arterial	17.0%	41.6%	25.9%	15.5%
Weekend	Freeway	13.1%	40.2%	25.4%	21.3%
	Arterial	11.8%	42.8%	25.4%	20.0%

#### 4.3.4 Time-space Network Specification

The optimization model is applied to the study network to seek optimal sensor location-relocation strategies with a one-week rolling time horizon. In addition to non-recurrent traffic incidents, the temporal-spatial traffic state patterns of a specific transportation network are measured determined by the hour-by-hour and day-by-day traffic volume patterns. Specifically, traffic state at a particular location not only exhibits an hourly temporal pattern but also follows a weekly based pattern. For example, for a particular highway corridor, the travel time pattern on Monday is highly similar to that observed from the previous Monday due to the recurrent traffic demand. Similarly, traffic states from spatially correlated locations exhibit such types of recurrent patterns both on an hourly basis and weekly basis as well. Therefore, to comprehensively study the surveillance benefit introduced by deploying sensors, we expand the target network to a one-week time horizon based time-space network. Due to the strong weekly traffic volume and traffic state patterns exhibited within the network, the optimal sensor deployment-relocation strategies can be flexibly and recurrently applied to a much longer time horizon.

The adopted time-space network in this case study is specified and illustrated in Figure 37. The entire horizon length is one week, i.e., seven consecutive days. Each day is discretized into four time periods. They are AM peak period, midday period, PM peak period and night period. Traffic patterns at most locations can be distinguished by different times of a particular day. In the morning peak period, most people are commuting to work, and during the afternoon peak, people are traveling home. During midday and night period, people travel for different types of purposes, but the network demand is lower than those from AM peak and PM peak. Here, we consider AM peak consists of four hours from 6:00 AM to 10:00 AM and PM peak from 4:00 PM to 8:00 PM. Other day times are classified as midday period and night period. Accordingly, traffic volume on each segment at each particular period of the day and particular day of the week is estimated based on the demand data and distribution given in the last section.

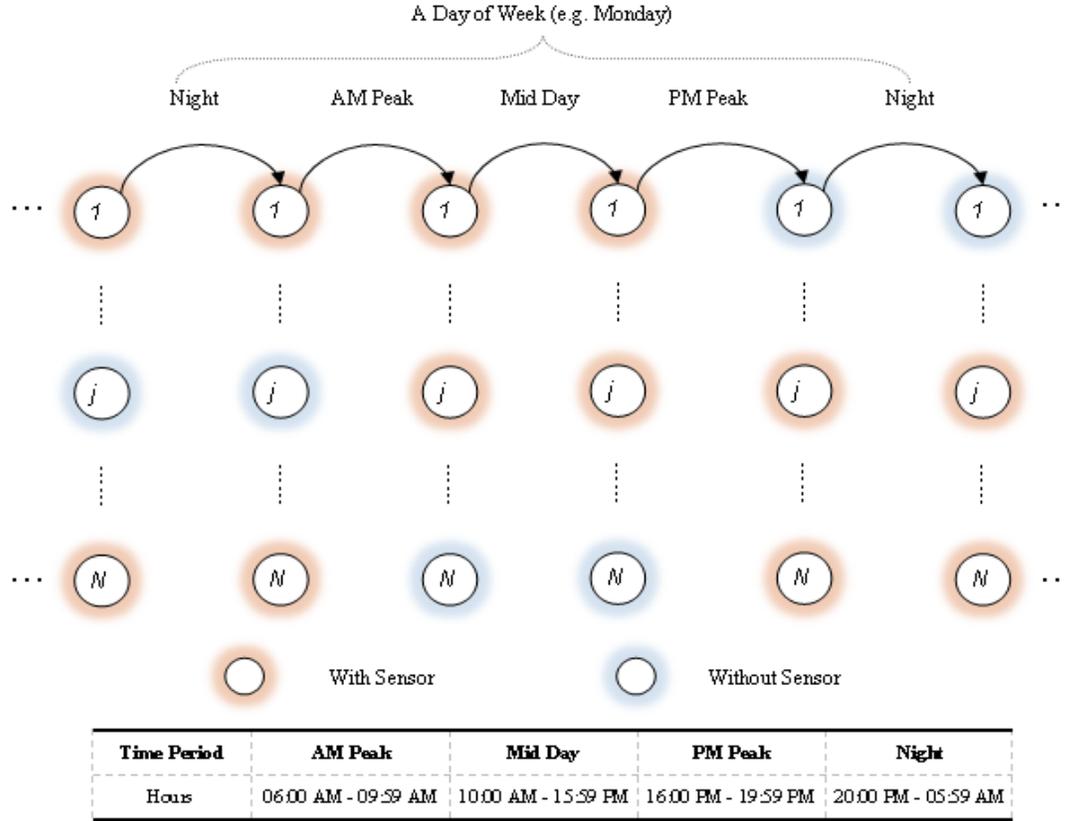


Figure 37: Time-space Network Specification of the Case Study.

#### 4.3.5 Numerical Experiments and Solution Time Investigation

The numerical experiments conducted in this section based on the real-world dataset aim to investigate and answer three major questions:

- (1) The impact of sensor numbers on network level real-time traffic state uncertainty reduction,
- (2) The additional traffic state uncertainty reduction induced by relocation operations, and,
- (3) The solution efficiency of the proposed optimization model.

We take advantage of the measurement-based state prediction uncertainty estimation model proposed in Chapter 3 to study the real-time traffic state uncertainty reduction. More concretely, the time-dependent travel time value is considered to represent the traffic state on a specific segment. As was discussed in Chapter 3, without any real-time data collection scheme, travel time information of a given corridor segment can only be statistically inferred based on the historical dataset, e.g., historical mean value during the same time of the day and day of the week.

Prediction uncertainty based on this type of inference can be quantified based on the historical distribution and denoted by  $H(T^t)$ . The broader the historical distribution is, the larger the estimation uncertainty is.

For such a corridor segment with some kinds of real-time data measurement system (e.g., traffic sensors), the prediction can be made not only based on the historical information but can also rely on the real-time information in past minutes. As is empirically validated in the study given in Chapter 3, the prediction uncertainty for each segment can be reduced to some extent given the real-time information is known. This type of prediction uncertainty can be viewed as the conditional prediction entropy mathematically written by  $H(T^t|\mathbf{X})$ , where  $\mathbf{X}$  represents the real-time data available for the prediction. Specifically, if  $H(T^t|\mathbf{X}) < H(T^t)$  for a particular highway segment, it means real-time measurements can improve the prediction. Otherwise, it indicates the particular real-time measurements  $\mathbf{X}$  does not benefit the prediction.

In this section, we adopted the above concept to investigate and evaluate the benefits introduced by a sensor network from a theoretical perspective. For each period, parameters used to evaluate the objective function are calculated based on prediction uncertainty reduction weighed by the corresponding traffic volume. Further, the predictions for a given location  $l$  are classified into two types. One is prediction based on direct real-time measurements, and the other one is prediction based on indirect real-time measurement. Total prediction uncertainty reduction with respect to direct and indirect measurements are given by the following two formulas.

$$B_l = Flow_l \cdot \max\{H_l(T^t) - H_l(T^t|\mathbf{X}_l), 0\}$$

$$B_{k,l} = Flow_l \cdot \max\{H_l(T^t) - H_l(T^t|\mathbf{X}_k), 0\}$$

where,  $\mathbf{X}_l$  means the real-time data is measured from location  $l$ , and  $\mathbf{X}_k$  means the real-time data is measured from another location  $k$ .  $Flow_l$  denotes the traffic volume at location  $l$ . For the study network with the one-week rolling time horizon, each of the above parameters are estimated for each discretized periods (i.e., AM Peak, mid-day Period, PM Peak, and night period).

Nearly every existing traffic sensor network optimization model only considers the surveillance benefit based on direct measurements. In other words, by deploying sensors to a particular location, they only consider the benefit brought to the location itself and ignore the additional benefits brought to its adjacent segments (e.g., downstream and upstream segments). In this study, we empirically calculate and consider both direct-surveillance benefit and indirect-surveillance benefit. For indirect surveillance, we consider the real-time data collected at each segment can also be used to make real-time predictions for its two-nearest upstream segments and two-nearest downstream segments. For example in the map given by Figure 31, the two-nearest downstream segments for '1516' are '1617' and '1718', and the two-nearest upstream segments are '1314' and '1415'.

The mathematical formulation was implemented with Python scripts and solved by MIP solver Gurobi 7.5.1. Platform for running the optimization is 64-bit Windows 10 operating system with Intel(R) Core(TM) i7 CPU @3.6 GHz and 16 GB RAM. Computational results are given and discussed in the next subsection. The computational results are summarized and given in Table 9.

As are shown in the summary table, seven sensor fleet sizes are considered, i.e., sensor number equal to 4, 8, 12, 16, 20, 24, 28, and five different relocation limits are investigated with respect to each particular size of sensor fleet. The solution time for cases in which the sensor fleet size is large (i.e., 24 and 28) or small (i.e., 4 and 8) are commonly lower than those of cases in which sensor fleet size is moderate (i.e., 12, 16, and 20). The basic reason is that the solution process of the location-relocation optimization model is affected by the symmetry structure of the location and relocation decision variables. When the sensor number is relatively smaller or larger compared to the network node set, it tends to be easier to find the optimal locations for deploying and relocating the sensors. When the sensor number is relatively moderate compared to the network size, size of the multiple closed-to-optimal solutions are large, which results in larger solution time.

More specifically, Figure 38 plots the solution time for each optimization case, and Figure 39 shows the median solution times with respect to the original formulation and the original formulation with integer cuts added. The worst solution time for solving the original formulation with the study network and time horizon is around 4600 seconds and the least time is around 800 seconds. Adding integer cuts can significantly reduce the solution time for each case. As is shown in Figure 39, the average solution time by adding the proposed cuts is no more than 500 seconds, whereas the average solution time without adding the cuts can be much higher reaching as 2400 seconds. The various numerical experiments indicate the high efficiency of the formulated optimization model. Although the optimization model is for planning purpose, the optimal solution can be found in a satisfactory time. This lays a solid foundation for running the location-relocation optimization in a real-time manner with even smaller rolling time horizon.

Table 9: Summary of Computational Results under Different Model Configurations.

Sensor Number	Relocation Limit	Obj-Part1 (Direct)	Obj-Part2 (Indirect)	Objective Value	Optimal Relocation Amount	Solution Time (seconds)	
						Original Formulation	Original Formulation (with Cuts)
4	2	1364138	3393119	4757256	2	1070.59	175.79
	4	1401573	3410535	4812109	3	2226.95	745.29
	6	1423581	3412329	4835909	6	1279.86	223.33
	8	1423581	3412329	4835909	6	1091.89	167.44
	10	1423581	3412329	4835909	6	2330.58	176.60
8	2	2730689	5474419	8205108	1	2358.06	734.19
	4	2730689	5474419	8205108	1	2595.04	336.50
	6	2730689	5474419	8205108	1	2408.73	285.00
	8	2730689	5474419	8205108	1	2562.51	277.60
	10	2730689	5474419	8205108	1	2330.58	298.14
12	2	3994893	6336183	10331076	2	2539.43	351.59
	4	4000879	6341525	10342404	4	3707.34	586.16
	6	3983809	6364421	10348230	6	4628.98	626.01
	8	4014868	6338189	10353057	8	4079.12	983.46
	10	3997799	6361085	10358884	10	4264.26	965.91
16	2	5367260	6463538	11830798	2	2549.90	352.76
	4	5397631	6459140	11856771	4	3360.02	430.97
	6	5307092	6557920	11865012	6	3831.03	437.58
	8	5390726	6486747	11877473	8	3261.83	481.91
	10	5394617	6488978	11883595	10	4476.59	492.18
20	2	6893881	5796740	12690621	2	1845.51	259.30
	4	7128181	5605618	12733800	4	2273.24	301.79
	6	7119371	5642378	12761749	6	2775.18	365.00
	8	7269770	5502254	12772024	8	3067.37	495.31
	10	7254472	5524945	12779417	10	3877.14	666.76
24	2	8751295	4619565	13370859	2	1347.97	201.30
	4	8902314	4479831	13382145	4	1275.18	185.48
	6	8905514	4480749	13386263	6	1581.39	193.71
	8	8992619	4397280	13389900	8	1994.18	311.94
	10	8965441	4429189	13394629	10	2007.75	233.37
28	2	10649130	2987019	13636149	2	517.17	67.10
	4	10839243	2808731	13647974	4	858.13	265.15
	6	10887288	2771915	13659203	6	924.30	154.32
	8	10518943	3147815	13666758	8	1035.06	110.28
	10	10700294	2972783	13673077	10	1142.90	159.39

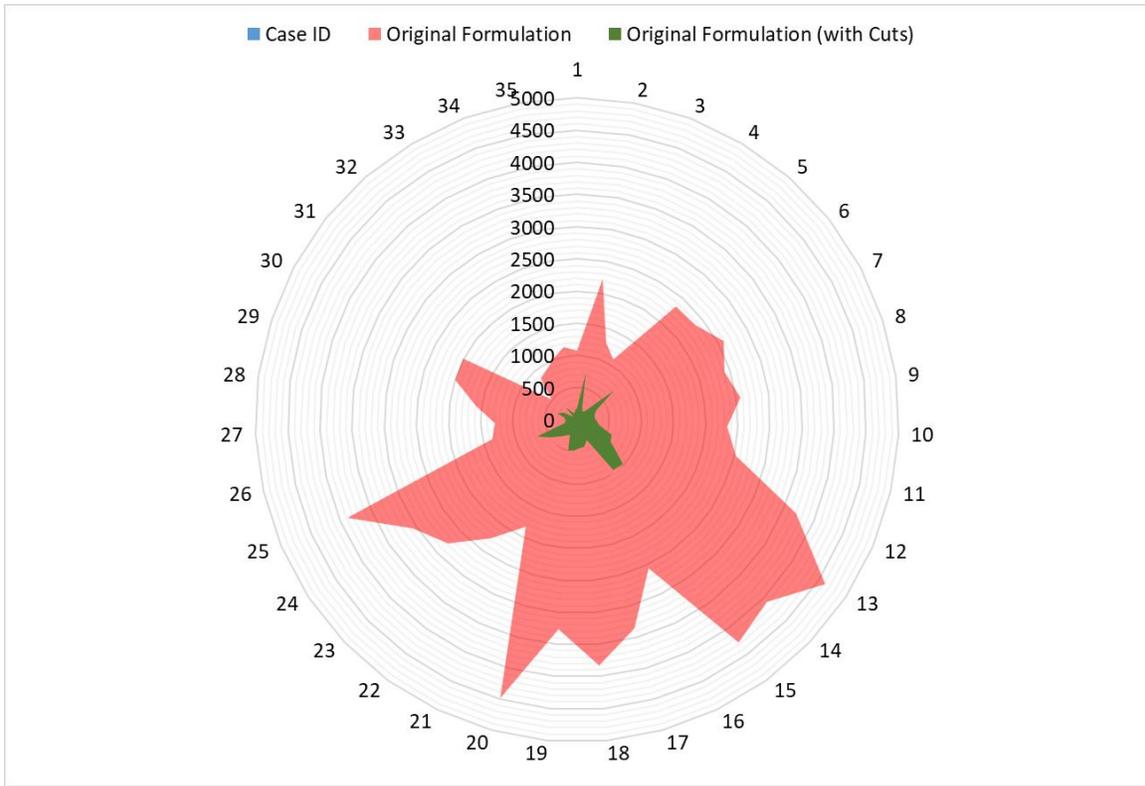


Figure 38: Solution Time Comparison of Each Optimization Case.

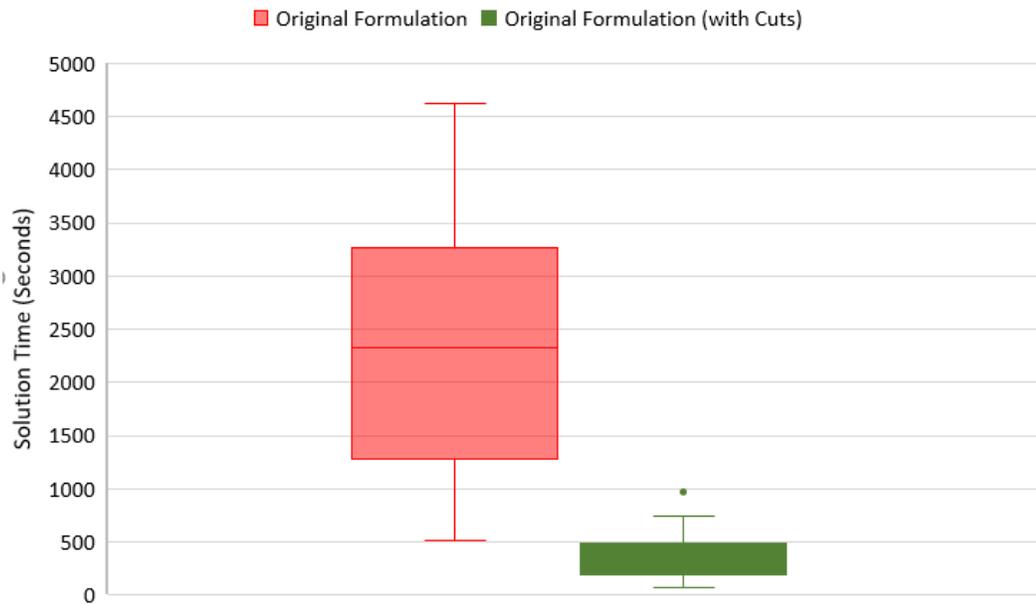


Figure 39: Overall Solution Time Comparison between the Original Formulation and the Cut based Formulation.

The fifth column in the summary table gives the optimal objective value for each case, i.e., the total network-level travel time prediction uncertainty reduction for the entire time horizon (i.e., one week). The objective value consists of two parts, the direct-measurement based surveillance benefit (i.e., Obj-Part1 in 3<sup>rd</sup> column) and the indirect-measurement based surveillance benefit (i.e., Obj-Part2 in 4<sup>th</sup> column). Figure 40 plots the total objective value, Obj-Part1 value, and Obj-Part2 value for each optimization case. There are three major findings by comparing the objective value of each case.

First, increasing sensor fleet size can significantly improve the total surveillance benefit, specifically when the fleet size is relatively small. However, when the fleet size is relatively large, although adding more sensors can improve the total surveillance benefit, the improvement is not high. For example, when the sensor fleet size is 4, increasing the fleet size to 8 can nearly double the surveillance benefit (i.e., prediction uncertainty reduction). Given the sensor number is 20, increasing it to 24 can only make a slight improvement.

Second, with the increase of the sensor fleet size, surveillance benefit improvement based on direct measurement (i.e., green dots in the graph) does increase as expected, but the increase rate becomes smaller. In other words, the marginal benefit with respect to the fleet size turns to be small. On the other hand, the indirect-measurement based surveillance benefit reaches a peak when the fleet size is moderate (i.e., 16), then decreases as the fleet size increases. The definition of the optimization model determines this. More specifically, given more sensors are available, more segments in the network can be directly monitored. Consequently, the prediction system for those segments will rely on the real-time data collected by itself. When sensors are fewer, predictions for more segments have to rely on the data collected from their upstream or downstream segments.

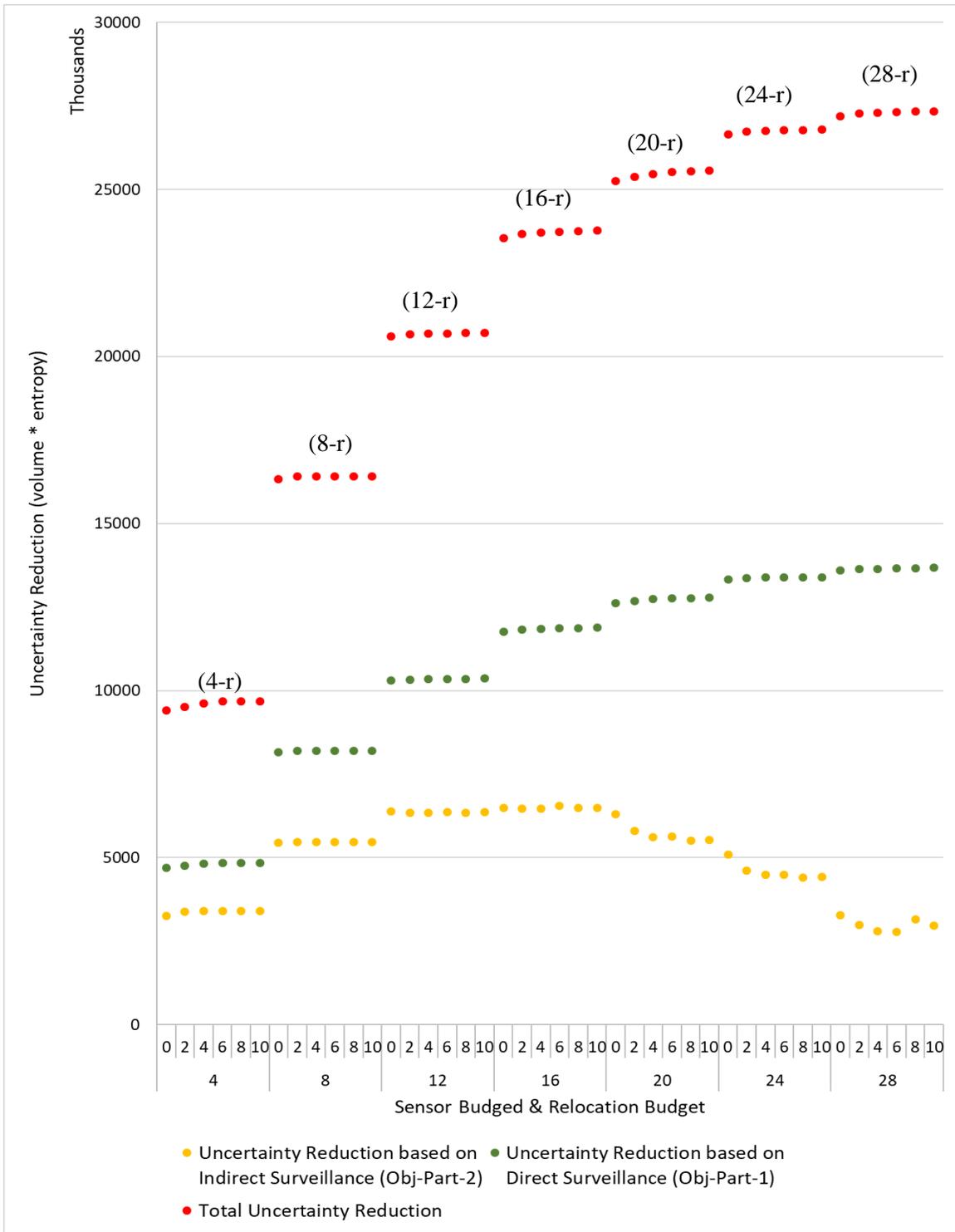
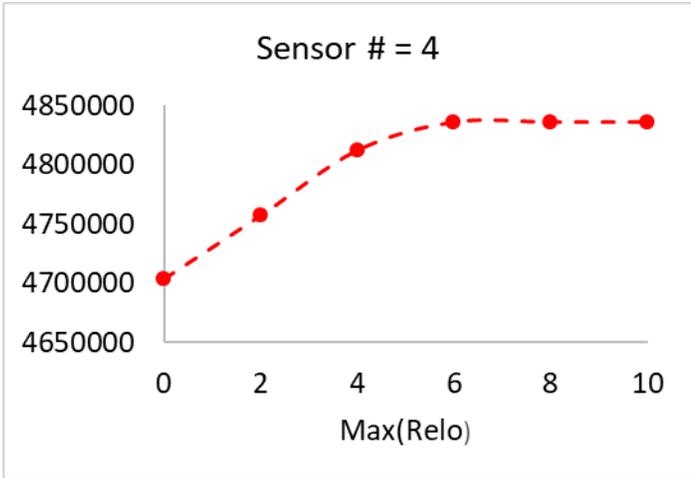


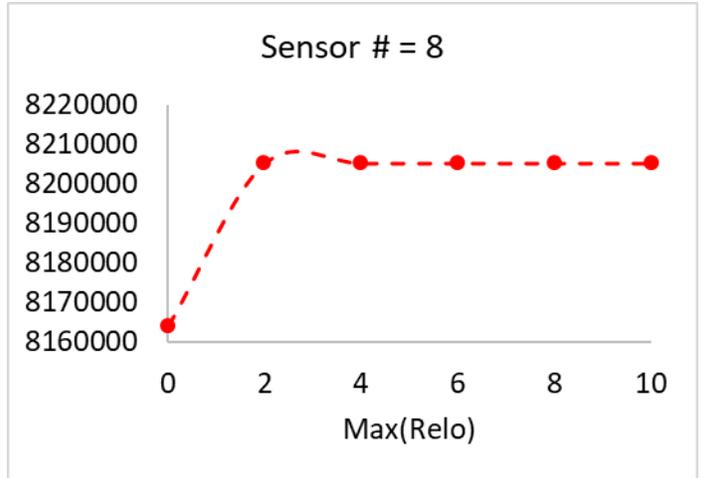
Figure 40: Objective Value (Total Prediction Uncertainty Reduction) Comparison for Different Cases.

A third finding from the above optimal solutions is that allowing flexibilities to relocate sensors in the time horizon can bring additional surveillance benefits. As is shown by the 6<sup>th</sup> column in the table, the constraint of the relocation operation limit is active for every case except a few in which sensor fleet size is 8. A set of more clear plots to demonstrate the objective trend with respect to relocation constraint under each fleet size is provided in Figure 41. As are shown in the subplots, relaxing the relocation constraint does reduce the total prediction uncertainty due to the heterogeneous temporal-spatial distribution of both traffic volumes and prediction performance in the time-space network. But as the relocation budget increases, the marginal benefit decreases. This is mainly because conducting more relocation operations results in more real-time surveillance loss.

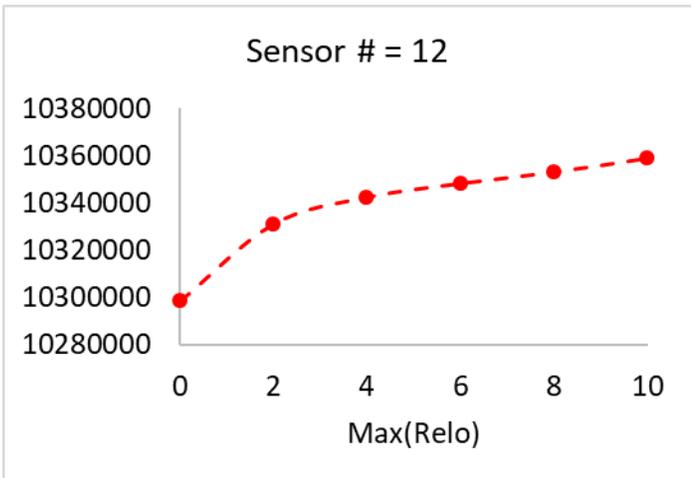
In this section, an overview of the results based on the application of the proposed dynamic sensor location-relocation optimization model is provided. The objective function is formulated based on the prediction uncertainty. Possibility and benefit of considering sensor relocation within a given time horizon are explored and validated based on the real-world data. At the current stage, one can only conclude that inducing relocation operations might improve the network-level surveillance benefit from a theoretical perspective since no monetary estimations are evaluated with respect to the additional uncertainty reduction benefits introduced by relocating sensors. In the next section, a more practical analysis will be given to answer the value of sensor relocation operations.



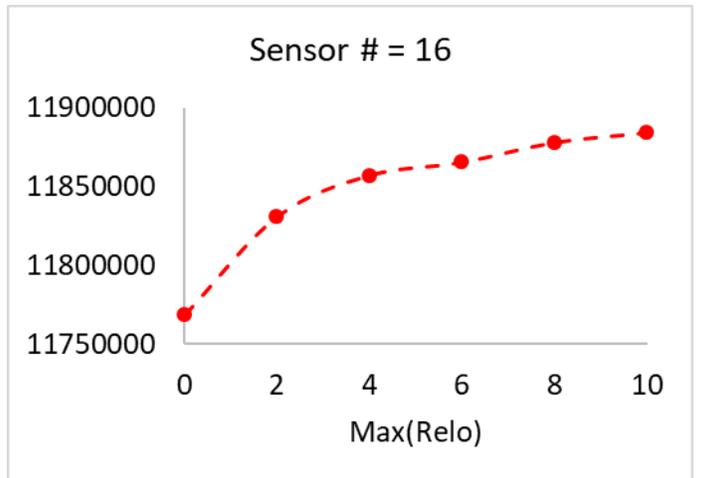
(a) 4-r



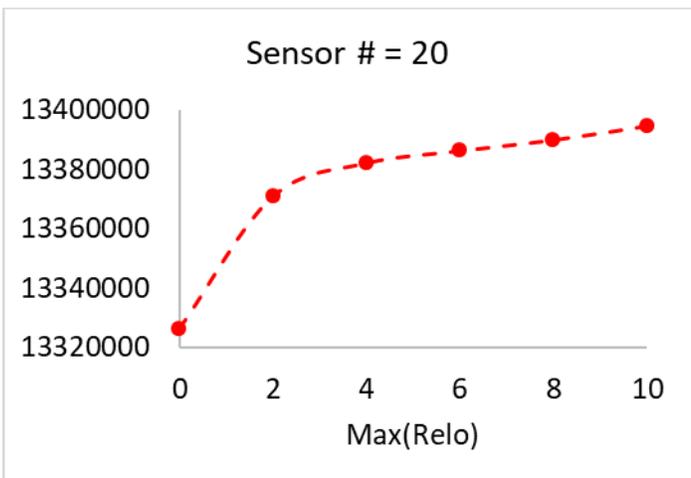
(b) 8-r



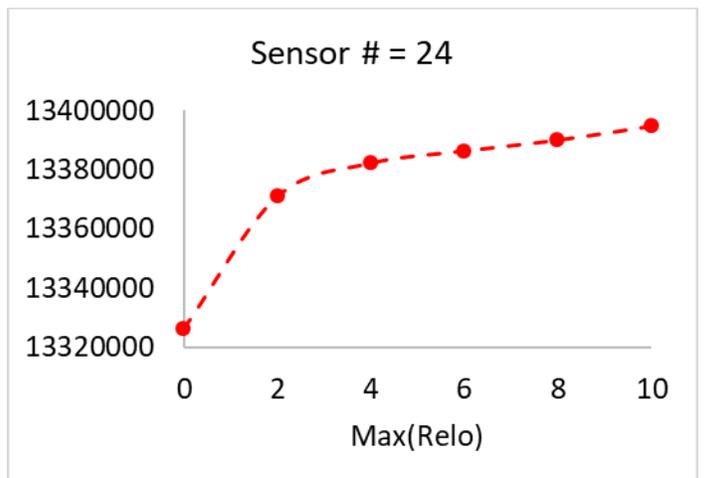
(c) 12-r



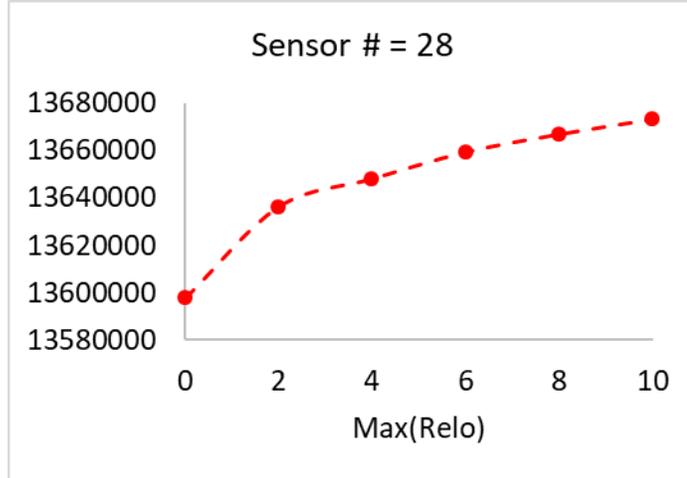
(d) 16-r



(e) 20-r



(f) 24-r



(g) 28-r

Figure 41: Effects of Relocation Operations on Surveillance Benefit Improvement: (a) 4-sensor Case, (b) 8-sensor Case, (c) 12-sensor Case, (d) 16-sensor Case, (e) 20-sensor Case, (f) 24-sensor Case, and (g) 28-sensor Case.

#### 4.3.6 Monetary Savings Estimation and Practical Policy Implications

Improving real-time traffic surveillance efficiency has a positive economic impact for travelers commuting in the highway network. In this case study, travel time is considered as the target variable for real-time prediction, thus improving real-time surveillance benefits indicates enhancing travelers' travel time reliability by reducing the real-time predicted travel time errors. In other words, improving the travel time information accuracy is of key importance in any advanced traveler information system (ATIS). Enhancing travelers' perception of the travel time of their upcoming trips can directly benefit their time savings since each traveler has a value of time associated with his/her daily commutes.

Suppose the average true travel time for a traveler traversing a roadway  $l$  is  $\bar{T}$ , and the expected relative prediction error on his/her travel time is  $e_l = \frac{1}{N} \sum_i \frac{|\hat{T}_i - T_i|}{T_i}$ , where  $\hat{T}_i$  is the predicted travel time given by a specific ATIS, and  $T_i$  is the ground truth value. The expected time wasted by the inaccurate prediction for the traveler on this segment can be estimated as  $e_l \cdot \bar{T}$ .

For underestimated prediction, it means the traveler should spend more unexpected time in his/her travel (i.e., late arrival case). This can be directly understood as the time waste resulting from the inaccurate prediction. For overestimation prediction, it indicates the traveler is faced with more unexpected time at his arrival (i.e., early arrival case). From the travel planning perspectives, either case is considered as an economic loss due to the unexpected travel time.

To study the economic effects induced by a particular sensor network, we update the objective function by explicitly considering total wasted travel time savings as the surveillance benefit. Specifically, the benefit parameters  $B_l^\tau$  and  $B_{k,l}^\tau$  in the objective function is calculated as below,

$$B_l^\tau = flow_l^\tau \cdot e_l^\tau \cdot \bar{T}_l^\tau$$

$$B_{k,l}^\tau = flow_l^\tau \cdot e_{k,l}^\tau \cdot \bar{T}_l^\tau$$

where,  $flow_l^\tau$  is the expected traffic volume through link  $l$ .  $e_l^\tau$  denotes the expected relative prediction error (i.e. MAPE) based on real-time data collected from  $l$ , and  $e_{k,l}^\tau$  denotes the expected relative error based on real-time data collected from link  $k$ .  $\bar{T}_l^\tau$  represents the average travel time value through  $l$  when the entrance time is within time window  $\tau$ . Therefore, the new objective function is interpreted as the total wasted time savings of the entire network in a given time horizon upon the surveillance of a particular sensor network. This value can also be used to estimate the monetary savings if the average value of time  $\theta$  for the travelers is assumed and considered.

Error terms  $e_l^\tau$  and  $e_{k,l}^\tau$  used to evaluate the above benefit parameters can either be directly evaluated by using specific prediction models with training data collected from each roadway location at each time period, or can be indirectly inferred based on the uncertainty estimation model proposed in Chapter 3. As is empirically validated in Chapter 3, data-driven based travel time prediction error on a specific highway segment has a monotonous increasing relationship with the entropy-based uncertainty index. Further, as it is empirically illustrated, the

uncertainty model can more robustly infer the prediction error reduction based on the stochastic state transition patterns with and without real-time traffic surveillance. Based on the empirical results obtained in Chapter 3, we consider the prediction error reduction in the studied highway system has the following relationship with the uncertainty reduction measurement (Table 10).

Table 10: Empirical Relationship between Relative Travel Time Prediction Error Reduction and Stochastic System Uncertainty by Introducing Real-time Measurements.

Prediction Relative Error Reduction (MAPE)	Entropy-based Prediction Uncertainty Reduction
1.0%	(0.0, 0.2]
2.0%	(0.2, 0.4]
4.0%	(0.4, 0.6]
7.0%	(0.6, 0.8]
12.0%	(0.8, 1.0]
18.0%	(1.0, 1.2]
30.0%	> 1.2

Optimization results with the total travel time prediction error savings as objective function are summarized and given in Table 11. To study the beneficial effect introduced by relocating sensors at specific timestamps, we calculated the static optimal sensor locations for benchmark comparisons. In the static deployment case, none of the deployed sensors can be relocated in subsequent periods. This indicates that once a set of sensors are deployed, they will function at the initial locations for the entire week.

The objective value is in the unit of hours, which represents the total travel time errors reduced for all travelers during the entire week. Also, the total error savings attributed to the direct surveillance and indirect surveillance are given in the 4<sup>th</sup> column and the 5<sup>th</sup> column of the table, respectively. As is shown in the table, indirect surveillance based error savings are higher when sensor budget is low. As the sensor budget increases, the network-level travel time error savings significantly increases. Indirect surveillance contributes a lot to the total error savings, but gradually direct surveillance overtakes the contribution.

Table 11: Monetary Saving based Optimization Results Summary.

<i>Case Configuration</i>		<b>Optimal Solution</b>			
<i>Sensor Number</i>	Relocation Limit	Relocation #	Direct-monitoring based Error Savings (Hours)	Indirect-monitoring based Error Savings (Hours)	Total Error Savings (Hours)
4	0	0	3,545	13,626	17,170
4	2	0	3,545	13,626	17,170
4	4	0	3,545	13,626	17,170
4	6	0	3,545	13,626	17,170
4	8	0	3,545	13,626	17,170
4	10	0	3,545	13,626	17,170
8	0	0	7,160	23,257	30,417
8	2	2	6,665	23,882	30,547
8	4	4	6,938	23,676	30,614
8	6	4	6,938	23,676	30,614
8	8	4	6,938	23,676	30,614
8	10	4	6,938	23,676	30,614
12	0	0	8,884	30,746	39,630
12	2	0	8,884	30,746	39,630
12	4	0	8,884	30,746	39,630
12	6	0	8,884	30,746	39,630
12	8	0	8,884	30,746	39,630
12	10	0	8,884	30,746	39,630
16	0	0	11,966	31,623	43,589
16	2	2	13,809	30,087	43,896
16	4	4	14,501	29,619	44,120
16	6	6	14,446	29,791	44,237
16	8	8	14,881	29,463	44,344
16	10	10	14,833	29,554	44,387
20	0	0	17,302	27,888	45,191
20	2	2	17,916	27,582	45,498
20	4	4	17,681	28,015	45,696
20	6	6	17,677	28,182	45,860
20	8	8	17,879	28,066	45,945
20	10	10	18,935	27,118	46,053
24	0	0	20,651	25,281	45,932
24	2	2	22,536	23,720	46,257
24	4	4	23,025	23,509	46,535
24	6	6	22,320	24,378	46,698
24	8	8	22,364	24,454	46,818

24	10	10	22,319	24,596	46,915
28	0	0	23,318	22,774	46,092
28	2	2	23,668	22,803	46,471
28	4	4	24,383	22,373	46,756
28	6	6	24,239	22,705	46,943
28	8	8	24,028	23,054	47,083
28	10	10	25,236	21,994	47,230

With surveillance benefits by static sensor networks as benchmarks, the trade-off and marginal analysis between sensor budget and relocation limit are plotted in Figure 42. As is displayed in the figure, the dashed lines with red markers, blue markers, purple markers and green markers provide the optimal objectives for optimization cases with sensor budget equal to 16, 20, 24, and 28, respectively. The x-axis denotes maximum relocation limit. When maximum relocation limit is equal to zero, the optimal solution corresponds to the static sensor network. Total travel time prediction error reduction is 4,3589 hours for the case in which 16 sensors are deployed, and relocation is prohibited.

As is indicated in the plot, both increasing sensor budget and allowing sensor relocations can increase the objective function value. For example, when the sensor budget is increased from 16 to 20, the total error savings can be increased by 3.67%. By maintaining the sensor budget as 16, allowing ten relocation operations can increase the objective by 1.84%. An interesting fact can be seen in the comparison plot. The marginal surveillance benefit becomes smaller when the number of sensors is relatively large compared to the network size. For example, by increasing the number of sensors from 20 to 24 and 24 to 28, only 1.64% and 0.35% additional surveillance benefits can be brought, while this marginal benefit is 3.67% when increasing sensor budget from 16 to 20.

But in those cases where the number of sensors is relatively large, allowing sensor relocations can introduce notable objective enhancement. In other words, the marginal benefit of relocation budget turns to be large when the number of sensors is already high. For instance in the

case study, for the optimization case with 24 sensors, allowing at most 10 relocation operations can bring an additional 2.14% increase in the surveillance benefit while only 0.35% increase is realized by adding four more sensors.

The interesting phenomenon revealed above is mainly because relocation can result in surveillance loss to some extent. The real-time surveillance loss is more significant when the number of sensors is relatively small compared to the network size. More specifically, when the fleet size of sensors used to monitor the network is small, it is better to maintain those sensors at some fixed locations where the total travel time error is periodically high considering the similar day-to-day patterns of the traffic volume. Although relocating some of the sensors at such locations to some other locations for some particular time periods can bring additional surveillance benefits, the surveillance benefit at such locations will be lost to some large extent as well. As a consequence, an increase of the total surveillance benefit is not significant. However, when the fleet size of sensors in the network is relatively large, one can guarantee the major locations at which the total travel time errors are large can be consistently covered. Further, relocating parts of the sensors from locations with relatively small surveillance benefits to locations with non-recurrently higher travel time error reductions can bring significant additional benefits.

Based on the optimization solutions obtained in this case study, we can conclude that the trade-off between surveillance benefit loss and relocation-induced benefit increase is high when sensor fleet size is small (i.e., below 20 sensors in this commuting network). But when the sensor fleet size is already large (i.e., over 20 sensors in this commuting network), conducting relocation operations can bring more notable surveillance benefits than buying more sensors. The 2-D comparison plot based on the optimal results obtained through the proposed dynamic sensor planning model provides a useful tool to evaluate the tradeoff between cost and benefits. Detailed impacts of relocation limits on total travel time error savings under sensor budgets are given in Figure 43.

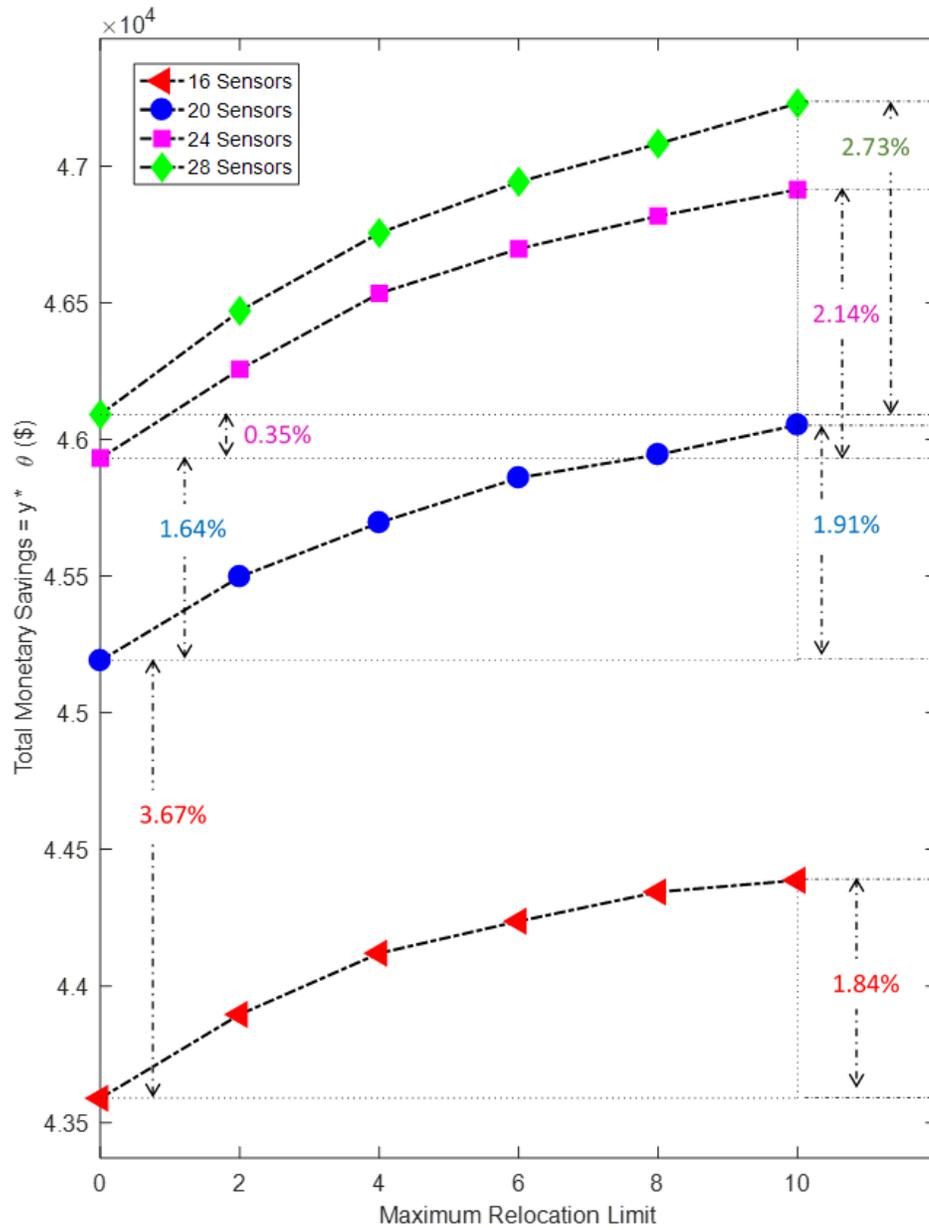
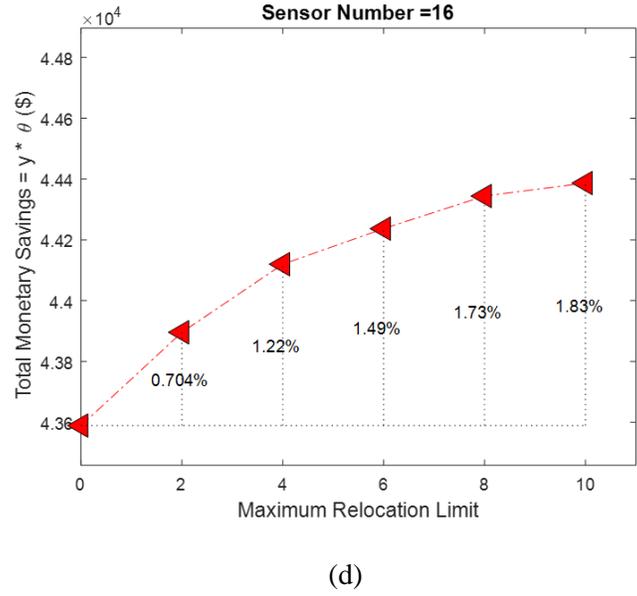
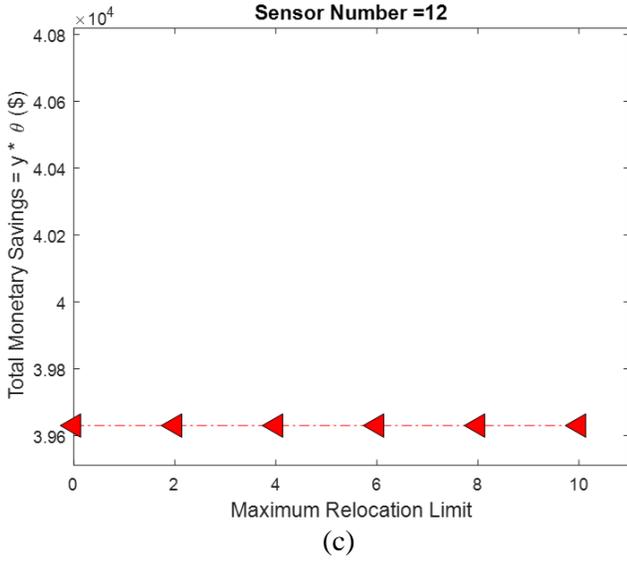
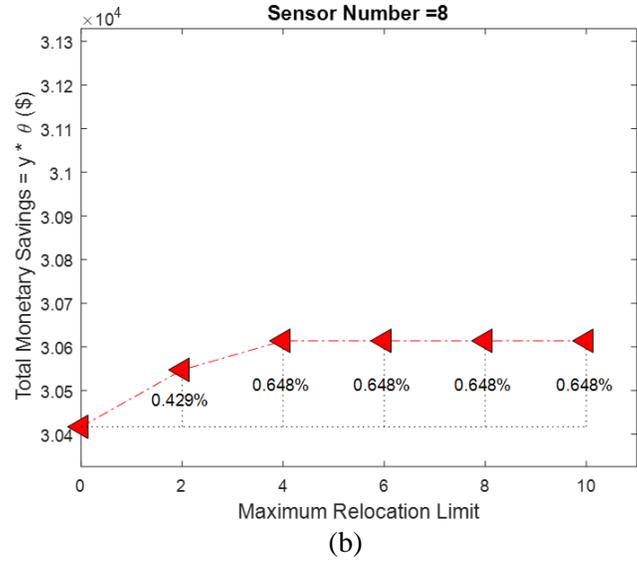
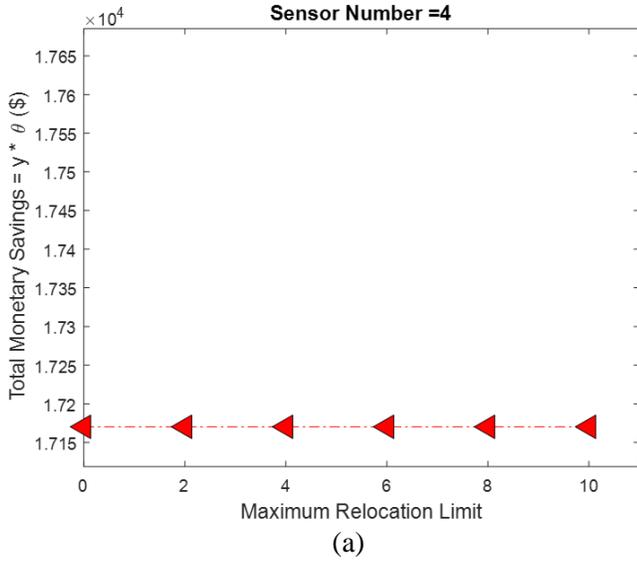


Figure 42: Combinatory Impacts of Sensor Budget and Relocation Limits on Total Network Travel Time Error Savings.



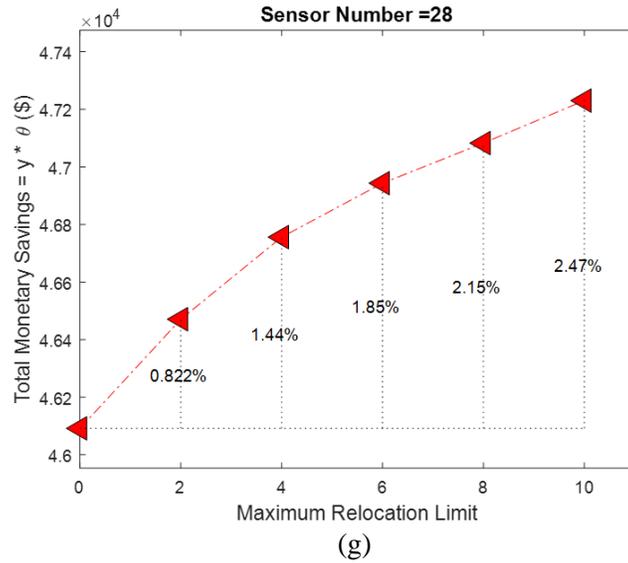
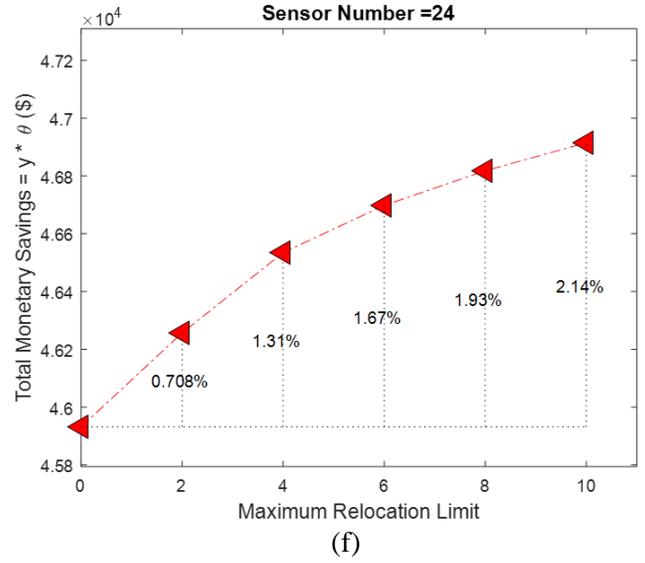
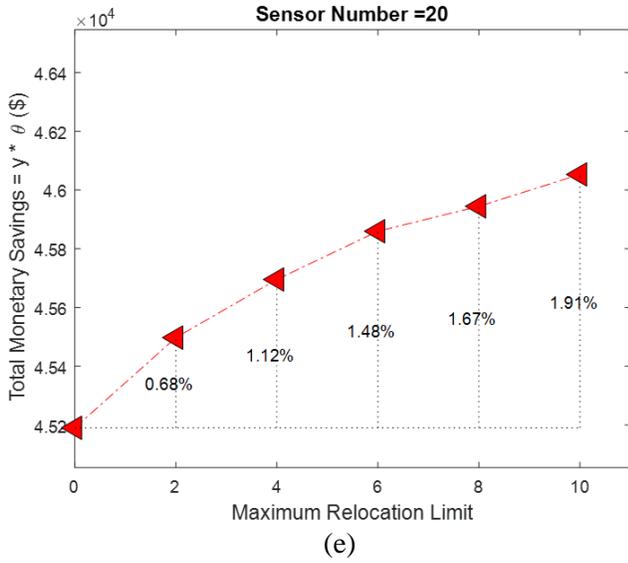


Figure 43: Impacts of Relocation Limits on Total Travel Time Error Savings for Different Fleet-Size Sensor Network: (a) 4-Sensor Network; (b) 8-Sensor Network; (c) 12-Sensor Network; (d) 16-Sensor Network; (e) 20-Sensor Network; (f) 24-Sensor Network; (g) 28-Sensor Network.

#### 4.3.7 Applications with the Existence of a Second Data Source

Numerical experiments conducted in previous sections investigated surveillance benefit improvement over the transportation network with a particular set of traffic monitoring sensors. Real-time travel time information collected by those sensors serves as an independent data source to estimate the surveillance benefit. As was pointed out early in this Chapter, there might exist another independent data source providing related real-time traffic information. Once the ATIS

has access to both the data collected from the deployed sensors and the access to the data provided by a second real-time data provider (e.g., GPS probe data), it can further evaluate the overall surveillance benefits by jointly considering the two data sources. The proposed optimization model provides flexibility to conduct the joint surveillance benefit evaluation. In such cases, the value of the binary parameter  $a_l$  indicating whether a segment has reliable traffic information provided by a second data source should be explicitly considered.

In this section, numerical experiments are conducted to investigate two major issues. Firstly, the impact of the existence of a second data source on the surveillance benefit improvement is quantitatively studied for the DC-Baltimore commuting network. We consider the existence of second reliable data source at different levels. In particular, the coverage level is expressed as the percentage of segments having reliable real-time data feeds from another data source. For examples displayed in the following results, by referring coverage probability equal to 0.5, it indicates half of the network segments have reliable data feeds from another data source in addition to the newly deployed sensor system.

Another interesting issue quantitatively investigated here is whether the total travel time prediction error reduction based on direct measurements is always better than that based on indirect measurements. In nearly every existing study about traffic sensor placement optimization, people mainly focused on the surveillance benefits at the locations where sensors are placed and considered the surveillance benefits at locations without sensors as a trivial improvement (based on other measurements). The optimization model developed in this research explicitly takes the indirect measurement based surveillance benefit into account (i.e., possibility of predicting a segment's travel time based on measurements from other segments). As is already revealed by the results from previous two sections, the contribution from the indirect measurement based predictions on the total network travel time error reduction is not trivial. Especially when the number of sensors is small, a large proportion of the total travel time error reduction is attributed to the indirect measurement based predictions.

Even though we consider the possibility of using measurements from one location to predict traffic state at another location without measurement, there exists one assumption for the optimization model. As one may notice, constraint (4-2-5) of the proposed sensor network optimization model indicates that the indirect measurement based prediction can be applied to a specific location only if this location is out of surveillance. Otherwise, traffic state prediction on this segment will rely on real-time measurements directly collected from it. This constraint empirically originates from the field of traffic prediction studies. In most of the past empirical traffic related prediction studies, it is more reliable and accurate to predict the state on a highway corridor based on the data collected from the same segment since traffic state is directly determined by the traffic volume through it. This might be valid for traffic volume detection. However, for real-time travel time prediction on a segment, the prediction based on related information measured from either downstream or upstream might outperforms the prediction based on travel time measured from the same segment in the past minutes.

For the above reason, we investigate the total network travel time prediction error reduction with respect to two different prediction policies in this section as well. The two real-time measurement-based prediction policies are defined as below:

- Policy 1: For real-time travel time prediction on a specific segment  $l$ , the priority of using data feeds measured from segment  $l$  is higher than that measured from any of its adjacent segment  $k \in I(l)$ .
- Policy 2: For real-time travel time prediction on a specific segment  $l$ , the priority of using data feeds measured from segment  $l$  is equal to that measured from any of its adjacent segment  $k \in I(l)$ .

Policy 1 is already enforced by the formulation given in section 4.2 (i.e., the original formulation). It is easier to reformulate the optimization model to enforce Policy 2. Here we

introduce a new auxiliary decision variable  $d_l$  indicating whether segment  $l$  is monitored by a pair of travel time detection sensors. Accordingly, the objective function can be updated as,

$$\text{Maximize: } \sum_{\tau} \sum_l f_l^{\tau} \cdot B_l^{\tau} \cdot d_l^{\tau} + \sum_{\tau} \sum_l \sum_{k \in I(l)} f_l^{\tau} \cdot B_{k,l}^{\tau} \cdot u_{k,l}^{\tau}$$

Also, two new constraints are incorporated, and constraint (4-2-5) in the original formulation is removed.

$$d_l^{\tau} \leq a_l^{\tau} + y_l^{\tau}, \forall l, \forall \tau$$

$$d_l^{\tau} + \sum_{\forall k} u_{k,l}^{\tau} \leq 1, \forall l, \forall \tau$$

The two newly added constraints and the maximization-based objective function together ensure that there is no strict priority on the use of direct measurements and indirect measurements for the traffic state prediction for any segment  $l$ . More specifically, the one that can bring the largest prediction error reduction is automatically chosen by the objective function.

The optimization results considering both prediction policies are given through Figure 44 to Figure 50. For each particular level (i.e.,  $p\%$ ) of the 2<sup>nd</sup> reliable data source coverage, we randomly generated 20 different scenarios, in which  $p$  percent of the segments have reliable travel time data from a 2<sup>nd</sup> data source. The following plots provide an intuitive view on the maximal total travel time error savings under different sensor budgets and the level of 2<sup>nd</sup> data source coverage.

In addition to the overall expectation that total error savings increases as the sensor budget increases and the 2<sup>nd</sup> data source coverage level increases, one counter-intuitive finding is obtained by comparing the optimal solutions under Policy 1. As is displayed by part (a) in each figure, when the coverage level of reliable real-time data feeds is over 60%, the increases in the coverage level result in the decrease of total travel time error savings to some minor extent. This is due to the constraint of Policy 1 on the prediction manner. In fact, using data feeds collected from upstream or downstream segments for prediction may be better than that directly collected

from the segment. For example, temporal data delay may result in inaccurate travel time prediction on an overly long segment. Given the requirement posted by Policy 1, when the coverage level of reliable data feeds is high, travel times on most segments are predicted based on the direct measurements. As a consequence, advantages of using upstream or downstream data feeds for prediction are completely ignored. Even though this loss is not large, it is notable from the numerical results given by Figure 47-(a), Figure 48-(a), and Figure 49-(a).

The optimization model under Policy 2 fairly evaluates the traffic stage prediction performance introduced by direct measurements and indirect measurements and aims to choose the best one for network-level surveillance benefit evaluation. As is shown by part (b) in each figure, the objective value (i.e., total travel time error savings) monotonically increases as the coverage level of the 2<sup>nd</sup> data source increases. Even though some segments are deployed with sensors, their traffic state predictions rely on the data feeds collected from other spatially correlated segments. Therefore, as the sensor budget and real-time data feed coverage increase, one can more intelligently evaluate and improve the network-level travel time error reduction.

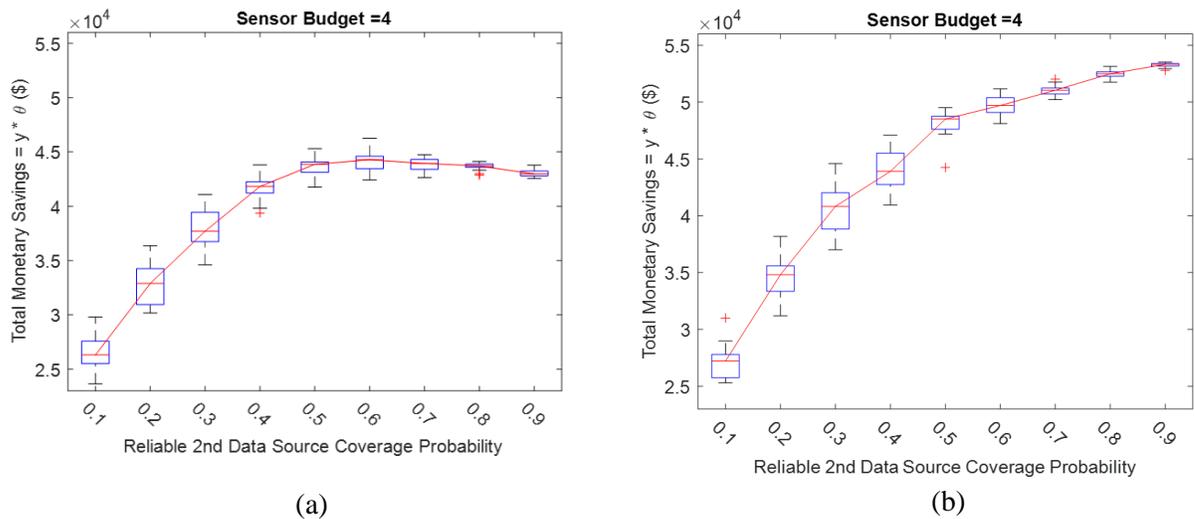
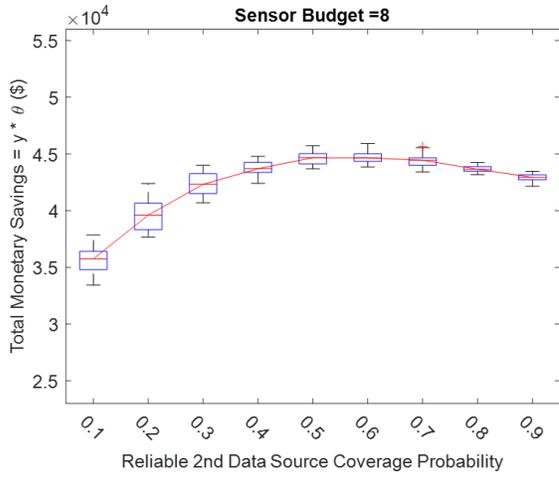
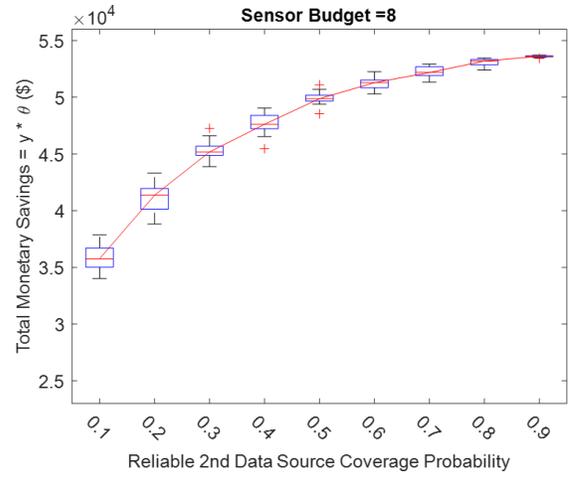


Figure 44: Total Travel Time Error Savings with 4 Sensors under Different Levels of 2<sup>nd</sup> Data Source Coverage: (a) Policy 1; (b) Policy 2.

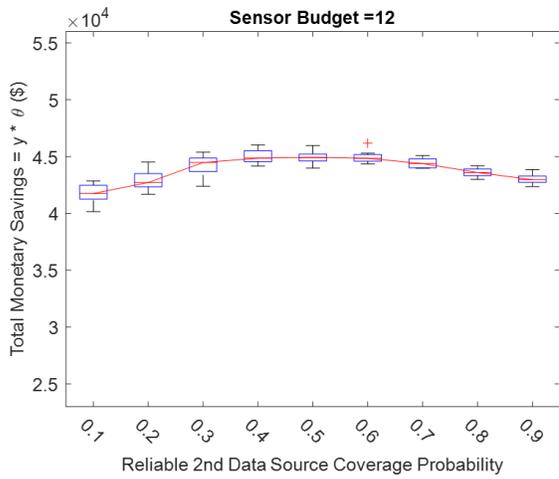


(a)

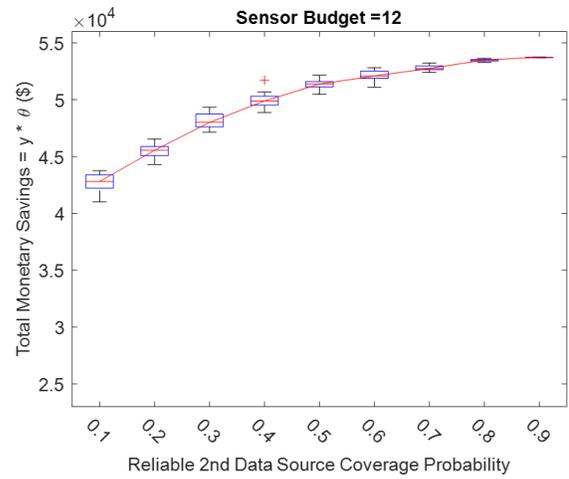


(b)

Figure 45: Total Travel Time Error Savings with 8 Sensors under Different Levels of 2<sup>nd</sup> Data Source Coverage: (a) Policy 1; (b) Policy 2.

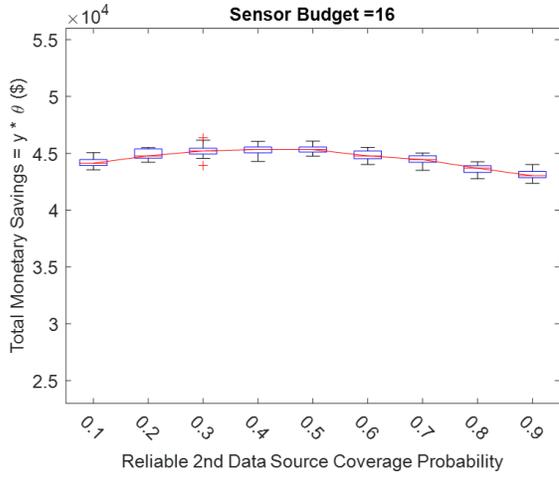


(a)

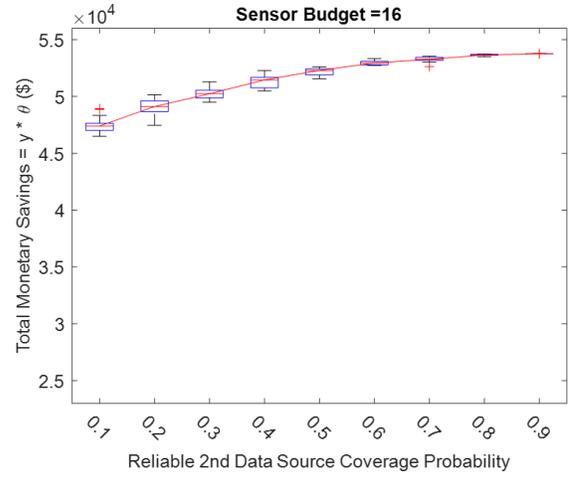


(b)

Figure 46: Total Travel Time Error Savings with 12 Sensors under Different Levels of 2<sup>nd</sup> Data Source Coverage: (a) Policy 1; (b) Policy 2.

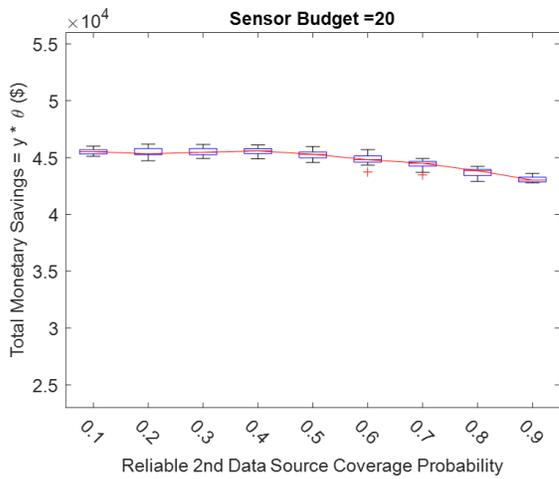


(a)

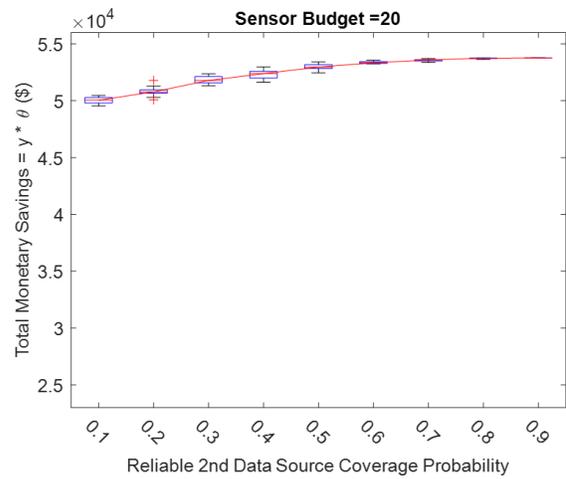


(b)

Figure 47: Total Travel Time Error Savings with 16 Sensors under Different Levels of 2<sup>nd</sup> Data Source Coverage: (a) Policy 1; (b) Policy 2.



(a)



(b)

Figure 48: Total Travel Time Error Savings with 20 Sensors under Different Levels of 2<sup>nd</sup> Data Source Coverage: (a) Policy 1; (b) Policy 2.

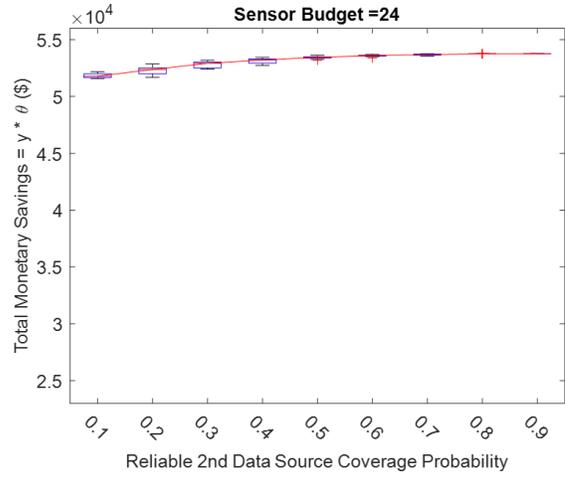
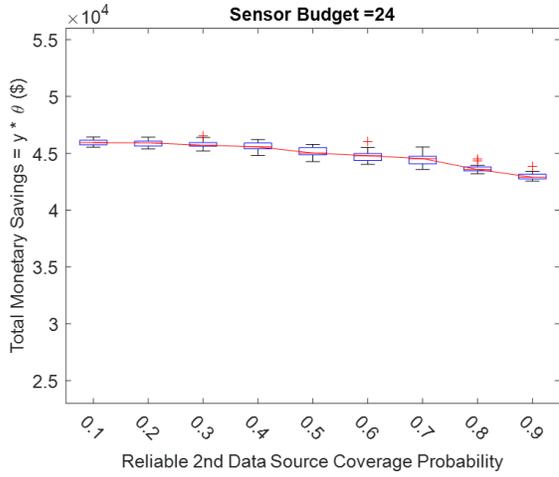


Figure 49: Total Travel Time Error Savings with 24 Sensors under Different Levels of 2<sup>nd</sup> Data Source Coverage: (a) Policy 1; (b) Policy 2.

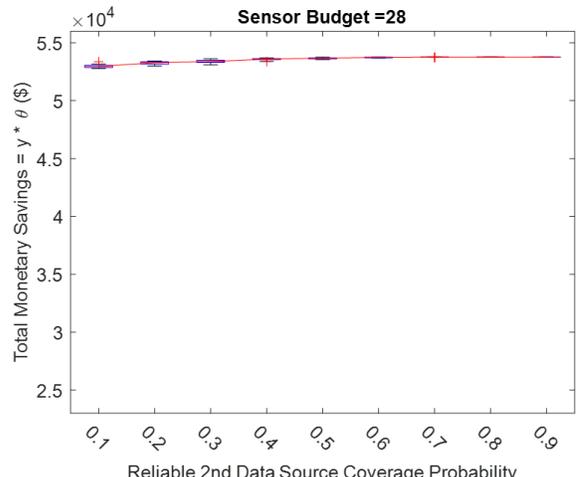
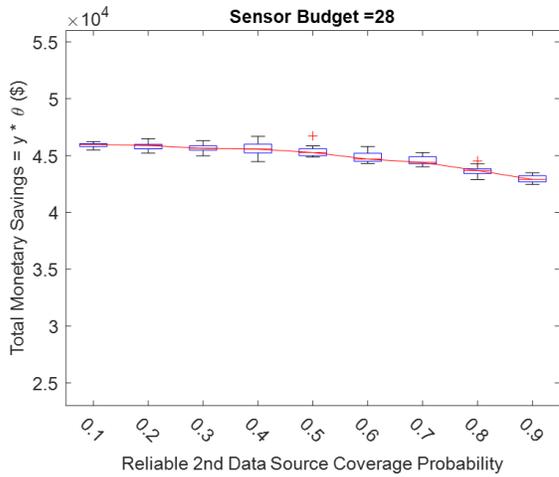


Figure 50: Total Travel Time Error Savings with 28 Sensors under Different Levels of 2<sup>nd</sup> Data Source Coverage: (a) Policy 1; (b) Policy 2.

#### 4.4 Summary and Conclusions

This chapter presented a traffic sensor network optimization model aiming to maximize the network-level surveillance benefit with real-time measurements. One important application of real-time traffic related measurements is to provide real-time travel time information for commuters. Travel state prediction for a given corridor can not only rely on data directly

measured from the location but can also be made based on data collected from other spatially correlated locations such as downstream and upstream highway segments. The optimization model presented here takes into account both types of predictions. With a given fleet of traffic sensors, the network-level traffic state prediction uncertainty reduction can be comprehensively evaluated based on empirical prediction performance for each corridor segment.

Moreover, considering prediction performance on each network segment as time-dependent, a dynamic sensor network planning model was formulated with a given time horizon. On the one hand, traffic state prediction performance is subject to the impacts of both recurrent and non-recurrent events. On the other hand, traffic volume traversing through the network exhibits a time-dependent pattern. Hence it is rational to explore the possibility of relocating sensors dynamically to bring more surveillance benefits over a time horizon.

Numerical experiments were conducted for a real-world network with travel time considered as the traffic state variable. Travel time prediction performance on each network segment within each representative particular time period was empirically obtained. In the case study, the goal of deploying sensors was to maximize the total travel time error reduction of the entire network. The optimization results in this research revealed three significant findings:

- Allowing relocation operations with a fixed fleet of traffic sensors can bring additional surveillance benefit. The marginal benefit of relocating sensors is more significant than that by adding new sensors when the fleet size is relatively large to the network size. However, when the fleet size is small, adding more sensors is preferable since relocating sensors will result in significant real-time information loss for some periods;
- Travel time prediction for one location based on real-time data feeds collected from other spatially correlated locations is not trivial when evaluating the network-level travel time error reduction. When the sensor fleet size is small compared to the network size, considering spatial information based prediction can bring significant total error savings;

- For some highway segments, spatial information based travel time prediction performance sometimes can outperform the one by temporal information based prediction. Therefore, we should not enforce that the real-time data collected from one segment must be used to predict the travel time for that segment. To maximize the network-level travel time prediction error reduction, one should intelligently make use of the spatial information based prediction and temporal information based prediction.

# Chapter 5 **Data Validation Oriented Sensor Network Planning:**

## **A Multistage Stochastic Optimization Approach**

### *5.1 Introduction and Problem Statement*

In the last chapter, an optimization framework to plan a particular sensor network was proposed with the objective to enhance the highway network real-time traffic state surveillance. This objective is of high interest in nearly every existing traffic sensor network planning studies. A common key component of such studies is a priori information should be learned at the beginning of each planning stage. Those a priori information is temporal traffic state variance and spatial covariance relationships across the entire highway segments within the deployment scope, prediction errors improvements from a post-deployment perspective, and traffic state estimation uncertainty enhancement by introducing particular types of monitoring devices. As it was thoroughly discussed in the previous two chapters, at least some historical data describing the traffic state variables should be obtained to estimate the surveillance benefits improvement.

In this chapter, we consider the sensor placement problem from a different perspective given the a priori information is completely missing. In other words, for a highway network with complete unknown historical traffic data and unknown GPS coverage, how operators should plan a sensor network to evaluate these a priori traffic information. Specifically, we consider the highway system is prone to GPS data coverage which might provide useful information and guidance for future planning work. But the reliability of the GPS data coverage is unknown and should be validated through other data sources (e.g., Bluetooth data collection based validation). This type of work has been intensively conducted in the I-95 Coalition Validation project in the past ten years. In that project, Bluetooth devices were deployed along major corridors of interest to validate the accuracy of aggregated GPS data provided by INRIX (i.e., Phase I project), and INRIX, HERE and TomTom (i.e., Phase II project). Based on the validation conducted with data

collected and aggregated through a two-week period, the GPS data reliability of a particular data provider was summarized and reported. Therefore, with the unknown knowledge on the GPS coverage reliability, one should deploy reliable traffic detection sensors in the system and conduct data collection operations for a given time period to obtain enough validation samples. In the problem studied here, we consider the deployment of sensors for GPS data reliability validation as a necessarily required operation to collect and learn the a priori information of a highway system. This is of key usefulness since reliable GPS data can further assist operators to evaluate the temporal-spatial traffic state patterns, and thus save the usage of additional physical traffic sensors in following stages.

Moreover, through the deployment of physical sensors into the highway system and the gradually revealed GPS data, one can collect and estimate both the temporal and spatial traffic state evolution patterns, which are particularly useful when planning a real-time surveillance sensor network. In particular, the temporal traffic state patterns of a corridor segment (e.g., travel time variance) can be learned when data collection devices are placed at the location. As well, when two different segments (e.g., upstream and downstream corridors) are monitored with such devices simultaneously, traffic state correlations can be obtained through some specific estimation approaches. Well-planned sensor deployment strategies can efficiently assist the operators to investigate and evaluate such types of a priori information.

Therefore, the problem targeted in this chapter can be briefly summarized as follows. For a given highway network with possibly reliable GPS data reports, determine the optimal number of traffic sensors and location-relocation operations to meet three requirements:

- (1) Validate GPS data reliability on each network link,
- (2) Collect reliable traffic state data and estimate the temporal state evolution patterns for each link, and,
- (3) Collect reliable traffic state data from any predefined pair of segments and estimate the spatial state evolution patterns.

This can be graphically demonstrated in Figure 51. The reason why GPS data should be validated at the same time is easy to understand. For the first two types of operations, i.e., GPS data validation and temporal state pattern evaluation, there are no additional operation efforts required. In other words, once devices are deployed along a segment for a period, and ground truth traffic state data is collected, both GPS data reliability and temporal traffic state evolution pattern can be learned. The additional benefit is that once GPS data is proved to be reliable for a specific segment, the spatial traffic state correlations between this segment and another segment (e.g., the downstream segment) can be learned by only placing sensing devices on the other segment in the next validation stage. In this way, usage of physical sensing devices can be saved in the overall planning horizon. Specific notations and details on model formulation will be given in the next section.

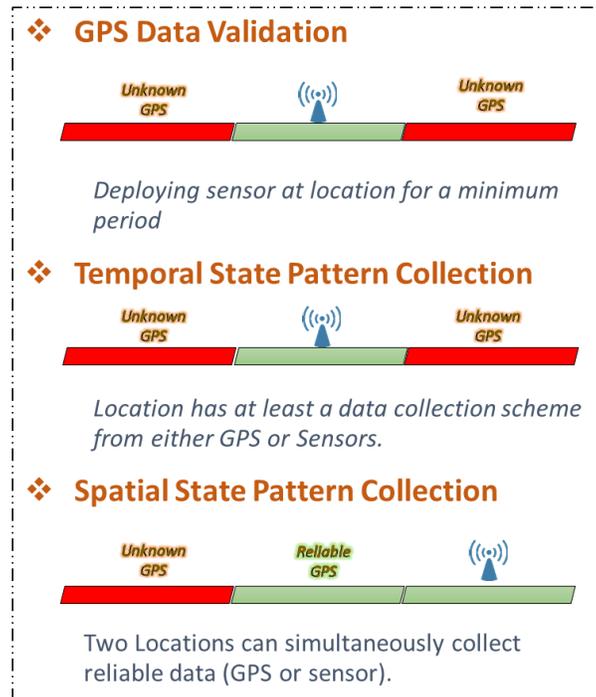


Figure 51: Data Validation based Network Optimization

## 5.2 *Preliminaries*

The mathematical formulation of the optimization model defined in this chapter is given in this section. It is noted that the GPS coverage and accuracy are considered as an unknown parameter in this problem. GPS data reliability on each segment within the network can only be revealed on the condition that data collection devices (i.e., sensors) have been deployed on the segment for a given time period  $t$ . Thus the problem itself turns to be a stochastic optimization with endogenous uncertainty on the GPS data reliability in each segment. Moreover, the temporal-spatial traffic state patterns across different network segments can be learned on the condition that sensing devices have been deployed to collect and monitor the traffic states for a minimum time period  $t$ . In this research, we assume the minimum duration of sensor monitoring period for all of the above three purposes are same and is denoted by a discrete variable  $t$ .

To begin with, associated parameters and variables are given and explained in the first subsection. A time-space network is adopted which assist in formulating the optimization problem as a multistage process in the second subsection. In the last subsection, the multistage stochastic programming model is given with the objective to minimize the total sensor cost and expected sensor relocation cost in the following subsequent stages.

### 5.2.1 Notations

A highway network  $Z$  can be represented by link set  $L$  and node set  $D$ , i.e.,  $Z = (L, D)$ . Each link  $l$  can be determined by a pair of nodes  $(l(e_1), l(e_2))$ , which are the endpoints of the link. For directed link, the sequence of the endpoints matters and can further indicate the link direction by specifying the start node and the end node. For the undirected links, the sequence of the end nodes does not matter. In the problem studied here, we consider the network as an undirected network since the identification and re-identification sensing system (e.g., Bluetooth detectors) is adopted for the data collection and validation jobs. Specifically, traffic data (e.g., travel time) on both directions of a highway link can be collected once a pair of detectors are deployed at the two

endpoints of the link. But this will not limit the generalization of the proposed sensor network placement model for directed networks. For example, sequence of the end nodes can be explicitly specified if one wants to distinguish the bi-directional links with an additional pair of sensors.

Considering the network size and limited budget on sensors, a time horizon  $T$  with discrete stages  $t$  is adopted here. This consideration will facilitate sensor relocation operations in different time periods in order to meet the data validation goal with limited number of sensors. In addition to the GPS data quality validation and temporal traffic state data collection (i.e., deploying sensors on a specific segment), we further introduce the set  $Q = \{(l, k) | l, k \in L\}$  to denote the set of link pairs on which traffic states are potentially correlated and are of interest for spatial data pattern validation. For example,  $l$  and  $k$  might be two adjacent corridor segments. Here, we consider this link pair set is predefined by the operators. The larger the set cardinality is, the more spatial traffic state evolution patterns can be obtained through the data collection process.

As is introduced early in this chapter, the GPS data reliability on each segment is unknown and can only be revealed after using the sensors to validate the data on site. Thus this information is considered as a stochastic parameter denoted as  $\xi_l$ . It has two realizations, i.e., 1 means reliable and 0 means unreliable. For the optimization model with such stochastic parameters, we use  $\omega$  to denote a particular random scenario, which will be used to decompose the stochastic optimization model to a scenario-aggregation based deterministic optimization model. Further, we use  $Relo_\omega(t)$  to denote the optimal number of sensor relocations at the end of the  $t^{th}$  stage with respect to scenario  $\omega$ .

There are two sets of primary decision variables. The first one is the number of sensors purchased at the beginning of the planning horizon. This is an integer decision variable indicating the capital cost for the initial stage. The second set of primary decision variables  $x_{i,j}^t$ , indicate whether the sensor deployed at location  $i$  during stage  $t$  should be moved to location  $j$  at the end

of stage  $t$ . There are two indications of this decision variable when it takes value 1. If  $i = j$ ,  $x_{i,j}^t = 1$  means there is no operation on the sensor deployed at location  $i$  at the end of period  $t$ . Otherwise,  $x_{i,j}^t = 1$  indicates the sensor should be relocated to location  $j$  at the end of period  $t$ . In the latter case, a positive operational cost will incur.

In addition to the two primary decision variables, some auxiliary binary variables are additionally defined to facilitate and linearize the model formulation. These variables and associated constraints will be further introduced in the formulation section. Detailed explanations of necessary notations are given in Table 12.

Table 12: Notation Preparation for the Type-II Sensor Network Planning Model.

$L$	Set of network segments (i.e., links)
$l, k$	Segment index
$D$	Set of network nodes (i.e., potential sensor installation locations)
$i, j$	Node index taking value from $D$
$l(e_1), l(e_2)$	Endpoints of segment $l$ (i.e. locations for the pair of identification sensors)
$o'$	The index of depot initially storing the identification sensors
$T$	Maximum number of discrete periods considered for the validation project
$t$	Time period index taking values from $\{1, 2, \dots, T\}$
$Q$	Set of segment pairs, the spatial traffic state relationship of which is of interest
$c_s$	The unit cost of an identification sensor
$c_r$	The unit cost of sensor relocation operation
$\xi_l(\omega)$	A stochastic binary indicator indicating whether segment $l$ has reliable GPS data report under scenario $\omega$
$\omega$	Scenario index
$Relo_\omega(t)$	Number of relocation operations in period $t$ under scenario $\omega$

Decision variables	
$N$	Number of identification sensors purchased for network information validation
$x_{o'i}^0 \in \{0,1\}$	Binary variable, $\forall i \in D$ , indicating deployment operation from depot to location $i$ at the beginning of period $t = 1$ ;
$x_{ij}^t \in \{0,1\}$	Binary variable, $\forall i, j \in D, \forall t \in \{1, 2, \dots, T - 1\}$ , indicating an identification sensor is relocated from $i$ to $j$ at the end of period $t$ ;
Auxiliary variables	
$y_i^t \in \{0,1\}$	Binary variable, $\forall i \in D, \forall t \in \{1, 2, \dots, T\}$ , indicating whether location $i$ is placed with an identification sensor in period $t$
$o_l^t \in \{0,1\}$	Binary variable, $\forall l \in L, \forall t \in \{1, 2, \dots, T\}$ , indicating whether segment $l$ is monitored with a pair of traffic identification sensors
$V_l^t \in \{0,1\}$	Binary variable, $\forall l \in L, \forall t \in \{1, 2, \dots, T - 1\}$ , indicating whether the traffic data of segment $l$ has been validated in any period from 1 to $t$ ;
$r_l^t \in \{0,1\}$	Binary variable, $\forall l \in L, \forall t \in \{1, 2, \dots, T\}$ , indicating whether the GPS data of segment $l$ in period $t$ can be reliably used;
$\phi_l^t \in \{0,1\}$	Binary variable, $\forall l \in L, \forall t \in \{1, 2, \dots, T\}$ , indicating whether one can obtain reliable traffic information from segment $l$ in period $t$ (either from GPS or sensors);
$\lambda_{lk}^t \in \{0,1\}$	Binary variable, $\forall (l, k) \in S, \forall t \in \{1, 2, \dots, T\}$ , indicating whether one can simultaneously obtain reliable traffic information from segment $l$ and $k$ in period $t$ ;

### 5.2.2 Time-space Network Representation

The multistage sensor location-relocation optimization problem can be represented using a time-space network given in Figure 52. As is shown, each optimization stage is represented by a stage

index  $t \in \{1, \dots, T\}$  and a set of target sensor installation locations. Further, to facilitate the location optimization process, a dummy node  $O'$  is constructed at the beginning of the planning stage.

For a location belonging to a particular time period  $t \in \{1, \dots, T\}$ , a sensor is deployed there given that an arc goes into the node. For the example in the given time-space network, two arcs are routed from the dummy node before stage 1 and this indicates a sensor is installed at location 1 and location 3 during the first stage. In between the first and second stage, one arc is routed from node 1 to node 1, and the other one is routed from node 3 to node  $N$ . This means location of first sensor will be unchanged in the next coming stage, i.e., it is still deployed at location 1, and the second sensor is relocated from location 3 to a different location, i.e., location  $N$ . In this way, the entire sensor location-relocation process can be fully represented and captured by the time-space network with predefined time horizon and location set.

The initial decision, which is the total number of sensors purchased, can be calculated as the summation of the arcs routing out from the dummy node (which can be viewed as a depot storing the sensors). The stage-dependent decisions in the future can be depicted by the arcs routing in between any two consecutive stages. Accordingly, the relocation cost in between any two adjacent stages  $t$  and  $t + 1$  can be estimated based on the relocation operations (i.e.

$\sum_{\forall i} \sum_{\forall j \neq i} x_{ij}^t$ ). Based on the time-space network representation and the notations previously defined, the mathematical representation of the sensor location-relocation optimization model is formulated and explained in the next subsection.

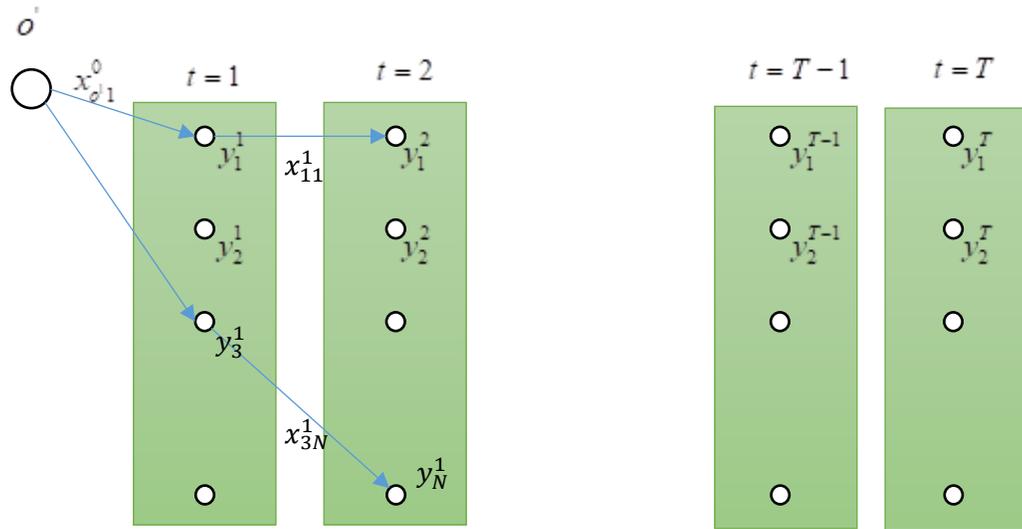


Figure 52: Time-space Representation.

### 5.3 Generalized Mathematical Formulation

Equation (5-1-1) gives the objective function of the multistage sensor location-relocation optimization model. The goal is to minimize the total cost associated with sensor purchase and subsequent relocation operations over a given time horizon. As is shown, the objective function consists of two parts. The first part denotes the capital cost determined by the number of sensors planned at the beginning of the deployment. The second part represents the expected total relocation cost spent at the end of each stage. Since the optimal relocation decisions are related to the location dependent GPS data reliability that is considered as a stochastic parameter in this model, the overall relocation cost of the entire time horizon is represented in an expectation form. In particular, the subscript  $\xi(\omega)$  in the objective function represents a random variable representing a specific scenario of GPS data coverage reliability. Therefore, the total cost of the multistage stochastic optimization model can be fully represented by equation (5-1-1) considering the existence of stochastic modeling parameters.

Objective function:

$$\text{Min: } c_s \cdot N + E_{\xi(\omega)} \left[ \sum_{t=1}^{T-1} \text{Relo}_{\xi(\omega)}(t) \right] \quad (5-1-1)$$

Subject to:

$$\text{Relo}(t) = \sum_{\forall i \in D} \sum_{\forall j \neq i} x_{ij}^t, \quad \forall t \in \{1, 2, \dots, T-1\} \quad (5-1-2)$$

$$N \leq |D| \quad (5-1-3)$$

$$y_i^1 = x_{oi}^0, \quad \forall i \in D \quad (5-1-4)$$

$$y_i^t = \sum_{\forall j \in D} x_{ji}^{t-1}, \quad \forall i \in D, \forall t \in \{2, 3, \dots, T\} \quad (5-1-5)$$

$$x_{oi}^0 = \sum_{\forall j \in D} x_{ij}^1, \quad \forall i \in D \quad (5-1-6)$$

$$\sum_{\forall j \in D} x_{ji}^{t-1} - \sum_{\forall j \in D} x_{ij}^t = 0, \quad \forall i \in D, \forall t \in \{2, 3, \dots, T-1\} \quad (5-1-7)$$

$$\sum_{\forall i \in D} y_i^t \leq N, \quad \forall t \in \{1, 2, \dots, T\} \quad (5-1-8)$$

$$2 \cdot o_l^t \leq y_{l(e_1)}^t + y_{l(e_2)}^t, \quad \forall l, \forall t \in \{1, 2, \dots, T\} \quad (5-1-9)$$

$$o_l^t + 1 \geq y_{l(e_1)}^t + y_{l(e_2)}^t, \quad \forall l, \forall t \in \{1, 2, \dots, T\} \quad (5-1-10)$$

$$V_l^t \leq \sum_{\tau=1}^t o_l^\tau, \quad \forall l \in L, \forall t \in \{1, 2, \dots, T-1\} \quad (5-1-11)$$

$$(|T|-1) \cdot V_l^t \geq \sum_{\tau=1}^t o_l^\tau, \quad \forall l \in L, \forall t \in \{1, 2, \dots, T-1\} \quad (5-1-12)$$

$$2 \cdot r_l^{t+1} \leq \xi_l + V_l^t, \quad \forall l \in L, \forall t \in \{1, 2, \dots, T-1\} \quad (5-1-13)$$

$$r_l^{t+1} + 1 \geq \xi_l + V_l^t, \quad \forall l \in L, \forall t \in \{1, 2, \dots, T-1\} \quad (5-1-14)$$

$$r_l^1 = 0, \quad \forall l \in L \quad (5-1-15)$$

$$r_l^t \geq r_l^{t-1}, \quad \forall l \in L, \forall t \in \{2, 3, \dots, T\} \quad (5-1-16)$$

$$V_l^t \geq V_l^{t-1}, \quad \forall l \in L, \forall t \in \{2, 3, \dots, T-1\} \quad (5-1-17)$$

$$2 \cdot \phi_l^t \geq r_l^t + o_l^t, \quad \forall l \in L, \forall t \in \{1, 2, \dots, T\} \quad (5-1-18)$$

$$\phi_l^t \leq r_l^t + o_l^t, \quad \forall l \in L, \forall t \in \{1, 2, \dots, T\} \quad (5-1-19)$$

$$2 \cdot \lambda_{lk}^t \leq \phi_l^t + \phi_k^t, \quad \forall (l, k) \in Q, \forall t \in \{1, 2, \dots, T\} \quad (5-1-20)$$

$$\lambda_{lk}^t + 1 \geq \phi_l^t + \phi_k^t, \quad \forall (l, k) \in Q, \forall t \in \{1, 2, \dots, T\} \quad (5-1-21)$$

$$\sum_{t=1}^T o_l^t \geq 1, \quad \forall l \in L \quad (5-1-22)$$

$$\sum_{t=1}^T \lambda_{lk}^t \geq 1, \quad \forall (l, k) \in Q \quad (5-1-23)$$

$$N \in \{1, 2, 3, \dots, |D|\} \quad (5-1-24)$$

$$y_i^t \in \{0, 1\}, \quad \forall i \in D, \forall t \in \{1, 2, \dots, T\} \quad (5-1-25)$$

$$x_{oi}^0 \in \{0, 1\}, \quad \forall i \in D \quad (5-1-26)$$

$$x_{ij}^t \in \{0, 1\}, \quad \forall i, j \in D, \forall t \in \{1, 2, \dots, T-1\} \quad (5-1-27)$$

$$o_l^t, r_l^t, \phi_l^t \in \{0, 1\}, \quad \forall l \in L, \forall t \in \{1, 2, \dots, T\} \quad (5-1-28)$$

$$V_l^t \in \{0, 1\}, \quad \forall l \in L, \forall t \in \{1, 2, \dots, T-1\} \quad (5-1-29)$$

$$\lambda_{lk}^t \in \{0, 1\}, \quad \forall (l, k) \in Q, \forall t \in \{1, 2, \dots, T\} \quad (5-1-30)$$

The constraints of the optimization model are given through equation (5-1-2) to equation (5-1-30). Equation (5-1-2) calculates the total amount of relocation operations at the end of stage  $t$  based on the decision variables  $x_{ij}^t$ . As is indicated, a relocation operation is taken from location  $i$  to  $j$  ( $j \neq i$ ) after stage  $t$  if the decision  $x_{ij}^t = 1$ . In such case, if  $i = j$ , it means no additional relocation operation is required at the end of stage  $t$  in terms of the sensor that has been deployed at location  $i$ .

Constraints ensure the total number of sensors should not exceed the number of optional installation locations in the network. Constraints (5-1-4) and (5-1-5) relate the operational decision variables (i.e.,  $x_{ij}^t$ ) and the installation decision variable (i.e.,  $y_i^t$ ). From the time-space network perspective, a location has a sensor installed at stage  $t$  if and only if a sensor was routed to this location at the end of stage  $(t - 1)$  from either itself or other locations.

Constraints (5-1-6) and (5-1-7) guarantee the location-relocation operations are consistent and feasible between any two consecutive stages. This pair of equality constraints describe the network flow conservation conditions from the time-space perspective. Also, constraint (5-1-8) guarantees the number of available sensors in each following stage should not exceed the number of sensors purchased at the beginning.

The rest of the constraints correlate decision variables and auxiliary decision variables, and linearly express the data validation requirements for the entire time horizon. Constraints (5-1-9) and (5-1-10) indicate a network segment is monitored through an identification-reidentification scheme if and only if the two sensors are deployed at the two endpoints of this segment, respectively. Accordingly, traffic data of a segment  $l$  can be collected (i.e.  $V_l^t = 1$ ) before stage  $(t + 1)$  if and only if sensors was installed for this segment. This data collection achievement indication is modeled by constraints (5-1-11) and (5-1-12). Constraints (5-1-13) and (5-1-14) describe that the GPS data source on a segment  $l$  can be viewed as a reliable data source for future (i.e.,  $[t + 1, T]$ ) validation purpose only if this segment does have a reliable GPS data coverage (i.e.,  $\xi_l = 1$ ) and the data from this segment has been validated in previous stages (i.e.,  $V_l^t=1$ ). But at the very beginning, the reliability of the GPS data source is unknown for every segment, and this is expressed by constraint (5-1-15). Constraints (5-1-16) and (5-1-17) are additionally incorporated to tighten the convex hull of the feasible solution region. In other words, if the GPS coverage reliability (or data validation) on segment  $l$  has already been revealed (or conducted) in previous stages, this information should be known as well in later stages.

Constraints (5-1-18) and (5-1-19) define whether the traffic data can be reliably extracted from a segment during stage  $t$  and indicate the successfulness of this fact by auxiliary variable  $\phi_l^t$ . Specifically, the traffic data from segment  $l$  can be successfully collected on the condition that at least one reliable data source can provide traffic information for this segment in stage  $t$ , i.e., either the GPS data source is reliable and known for segment  $l$  in stage  $t$  or segment  $l$  is exactly monitored with a pair of sensors.

Variable  $\lambda_{lk}^t$  denotes whether necessary data from both segment  $l$  and  $k$  can be simultaneously collected to investigate the spatial traffic state pattern between them in stage  $t$ . Constraints (5-1-20) and (5-1-21) express the above logical relationship. As is indicated, the spatial traffic state pattern can only be investigated on the condition that traffic state data from these two segments can be simultaneously collected in the same time period.

As is required, constraint (5-1-22) ensures that traffic information on each network segment should be validated through the overall time horizon. This is necessarily important to investigate the traffic patterns within a given highway network since the entire network-level traffic pattern consists of the traffic state evolution patterns from everywhere. Moreover, collecting traffic data from each network segment is necessary if one wants to validate the reliability and accuracy of the GPS probe data reported in such segments. Constraint (5-1-23), deals with spatial traffic state pattern collection and validation. In other words, in terms of the spatially correlated network segments (i.e.,  $\forall(l, k) \in Q$ ) that are of interest, the above constraint ensures the spatial traffic state patterns of those segments should be investigated as well within the time horizon. In practice, the target set consisting of spatially correlated network segments is usually declared by the traffic operators based on their interests, and are flexible in the optimization model. For example, every downstream-upstream segment pair should be taken into account as input to the optimization model if one wants to learn the upstream-downstream traffic state patterns everywhere within the network.

Finally, constraints (5-1-24) to (5-1-30) mathematically define the primary decision variables and the auxiliary decision variables.

#### 5.4 Solution Approach

In this section, the optimization model formulated above is analyzed, and a scenario decomposition-based solution approach is provided and discussed. At first, we reformulate the above multistage stochastic optimization model in a scenario decomposition-aggregation form. Then suitable non-anticipativity constraints are introduced and added into the scenario decomposition-based formulation specifically to tackle the endogenous stochasticity within the optimization process. In the end, we introduce the Monte Carlo simulation-based method to estimate the optimal solutions of the stochastic optimization model considering various sub-scenarios.

##### 5.4.1 Scenario Decomposition-Based Reformulation

The uncertainty of the proposed optimization model comes from the knowledge of the GPS probe data coverage reliability. This is represented by the uncertain parameter  $\xi_l$  displayed in the above mathematical formulation. This parameter takes value either 1 or 0.  $\xi_l = 1$  means network segment  $l$  has reliable GPS probe data as an alternative data source, and  $\xi_l = 0$  indicates the GPS probe data on the segment cannot serve as a reliable data source. This information is of key importance in the planning model. Specifically, once GPS probe data on a particular network segment is validated to be accurate and reliable, it can be used for plenty of purposes such as temporal traffic state pattern estimation and spatial traffic state correlation estimation without continuously deploying other traffic sensors at the same location.

Here, we use notation  $\omega$  to denote a specific realization of the uncertain parameters. With respect to each scenario  $\omega$ , the uncertain parameter  $\xi_l, \forall l \in \{L\}$  has a specific value indicating whether segment  $l$  has reliable GPS probe data source to reflect its traffic condition. Therefore,

the objective function defined in last section can be reformulated as shown in expression (5-2-1).

The first part of the objective function remains the same and represents the total capital cost to purchase traffic sensors. The operational cost given by the second part is updated by explicitly calculating and aggregating the sensor relocation cost under each scenario. In the scenario-aggregation based objective function,  $\pi_\omega$  is the occurrence probability of scenario  $\omega$ . Thus

$\sum_{\omega \in S} \pi_\omega \sum_{t=1}^{T-1} \text{Relo}^t(\omega)$  is interpreted as the expected sensor relocation cost over the entire planning

horizon summarizing various scenarios with anticipated scenario occurrence distributions.

In terms of the model constraints, a scenario index  $\omega$  is added to the formulation. As is provided below, each decision variable displayed in the original formulation is associated with a scenario index. Under each scenario, the original uncertain parameters turn to be deterministic. Thus the decisions associated with each particular scenario should be evaluated separately based on the parameters define by the scenario. Mathematical formulations of necessary constraints explicitly incorporating the scenario index are given in equations (5-2-2) to (5-2-28). As is noted, the basic structure of the entire constraints remains the same with that in the original formulation. The major difference is the application of the scenario index  $\omega$ . In other words, the original formulation is equivalently transformed into a scenario-aggregation form by specifying the occurrence probability of each scenario.

Scenario Aggregation based Formulation:

$$\text{Min: } c_s \cdot N + \sum_{\omega \in S} \pi_\omega \sum_{t=1}^{T-1} \text{Relo}^t(\omega) \quad (5-2-1)$$

Subject to:

$$\text{Relo}^t(\omega) = \sum_{\forall i \in D} \sum_{\forall j \neq i} x_{ij}^t(\omega), \quad \forall \omega \in S, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-2)$$

$$N \leq |D| \quad (5-2-3)$$

$$y_i^1(\omega) = \sum_{\forall j \in D} x_{ij}^1(\omega), \quad \forall \omega \in S, \forall i \in D \quad (5-2-4)$$

$$y_i^t(\omega) = \sum_{\forall j \in D} x_{ji}^{t-1}(\omega), \quad \forall \omega \in S, \forall i \in D, \forall t \in \{2, 3, \dots, T\} \quad (5-2-5)$$

$$\sum_{\forall j \in D} x_{ji}^{t-1}(\omega) - \sum_{\forall j \in D} x_{ij}^t(\omega) = 0, \quad \forall \omega \in S, \forall i \in D, \forall t \in \{2, 3, \dots, T-1\} \quad (5-2-6)$$

$$\sum_{\forall i \in D} y_i^t(\omega) \leq N, \quad \forall \omega \in S, \forall t \in \{1, 2, \dots, T\} \quad (5-2-7)$$

$$2 \cdot o_l^t(\omega) \leq y_{l(e_1)}^t(\omega) + y_{l(e_2)}^t(\omega), \quad \forall \omega \in S, \forall l, \forall t \in \{1, 2, \dots, T\} \quad (5-2-8)$$

$$o_l^t(\omega) + 1 \geq y_{l(e_1)}^t(\omega) + y_{l(e_2)}^t(\omega), \quad \forall \omega \in S, \forall l, \forall t \in \{1, 2, \dots, T\} \quad (5-2-9)$$

$$V_l^t(\omega) \leq \sum_{\tau=1}^t o_l^\tau(\omega), \quad \forall \omega \in S, \forall l \in L, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-10)$$

$$(|T|-1) \cdot V_l^t(\omega) \geq \sum_{\tau=1}^t o_l^\tau(\omega), \quad \forall \omega \in S, \forall l \in L, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-11)$$

$$2 \cdot r_l^{t+1}(\omega) \leq \xi_l^t(\omega) + V_l^t(\omega), \quad \forall \omega \in S, \forall l \in L, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-12)$$

$$r_l^{t+1}(\omega) + 1 \geq \xi_l^t(\omega) + V_l^t(\omega), \quad \forall \omega \in S, \forall l \in L, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-13)$$

$$r_l^1(\omega) = 0, \quad \forall \omega \in S, \forall l \in L \quad (5-2-14)$$

$$r_l^t(\omega) \geq r_l^{t-1}(\omega), \quad \forall \omega \in S, \forall l \in L, \forall t \in \{2, 3, \dots, T\} \quad (5-2-15)$$

$$V_l^t(\omega) \geq V_l^{t-1}(\omega), \quad \forall \omega \in S, \forall l \in L, \forall t \in \{2, 3, \dots, T-1\} \quad (5-2-16)$$

$$2 \cdot \phi_l^t(\omega) \geq r_l^t(\omega) + o_l^t(\omega), \quad \forall \omega \in S, \forall l \in L, \forall t \in \{1, 2, \dots, T\} \quad (5-2-17)$$

$$\phi_l^t(\omega) \leq r_l^t(\omega) + o_l^t(\omega), \quad \forall \omega \in S, \forall l \in L, \forall t \in \{1, 2, \dots, T\} \quad (5-2-18)$$

$$2 \cdot \lambda_{lk}^t(\omega) \leq \phi_l^t(\omega) + \phi_k^t(\omega), \quad \forall \omega \in S, \forall (l, k) \in Q, \forall t \in \{1, 2, \dots, T\} \quad (5-2-19)$$

$$\lambda_{lk}^t(\omega) + 1 \geq \phi_l^t(\omega) + \phi_k^t(\omega), \quad \forall \omega \in S, \forall (l, k) \in Q, \forall t \in \{1, 2, \dots, T\} \quad (5-2-20)$$

$$\sum_{t=1}^T o_l^t(\omega) \geq 1, \quad \forall \omega \in S, \forall l \in L \quad (5-2-21)$$

$$\sum_{t=1}^T \lambda_{lk}^t(\omega) \geq 1, \quad \forall \omega \in S, \forall (l, k) \in Q \quad (5-2-22)$$

$$y_i^t(\omega) \in \{0,1\}, \quad \forall \omega \in S, \forall i \in D, \forall t \in \{1,2,\dots,T\} \quad (5-2-23)$$

$$x_{oi}^0(\omega) \in \{0,1\}, \quad \forall \omega \in S, \forall i \in D \quad (5-2-24)$$

$$x_{ij}^t(\omega) \in \{0,1\}, \quad \forall \omega \in S, \forall i, j \in D, \forall t \in \{1,2,\dots,T-1\} \quad (5-2-25)$$

$$o_l^t(\omega), r_l^t(\omega), \phi_l^t(\omega) \in \{0,1\}, \quad \forall \omega \in S, \forall l \in L, \forall t \in \{1,2,\dots,T\} \quad (5-2-26)$$

$$V_l^t(\omega) \in \{0,1\}, \quad \forall \omega \in S, \forall l \in L, \forall t \in \{1,2,\dots,T-1\} \quad (5-2-27)$$

$$\lambda_{lk}^t(\omega) \in \{0,1\}, \quad \forall \omega \in S, \forall (l, k) \in Q, \forall t \in \{1,2,\dots,T\} \quad (5-2-28)$$

#### 5.4.2 Non-Anticipativity Constraints Generation

Stochastic optimization problems can be classified into two categories in terms of the stochasticity property, namely exogenous uncertainty and endogenous uncertainty. An optimization model with exogenous uncertainty is a type of problem in which the stochastic parameters are independent of the decisions. In other words, the realization of the uncertain parameters is not affected by the internal optimization process. On the contrary, endogenous uncertainty means realizations of the stochastic parameters are somehow related to the decisions made during the optimization process. To be specific, for a multistage optimization model with endogenous stochastic parameters, knowledge of the uncertain parameters is gradually revealed with the stage-wise decisions making.

The mathematical sensor location-relocation model formulated in this study is intrinsically a multistage optimization model with endogenous stochastic parameters. Uncertainty originates from the knowledge of the GPS data source reliability for each network segment. Before conducting any data validation job on a segment, the GPS data source reliability remains

unknown. This information will be learned when sensors are deployed at the target location for a period to validate the probe data reliability. Therefore, the value of this stochastic parameter will be learned when a decision directly related to this parameter is made. Then, this parameter becomes a fixed parameter in the following optimization process.

Non-anticipativity constraints are required for scenario decomposition-aggregation based solution approach for the multistage optimization model with endogenous stochastic parameters. As is easy to understand, solutions for the above scenario aggregation-based formulation are prone to be over-optimistic if the dependency of decisions from different scenarios is ignored. In other words, optimal decisions for each scenario might be of high heterogeneity by solving the formulation given in the last subsection. The relationships of the stage-wise decisions in any two different scenarios should be explicitly considered to estimate the expectation of the stage-wise operational cost more conservatively. Therefore, non-anticipativity constraints are necessarily generated and should be added to the formulation.

The basic concept of non-anticipativity constraints illustrates that for any two different scenarios, if their decisions in previous stages are same, then the expected optimal decision in the next coming stage should be same. For the proposed formulation, non-anticipativity constraints are designed and explained as below.

In this problem, each network segment  $l$  is related to a binary stochastic parameter representing whether the GPS data source on the segment is reliable. Thus, for a network with  $|L|$  number of segments, there are  $|L|$  such parameters with uncertain value. Suppose  $\omega$  and  $\hat{\omega}$  are two scenarios with different realizations on some of the stochastic parameters, and  $L_d^{\omega, \hat{\omega}}$  is the set of stochastic parameters that have different realizations comparing  $\omega$  against  $\hat{\omega}$ . As is illustrated in the example with six stochastic binary parameters shown by Figure 53,  $\omega$  and  $\hat{\omega}$  are two scenarios with parameter  $\xi$  realized as  $\{0,1,0,1,1,1\}$  and  $\{1,1,1,0,1,1\}$ , respectively. Based on our definition,  $L_d^{\omega, \hat{\omega}} =$

$\{1,3,4\}$ , and consists of all parameters with different realizations across  $\omega$  and  $\hat{\omega}$ . On the contrary, the rest of the parameters  $L \setminus L_d^{\omega, \hat{\omega}} = \{2,5,6\}$  and consists of all stochastic parameters with exactly the same realizations across  $\omega$  and  $\hat{\omega}$ .

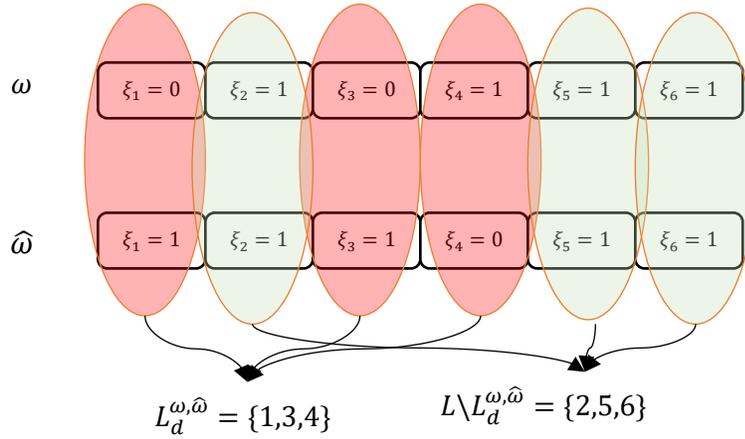


Figure 53: Non-anticipativity Constraint Generation Example in the Study.

Non-anticipativity constraints describe that if the decisions for  $\omega$  and  $\hat{\omega}$  in the previous stages  $\{1,2, \dots, t\}$  are same and they are only made associated with parameters  $\xi_l \in L \setminus L_d^{\omega, \hat{\omega}}$ , then the expected optimal decisions of  $\omega$  and  $\hat{\omega}$  in stage  $(t + 1)$  should be same. Following this concept and rule, we linearly formulate all the non-anticipativity constraints required in this optimization model. Necessary notations and auxiliary variables are introduced in Table 13. Mathematical expressions of the non-anticipativity constraints are given in equation (5-2-29) to equation (5-2-43).

Table 13:Notations for Non-anticipativity Constraints Generation.

$S$	Set of stochastic scenario realizations
$\omega, \hat{\omega}$	Scenario index, i.e., and $\hat{\omega} \in S$
$L_d^{(\omega, \hat{\omega})}$	Set of segments that have different realizations on the uncertainty parameter comparing scenario $\omega$ and $\hat{\omega}$
Auxiliary variable	

$Z_t^{\omega, \hat{\omega}} \in \{0, 1\}$	Binary variable indicating whether all the previous decisions related to the set $L_d^{(\omega, \hat{\omega})}$ before the end of time period $t$ are completely same for scenario $\omega$ and $\hat{\omega}$ .
---	--

$$Z_t^{\omega, \hat{\omega}} \leq 1 - o_l^\tau(\omega), \quad \forall(\omega, \hat{\omega}), \forall t \in \{1, 2, \dots, T-1\}, \forall l \in L_d^{\omega, \hat{\omega}}, \forall \tau \leq t \quad (5-2-29)$$

$$Z_t^{\omega, \hat{\omega}} \leq 1 - o_l^\tau(\hat{\omega}), \quad \forall(\omega, \hat{\omega}), \forall t \in \{1, 2, \dots, T-1\}, \forall l \in L_d^{\omega, \hat{\omega}}, \forall \tau \leq t \quad (5-2-30)$$

$$Z_t^{\omega, \hat{\omega}} \geq 1 - \sum_{l \in L_d^{\omega, \hat{\omega}}} \sum_{\tau=1}^t o_l^\tau(\omega), \quad \forall(\omega, \hat{\omega}), \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-31)$$

$$Z_t^{\omega, \hat{\omega}} \geq 1 - \sum_{l \in L_d^{\omega, \hat{\omega}}} \sum_{\tau=1}^t o_l^\tau(\hat{\omega}), \quad \forall(\omega, \hat{\omega}), \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-32)$$

$$Z_t^{\omega, \hat{\omega}} \in \{0, 1\}, \quad \forall(\omega, \hat{\omega}), \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-33)$$

$$x_{ij}^t(\omega) - x_{ij}^t(\hat{\omega}) \leq Z_t^{\omega, \hat{\omega}}, \quad \forall(\omega, \hat{\omega}), \forall i \in D, \forall j \in D, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-34)$$

$$x_{ij}^t(\hat{\omega}) - x_{ij}^t(\omega) \leq Z_t^{\omega, \hat{\omega}}, \quad \forall(\omega, \hat{\omega}), \forall i \in D, \forall j \in D, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-35)$$

$$y_i^{t+1}(\omega) - y_i^{t+1}(\hat{\omega}) \leq Z_t^{\omega, \hat{\omega}}, \quad \forall(\omega, \hat{\omega}), \forall i \in D, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-36)$$

$$y_i^{t+1}(\hat{\omega}) - y_i^{t+1}(\omega) \leq Z_t^{\omega, \hat{\omega}}, \quad \forall(\omega, \hat{\omega}), \forall i \in D, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-37)$$

$$r_l^{t+1}(\omega) - r_l^{t+1}(\hat{\omega}) \leq Z_t^{\omega, \hat{\omega}}, \quad \forall(\omega, \hat{\omega}), \forall l \in L, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-38)$$

$$r_l^{t+1}(\hat{\omega}) - r_l^{t+1}(\omega) \leq Z_t^{\omega, \hat{\omega}}, \quad \forall(\omega, \hat{\omega}), \forall l \in L, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-39)$$

$$\phi_l^{t+1}(\omega) - \phi_l^{t+1}(\hat{\omega}) \leq Z_t^{\omega, \hat{\omega}}, \quad \forall(\omega, \hat{\omega}), \forall l \in L, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-40)$$

$$\phi_l^{t+1}(\hat{\omega}) - \phi_l^{t+1}(\omega) \leq Z_t^{\omega, \hat{\omega}}, \quad \forall(\omega, \hat{\omega}), \forall l \in L, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-41)$$

$$\lambda_{lk}^{t+1}(\omega) - \lambda_{lk}^{t+1}(\hat{\omega}) \leq Z_t^{\omega, \hat{\omega}}, \quad \forall(\omega, \hat{\omega}), \forall(l, k) \in Q, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-42)$$

$$\lambda_{lk}^{t+1}(\hat{\omega}) - \lambda_{lk}^{t+1}(\omega) \leq Z_t^{\omega, \hat{\omega}}, \quad \forall(\omega, \hat{\omega}), \forall(l, k) \in Q, \forall t \in \{1, 2, \dots, T-1\} \quad (5-2-43)$$

### 5.4.3 Monte Carlo Scenario Aggregation Based Estimation

#### 5.4.3.1 Sampling-based Solution Evaluation

The stochastic optimization problem can be generalized and mathematically expressed in the following form:

$$g^* = \min_{x \in S} \{g(x) := E_P[G(x, \xi)]\} \quad (5-3-1)$$

where,  $x$  is the decision vector taking values from a finite set  $S$ , which is also called feasible solution space in optimization.  $\xi$  is a random vector representing a set of model parameters with uncertainty, and  $P$  denotes the joint probability distribution of these random parameters.  $G(x, \xi^l)$  is a real valued function with respect to vector  $x \in S$  and vector  $\xi^l \in \xi$ . Here  $E_P[\cdot]$  is the expectation calculator having the following form with respect to a set of random variables:

$$E_P[G(x, \xi)] = \int G(x, \xi) \cdot P(\xi) d\xi \quad (5-3-2)$$

Therefore, the optimization problem can be interpreted as finding the optimal solution  $x^*$  from the feasible space  $S$  by considering the statistical distribution of the random parameters  $\xi \sim P(\xi)$ .

In practice, the random vector  $\xi$  usually either takes values from an infinite set (i.e., continuous distribution) or a finite set with extremely huge enumerations (i.e., discrete distribution with many parameters and realizations). Moreover, in most cases, the estimator  $E_P[G(x, \xi)]$  cannot be expressed in closed form due to the complexity and high nonlinearity of the evaluation function  $G(\cdot)$ . Thus, optimizing the function with respect to decision vector  $x$  by considering all possible random realizations is not possible. Hence, sampling techniques should be taken into account to evaluate the optimization function.

Let  $\xi^l$  denote a specific realization of the random vector  $\xi$ , the optimal solution given by Equation (5-2-1) can be estimated by the following sampling-based estimator.

$$\widehat{g}_N^* = \min_{x \in S} \left\{ g(x) := \frac{1}{N} \sum_{l=1}^N G(x, \xi^l) \right\} \quad (5-3-3)$$

The optimization estimator provided by Equation (5-3-3) is named as sample average approximation (SAA) method (Kleywegt, Shapiro, and Homem-de-Mello 2002), and aims to find the optimal solution  $\hat{x}^*$  by considering a finite set of randomness realizations. Specifically,  $N$  is the amount of randomness realizations with respect to the random variable  $\xi$ , and the optimal function value is evaluated and aggregated based on the effectiveness of those  $N$  different set of random parameters.

As is theoretically proved in the study by Kleywegt, Shapiro, and Homem-de-Mello (2002),  $\widehat{g}_N^*$  converges to  $g^*$  with probability 1 as  $N \rightarrow \infty$ . Further, the convergence rate increases exponentially fast as  $N$  increases based on the theory of large deviations (LD). Therefore, one can evaluate the function optima under stochasticity with a finite set of random scenarios based on the estimator given by Equation (5-3-3). For optimization problem defined in this research, the objective function defined by Equation (5-2-1) is actually in agreement with this sampling based estimator when  $\omega$  takes values from a finite set of random realizations. To obtain a close-to-optimum solution with high quality (i.e. smaller optimal gap), one can dynamically increase the sampling power and incorporate more samples to solve the formulation.

As is suggested by Kleywegt, Shapiro, and Homem-de-Mello (2002), if the computational complexity of evaluating the SAA problem increases exponentially fast with the increase of sample size  $N$ , one should consider choosing a smaller sample size  $N$  and conduct multiple replications to evaluate the SAA problems. For example, the computational complexity of the scenario-aggregation based formulation given in this study increases exponentially fast as the sample size  $N$  increases due to the existence of the non-anticipativity constraints. Specifically, even though the number of decision variables linearly increases as the sample size  $N$  increases, the number of non-anticipativity constraints increases exponentially fast. As a consequence, the optimization complexity increases nonlinearly. In other words, one may not be able to resolve the

integer programming model within a reasonable time horizon. Therefore, a sampling-replication scheme should be designed in order to solve the optimization problem efficiently.

Here we define solving the optimization problem formulated in the last section (Equation (5-2-1) to (5-2-43)) as the Single-Replication SAA (SR-SAA) with sample size equal to  $N$ . As is discussed above, the general form of the SR-SAA can be expressed as Equation (5-3-3). To introduce the sampling-replication scheme, we define  $m \in \{1, 2, \dots, M\}$  as the replication index and  $N_m$  as the number of random samples generated in the  $m^{\text{th}}$  replication. Then the optimum estimator based on the Multiple-Replication SAA (MR-SAA) is given as the following equation:

$$\widehat{g}_{M,N}^* = \min_{x \in S} \left\{ g(x) := \frac{1}{M} \sum_{m=1}^M \frac{1}{N_m} \sum_{l=1}^{N_m} G(x, \xi^l) \right\} \quad (5-3-4)$$

As is provided in Equation (5-2-4), the optimum estimator consists of two nested parts. The inner-nested part calculates the expected optimal solution given by a set of random samples in a specific replication, and the outer part estimates the expected optima by aggregating the overall replications. By plugging Equation (5-3-3) into Equation (5-3-4), we can obtain the MR-SAA estimator as  $\widehat{g}_{M,N}^* = \min_{x \in S} \left\{ g(x) := \frac{1}{M} \sum_{m=1}^M \widehat{g}_{N_m}^* \right\}$ .

The remarkable advantage of MR-SAA is that it can significantly increase the optimization efficiency by breaking the sampling process into different replications when the computational complexity of the scenario-aggregation optimization problem exponentially increases as the sample size  $N$  increases (Kleywegt, Shapiro, and Homem-de-Mello 2002; Solak et al. 2010). However, unlike the estimator given by SR-SAA, for a finite sample size in each replication, the MR-SAA estimator is not guaranteed to converge to the true optima  $g^*$  when the replication number tends to infinite (i.e.  $N < \infty, M \rightarrow \infty \not\Rightarrow \widehat{g}_{M,N}^* \rightarrow g^*$ ). Instead, what obtained through the MR-SAA estimator is a lower bound of the global optima. The usage of MR-SAA is very problem dependent and is prone to underestimation (overestimation) risk for minimization (maximization) problem. Therefore, one should carefully evaluate the optimization gap between

the estimator and the true optima during the sampling-replication process. In other words, if estimation values still fluctuate to a large extent after huge replications or the gap to the optima is very large, the estimation from the MR-SAA scheme might not be reliable.

It is beneficial to learn the upper bound and lower bound of the stochastic optimization problem given structure of random parameters. In terms of the minimization-oriented optimization model formulated in this study, we can obtain both the lower bound  $g_L^*$  and the upper bound  $g_U^*$  of the objective function defined in Equation (5-1-1) by explicitly considering the positive and negative effects of the random parameters  $\xi$ , respectively. The following propositions provide two exact estimators of  $g_L^*$  and  $g_U^*$  on  $g^*$  in terms of the optimization problem formulated in this study:

**Proposition:**  $g_L^* = \min_{x \in S} \{g(x) := E_P[G(x, \xi = \vec{\mathbf{1}})]\}$  and  $g_U^* = \min_{x \in S} \{g(x) := E_P[G(x, \xi = \vec{\mathbf{0}})]\}$  are lower bound and upper bound of  $g^* = \min_{x \in S} \{g(x) := E_P[G(x, \xi)]\}$ , respectively.

Proof and interpretation of the above lower-upper bound proposition are obvious for the formulation given in this study. The random vector  $\xi$  here is a 0-1 vector and indicates the GPS data reliability for each network segment.  $\xi = \vec{\mathbf{1}}$  represents the ideal case that each segment has reliable GPS data source. In such a case, the number of sensors and the relocation frequency can be reduced to the largest extent compared to all other cases with  $\xi \leq \vec{\mathbf{1}}$ . Thus the optimal objective value under  $\xi = \vec{\mathbf{1}}$  is a lower bound when considering the expectation of all other possibilities. On the contrary,  $\xi = \vec{\mathbf{0}}$  is the worst case where none of the segments have reliable GPS data source. In such a case, the objective value tends to be the largest under the constraints to validate all of the spatial-temporal information across the entire network. Hence  $E_P[G(x, \xi = \vec{\mathbf{0}})]$  provides an upper bound of the optimization problem considering all different parameters realizations.

With the lower and upper bounds defined above, a Monte Carlo Simulation-based MR-SAA algorithm is designed to solve the stochastic optimization problem in this study. The solution procedure is described as below in Table 14:

Table 14: Monte Carlo Simulation Based MR-SAA Algorithm.

---

**Step 1:** Solve the optimization model for the lower bound  $g_L^*$  and upper bound  $g_U^*$  by setting the random parameter vector as  $\xi = \vec{\mathbf{1}}$  and  $\xi = \vec{\mathbf{0}}$ , respectively;

**Step 2:** Choose the replication limit  $M$  for conducting MR-SAA optimizations, such that  $\frac{1}{M+1}$  is sufficiently small;

**Step 3:** For  $m = 1, 2, \dots, M$ , run steps through 3.1 to 3.4:

**Step 3.1:** Randomly generate a set of samples with a size  $N_m$  based on the joint probabilistic distribution of the stochastic parameters  $P(\xi)$ ;

**Step 3.2:** Solve the SAA problem with objective value  $\widehat{g}_{N_m}^*$ , and estimate the variance of the objective values obtained so far as  $Var_{k=1\dots m}(\widehat{g}_{N_m}^*)$ ;

**Step 3.3:** Evaluate the estimator variance  $Var_{k=1\dots m}(\widehat{g}_{N_m}^*)$  against the  $\varepsilon$  small lower-upper bound gap  $\varepsilon \cdot |g_L^* - g_U^*|^2$ ;

**Step 3.4:** Check the stopping criteria: (1) whether  $m \geq M$  and (2)  $Var_{k=1\dots m}(\widehat{g}_{N_m}^*) \leq \varepsilon \cdot |g_L^* - g_U^*|^2$ . If any of the above two conditions is met, go to step 4. Otherwise, go back to step 3;

**Step 4:** If the variance of the objective values obtained through the above  $M$  replications is too large, increase the sample size  $N_m$  for each SAA with  $m = 1, 2, \dots, 3$  and redo Step 3; Otherwise, summarize and output the results from the MR-SAA process.

---

#### 5.4.3.2 Probabilistic Scenario Generation

In the above MR-SAA solution approach, an important part is the random scenario generation process used in Step 3.1. This random sample generation part is problem dependent and should be

carefully designed based on the distribution of the overall random parameters. For the problem studied in this research, the random parameters constitute a set of bool vector indicating the GPS data source reliability on each network segment. Considering the spatial correlations of vehicle trajectories within a transportation network, we propose a specific random scenario generation scheme to simulate the network-level GPS data source reliability distribution (Table 15).

Table 15: Random Scenario Generation Algorithm.

- 
- Step 1:** Set up the random number generator feed and the total sample number  $|S|$ ;
- Step 2:** Cluster the network segments into different classes:  $C_1 = \{l_1^1, l_1^1, \dots, l_{k_1}^1\}$ ,  $C_2 = \{l_1^2, l_2^2, \dots, l_{k_2}^2\}$ , ...,  $C_N = \{l_1^N, l_2^N, \dots, l_{k_N}^N\}$ , based on the roadway type (i.e., arterial or freeway) and the upstream-downstream relationship;
- Step 3:** Choose the probability threshold  $\{P_{C_1}, P_{C_2}, \dots, P_{C_N}\}$  indicating the occurrence of reliable GPS data source for each segment cluster;
- Step 4:** For each sample  $i = 1, 2, \dots, |S|$ , and each roadway class  $C_i$  in  $\{C_1, C_2, \dots, C_N\}$ , run step 4.1 to step 4.2:
- Step 4.1:** Based on  $P_{C_i}$ , randomly generate either  $\vec{\mathbf{1}}$  (i.e.  $\{r(l_1^i) = 1, r(l_2^i) = 1, \dots, r(l_{k_i}^i) = 1\}$ ) or  $\vec{\mathbf{0}}$  (i.e.  $\{r(l_1^i) = 0, r(l_2^i) = 0, \dots, r(l_{k_i}^i) = 0\}$ ) representing the GPS data source reliability for the entire class of segments;
- Step 4.2:** Separately conduct mutation operation on the segmental reliability indicator of each segment within the current class by a small probability threshold (e.g.  $p_{C_i}^m = 0.1$ ). That is,  $r(l_x^i) = \overline{r(l_x^i)}$  with probability  $p_{C_i}^m$ .
- Step 5:** If  $i = |S|$ , stop and output samples. Otherwise, set  $i = i + 1$  and go back to Step 4.
- 

The scenario generation algorithm randomly generates segmental level GPS data source reliability indicators by considering the overall network links into several clusters based on their location and geometric characteristics. Specifically, for highway segments belonging to the same

cluster such that they share similar traffic flow patterns and are spatially close to each other, they are likely having reliable (or non-reliable) GPS data coverage at the same time. This is described by the Step 5.1 in the above procedures. Moreover, we incorporate the mutation operations in the random scenario generation process as expressed in Step 4.2. This operator indicates that there is a chance that parts of the segments within the same cluster have the opposite GPS data source reliability compared with others. This additional operator adds more randomness to the generated samples and has meaningful real-world interpretation. For example, for a freeway corridor, there is no guarantee that each sub-segment has the same level of the GPS data report reliability even if they have similar roadway geometric characteristics and spatially correlated traffic volumes. Instead, there might be a small portion of segments having rare GPS data reports or highly fluctuating traffic reports due to the existence of large intersections or interchanges.

## 5.5 *Case Study and Numerical Experiments*

### 5.5.1 Preliminaries and Experimental Design

The proposed data validation-oriented sensor deployment optimization model is applied to a case study with the Washington D.C.-Baltimore commuting network that has been introduced in the last chapter. The real-world highway network consists of 88 directional corridor segments and 45 nodes. Each node denotes either a major intersection or an interchange within the commuting network, and each link represents the corridor segment between two consecutive nodes. The practical purposes of applying the proposed optimization model to the real-world network are threefold. First, deploying a particular type of traffic sensors to the network to validate the GPS data reports reliability on each corridor segment. Second, jointly using the deployed traffic sensors and reliable GPS data sources to collect and verify the traffic state evolution pattern on each segment. Third, using the sensors and GPS data reports to collect and verify the spatially correlated traffic states patterns across different segments. As was introduced early in this chapter, the GPS data source can be taken advantage of to conduct the following temporal-spatial

state pattern verification on the condition that the GPS data reports are validated to be reliable. Therefore, the first data validation purpose mentioned above is functionally correlated with the latter two purposes.

In this study, we choose the identification-reidentification technology-based sensors for data collection and validation. This type of sensors, such as Bluetooth sensors, have been widely adopted and applied for real-time traffic speed and travel time collection in past years due to the low-cost and high flexibility properties. Specifically, to monitor and collect traffic states on a highway segment, two sensors should be deployed and functioning at the two endpoints of the segment simultaneously. Applications of using identification-reidentification technology to collect and validate multi-source travel time and traffic speed data are described in Haghani, Hamed, and Sadabadi (2009), and Zhang, Hamed, and Haghani (2015).

In the related data validation projects, a two-week time horizon was empirically chosen as the minimum period to collect and validate GPS probe data from other independent data sources. This minimum data collection and validation period is referred as the minimum time period of each stage used in the proposed optimization model. In the developed stochastic optimization model, when we mention one sensor placement stage, we mean such a minimum sensor placement period. For example, by referring to the data validation project conducted by Haghani, Hamed, and Sadabadi (2009), one stage denotes two weeks, and two stages denote four weeks. In the following case studies, we applied and examined the proposed multi-stage optimization model for scenarios with three different time-horizon configurations, i.e., two-stage scenarios, three-stage scenarios, and four-stage scenarios. Here, two-stage based optimization means that one plans to finish all of the required data collection and validation processes within two stages (e.g., a four-week period). Accordingly, sensors can be relocated to new locations at the end of stage 1 and reused in stage 2.

For the GPS data source reliability on each segment, we consider zero a priori knowledge on its probability. In other words, any segment is considered to have reliable GPS data reports

with a probability equal to 0.5 and non-reliable GPS data reports with a probability equal to 0.5. In terms of GPS data validation and temporal-spatial traffic state patterns investigation, we set the coverage constraints in the way that, (1) each network segment should be covered by sensors at least for one-stage period (i.e., GPS data validation and temporal traffic state pattern investigation), and (2) the spatial traffic state evolution patterns between any two consecutive roadway segments (i.e., downstream and upstream segments) should be collected. The latter constraint requires that each pair of upstream and downstream segments should be simultaneously monitored by reliable traffic ‘sensors’ (i.e., either sensors or reliable GPS data reports).

To solve the optimization model proposed in this study, the mixed integer programming solver Gurobi 7.5.1 was used. The mathematical formulation was implemented with Python scripts. Platform for running the optimization is 64-bit Windows 10 operating system with Intel(R) Core(TM) i7 CPU @3.6 GHz and 16 GB RAM. Computational results are given and discussed in the next subsection.

### 5.5.2 Results Summary and Analysis

The proposed optimization model was applied to the DC-Baltimore commuting network to calculate the multi-stage traffic sensor deployment strategies and to test the optimization efficiency. In addition to the stage parameter, we also consider the purchase cost of the sensors (i.e., capital cost) and the relocation cost in subsequent stages (i.e., operational cost) and run the optimization model for different cost scenarios. Instead of fixing the value of each cost type separately, we consider the sensor-relocation cost ratio in the following experiments. Specifically, for sensor-relocation cost ratio equal to 1.0, we mean the unit sensor purchase cost is equal to the cost of relocating one sensor once. In other words, only the ratio of the cost matters and the absolute value of each cost does not affect the final solution. As is explicitly written in the objective function, the scenario that sensor-relocation cost is equal to \$1,000-\$100 should have the same optimal solution with the scenario that the sensor-relocation cost is equal to \$100-\$10.

Optimization results for 2-stage, 3-stage, and 4-stage planning problems are plotted and given in Figure 54, Figure 55 and Figure 56, respectively. In the case study, we further classify the problems based on four different sensor-relocation cost ratio, i.e., 1.0, 2.0, 5.0, and 10.0. Thus, the following figures display 12 optimization decisions. Specifically, for a case study with a particular stage number, there are further 4 set of optimization problems considering different cost parameters. As is shown in Figure 54, subplots given by (a), (b), (c), and (d) displays the decisions calculated by the MR-SAA approach with respect to sensor-relocation cost ratio equal to 1.0, 2.0, 5.0, and 10.0, respectively.

Considering the extremely heavy solution time imposed by the network size and the large number of constraints in the SAA formulation (e.g., over 50,000 constraints for the 2-scenario aggregation formulation), we set the sample size equal to 2 in each Monte Carlo replication and incorporated the non-anticipativity constraints for the two scenarios in the SAA formulation. Also, in each replication, we set the optimization stopping criterion as 5.0% optimum gap, which is also named as the 0.05-optimal aggregation gap (Kleywegt, Shapiro, and Homem-de-Mello 2002). As are displayed by the x-axis in the following three figures, the maximum replication number is 100, which indicates there are  $100 \times 2$  random scenarios generated and aggregated in the MR-SAA solution approach. For convenience, we denote this solution configuration as MR-SAA-N2.

Both sensor number decision variance and relocation decision variance are plotted in each figure to provide more details for the solution calculation process by MR-SAA algorithm. As are displayed in the optimization results for 2-stage, 3-stage and 4-stage problems, both of the two variances are very small compared with their expected values for scenarios in which the sensor-relocation cost ratio is greater than 1.0. For example, for the 2-stage optimization problem with sensor/relocation cost ratio equal to 5.0, the optimal sensor number decision is 20, and the expected relocation number is 14. Accordingly, the variance of the optimal sensor number and relocation number based on the MR-SAA algorithm are 0.16 and 0.18, respectively. Moreover, as

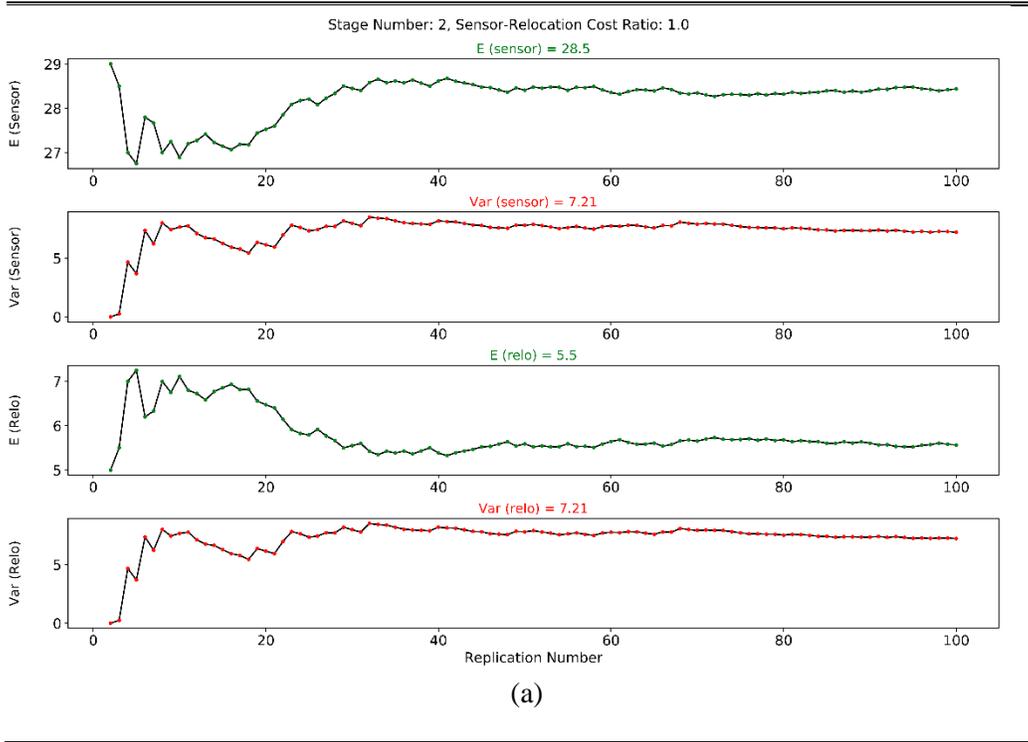
are displayed, these two variances become relatively stable compared with the expected values as the replication number increases. Therefore, the optimization results obtained through the MR-SAA approach are robust for the cases in this study.

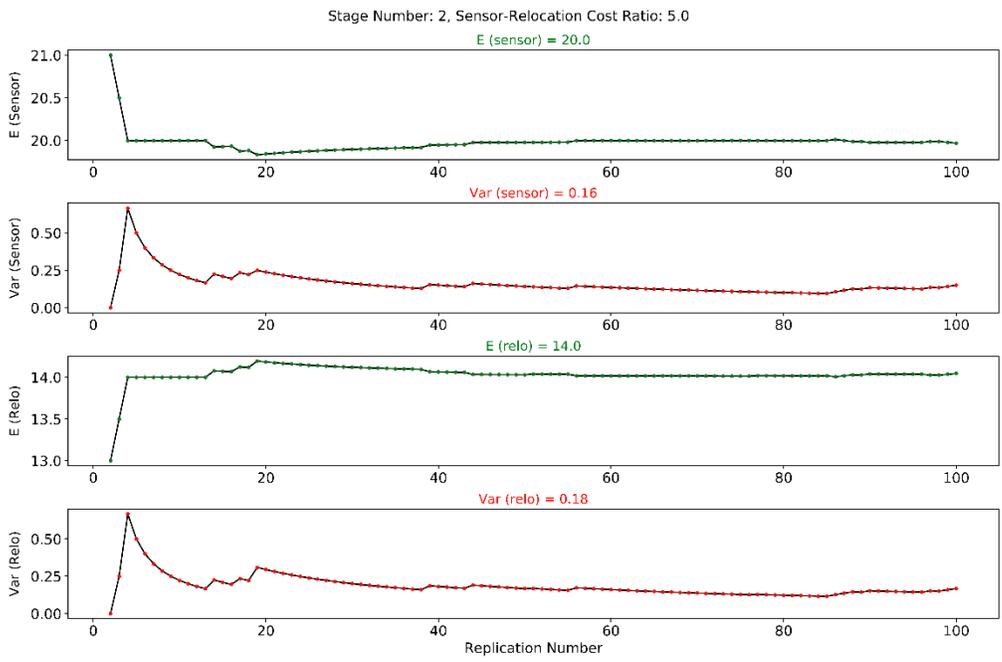
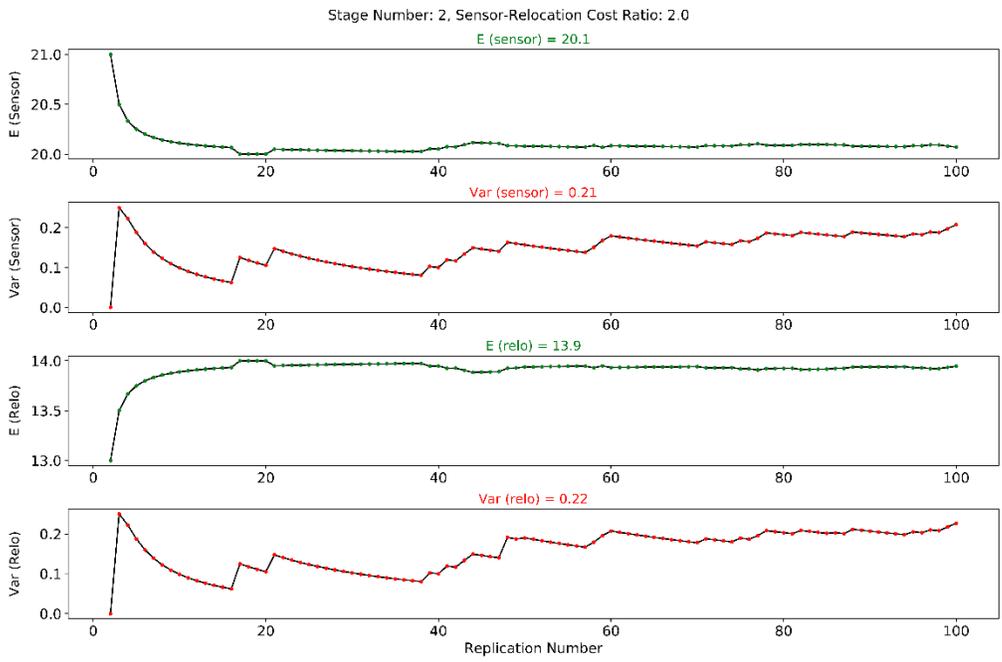
As one may note, for problems in which the sensor/relocation cost equal to 1, the optimization variances for both sensor number and expected relocation frequency are significantly large. When the cost ratio is 1, the variance of optimal sensor number and relocation frequency is 7.21 and 7.21, respectively, while the expected optimal sensor number and relocation frequency are 28.5 and 5.5, respectively. There is a very interesting fact one may find in Figure 54, Figure 55, and Figure 56 that the variance of sensor number and the variance of the total relocation frequency are the same for 2-stage, 3-stage, and 4-stage optimization problems.

This phenomenon also implicitly indicates why these two variances in the solution aggregation process are significantly large. If the unit sensor cost is equal to the unit sensor relocation cost, the optimal decisions impose no preference on purchasing more sensors or conducting more relocations. There might be two significantly different decisions that both are optimal considering the symmetry issue introduced by the problem itself. One solution might suggest purchasing more sensors but to conduct very few relocations in following stages, and the other solution might suggest purchasing few sensors but to conduct more relocations to achieve the same network coverage goal. Since the two costs are equal, both solutions have the same objective value.

This interesting phenomenon found by this case study provides an important insight for the multistage stochastic optimization problem and the application of the MR-SAA solution approach. Specifically, if the cost (or benefit) parameters of the initial-stage decisions are equal to the cost (or benefit) parameters of the subsequent-stage decisions, MR-SAA solution approach might result in a high variance for each type of decisions due to the existence of the extremely high number of optimal solutions. On the contrary, if there is a difference between the objective parameters of different types of decisions, MR-SAA with enough replications can provide robust

estimators for each type of decision even though the sample size in each replication is small. In other words, the symmetry issue of the optimal solution can significantly affect the sample aggregation-based solution process.





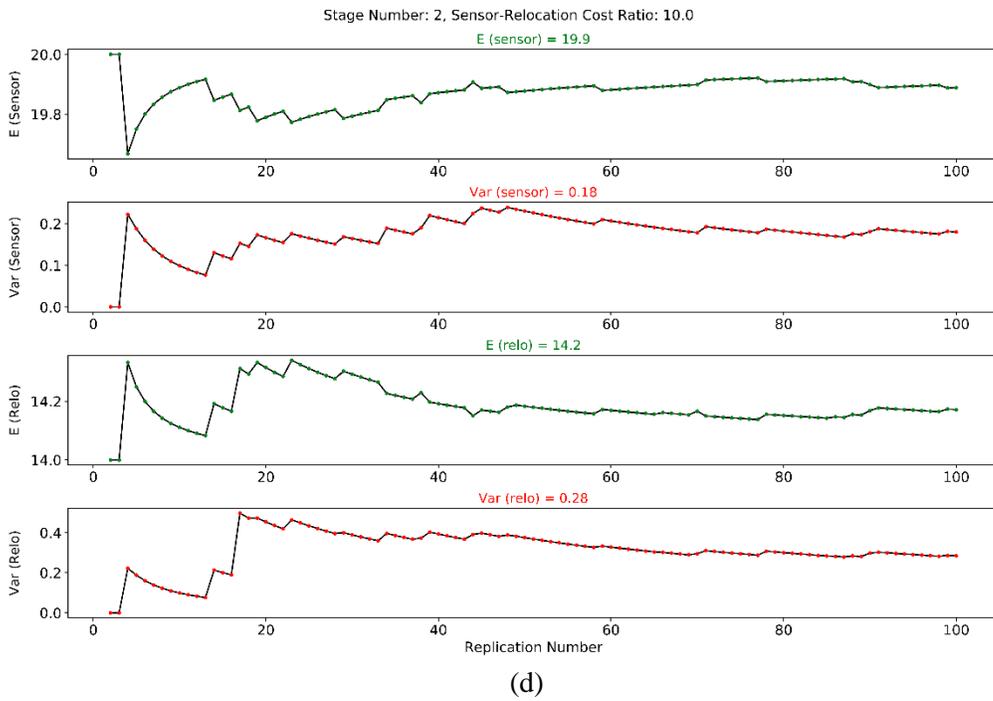
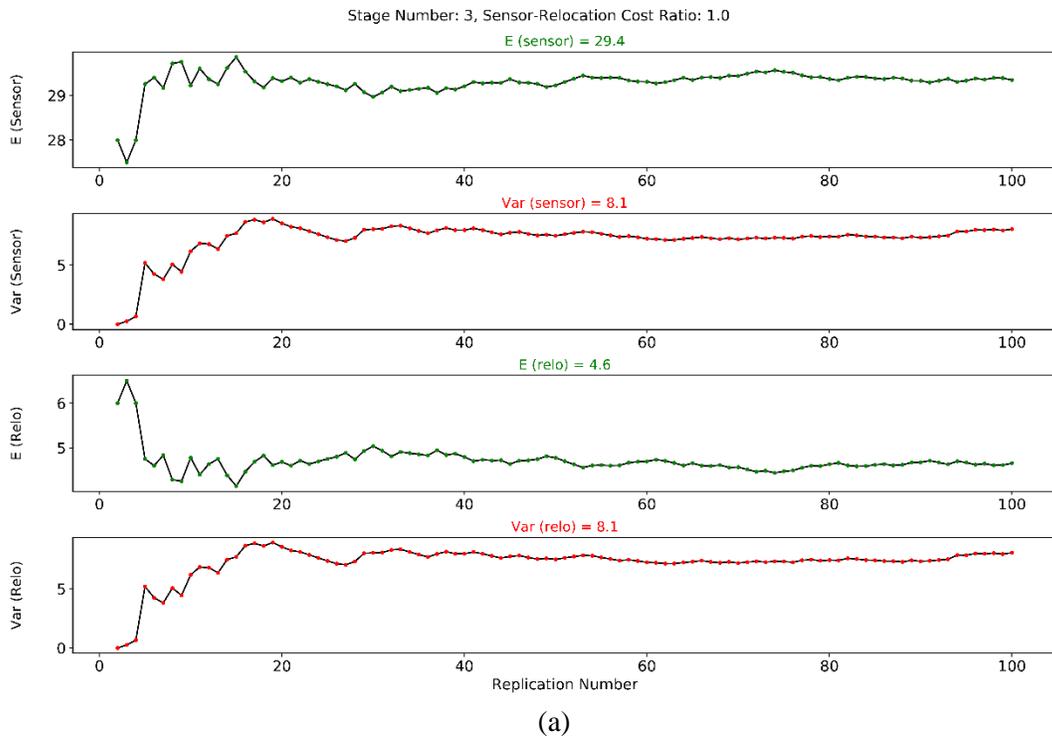
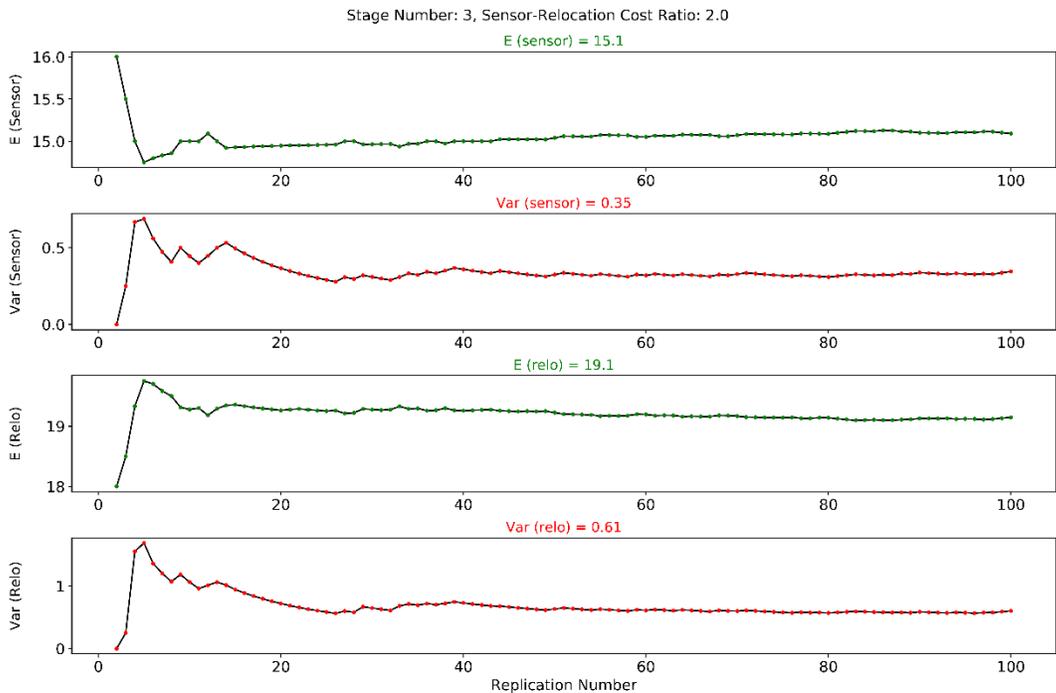
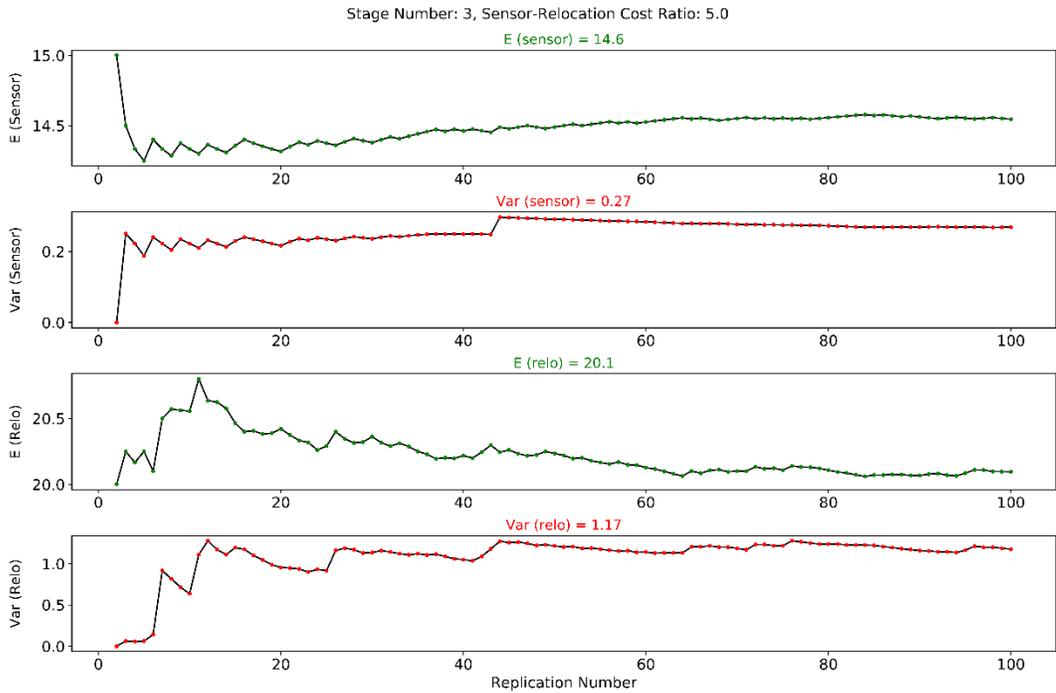


Figure 54: Solutions for 2-Stage Optimization by MR-SAA-N2 Algorithm: (a) Sensor/Relocation Cost Ratio=1.0; (b) Sensor/Relocation Cost Ratio=2.0; (c) Sensor/Relocation Cost Ratio=5.0; and (d) Sensor/Relocation Cost Ratio=10.0.

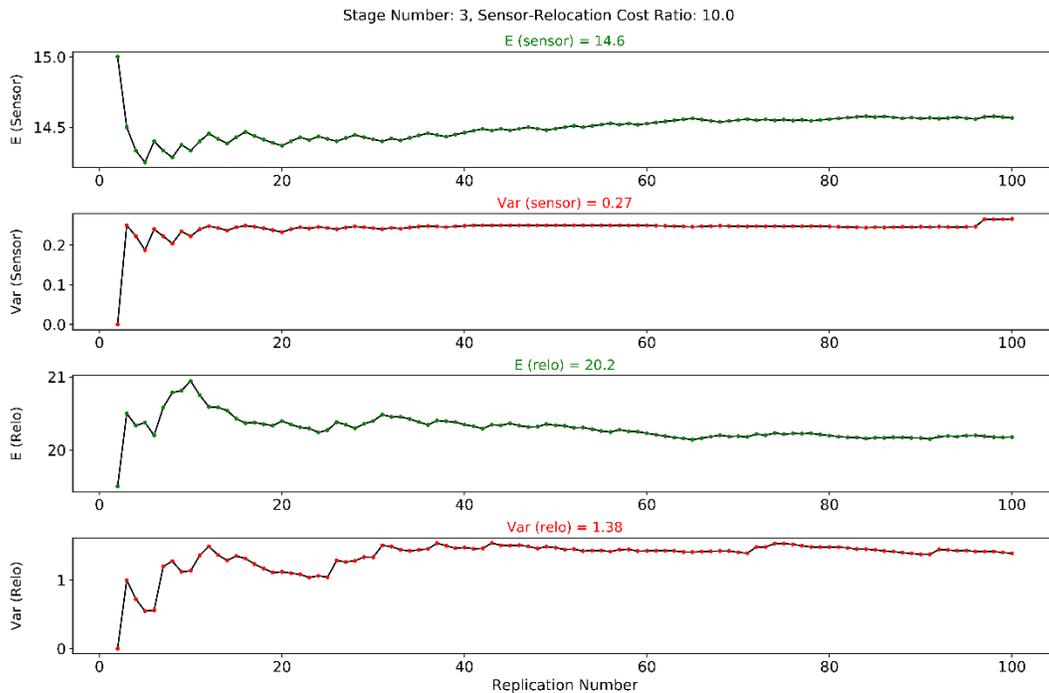




(b)

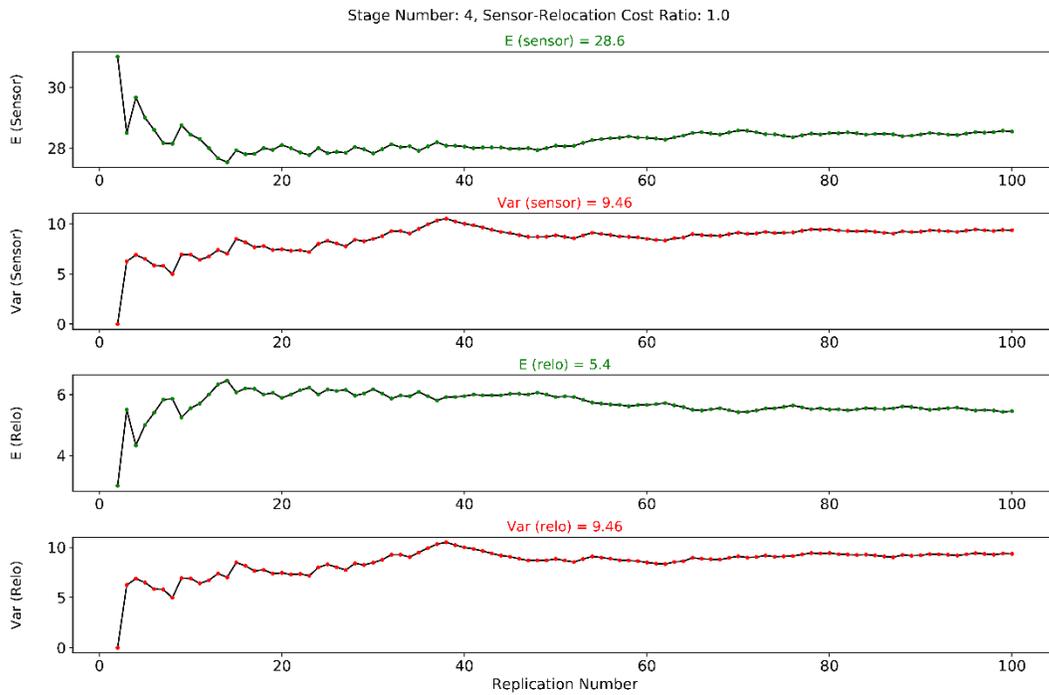


(c)

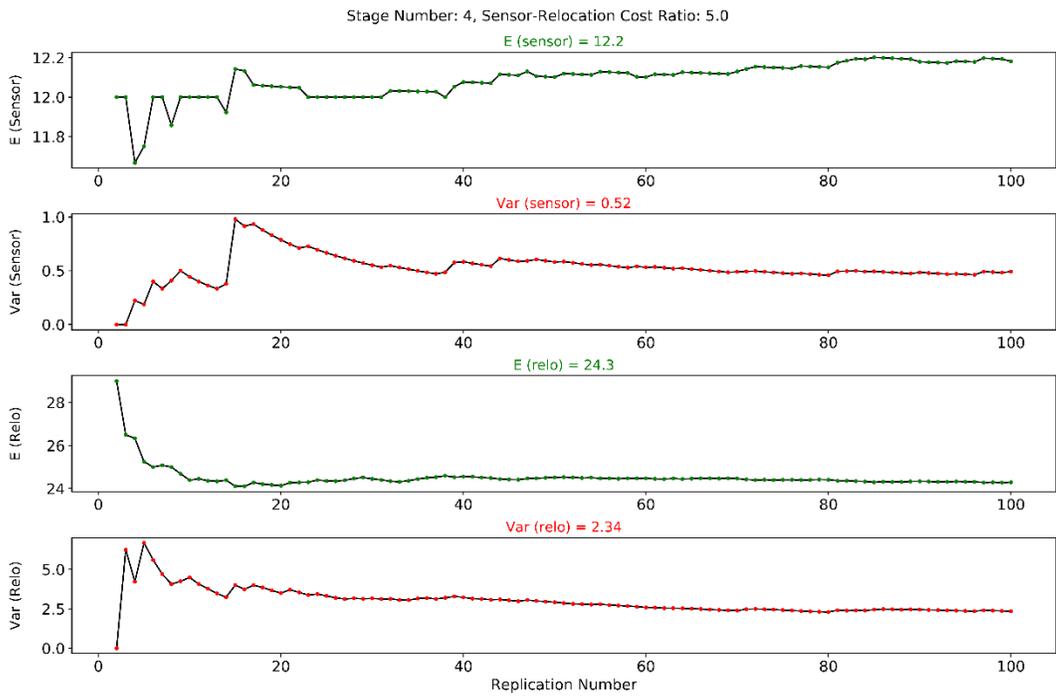
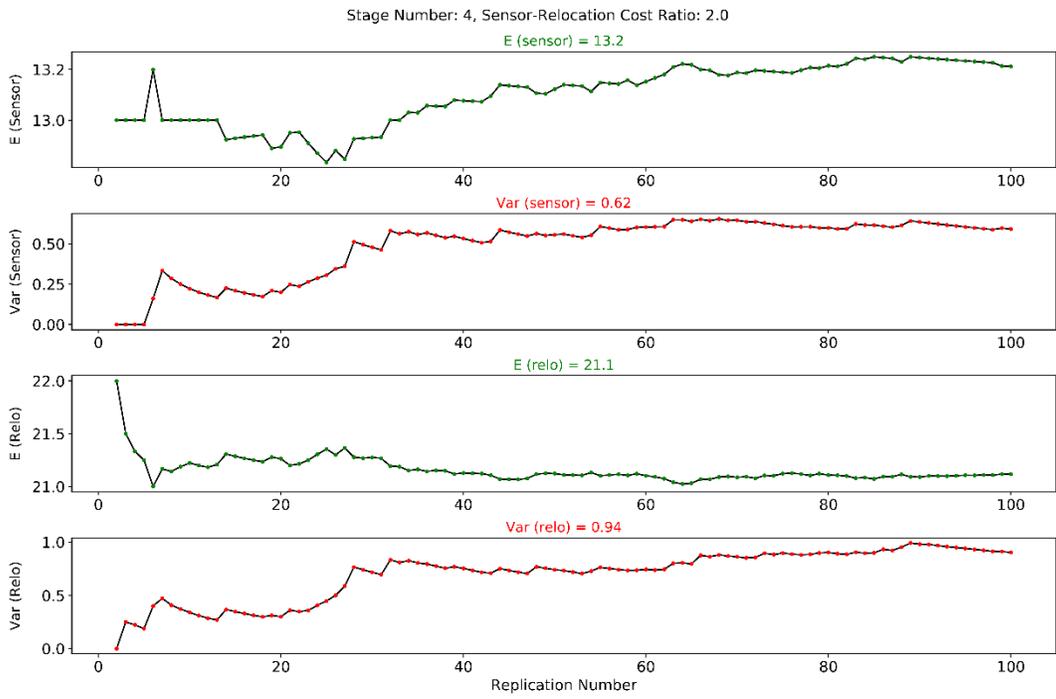


(d)

Figure 55: Solutions for 3-Stage Optimization by MR-SAA-N2 Algorithm: (a) Sensor/Relocation Cost Ratio=1.0; (b) Sensor/Relocation Cost Ratio=2.0; (c) Sensor/Relocation Cost Ratio=5.0; and (d) Sensor/Relocation Cost Ratio=10.0.



(a)



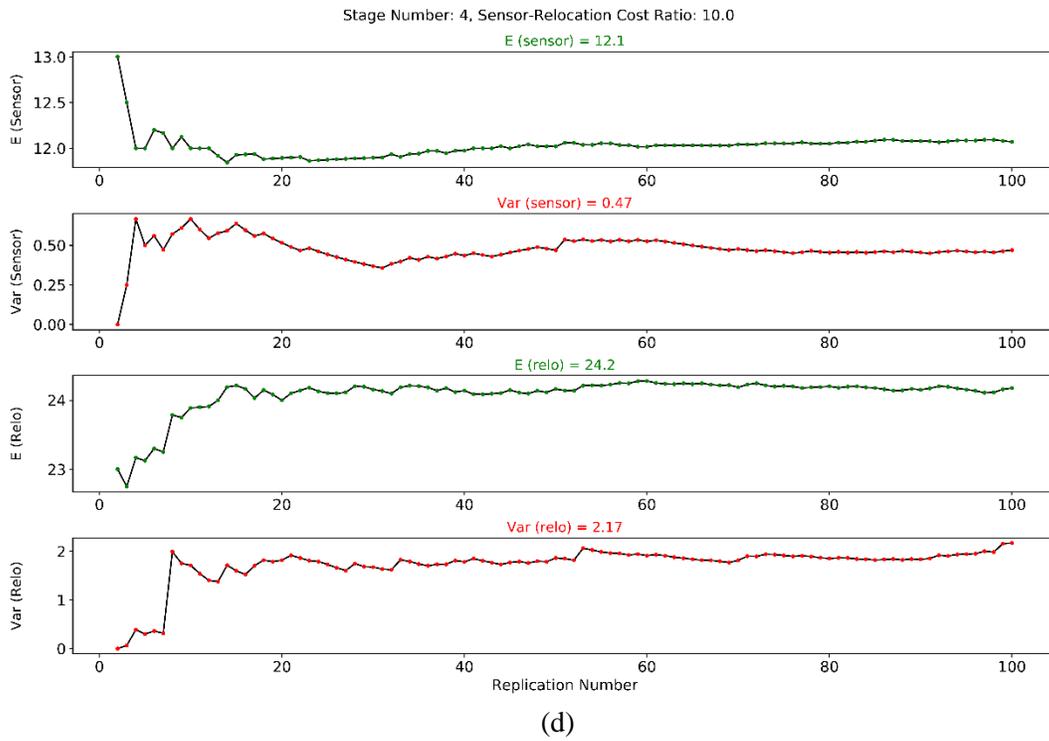
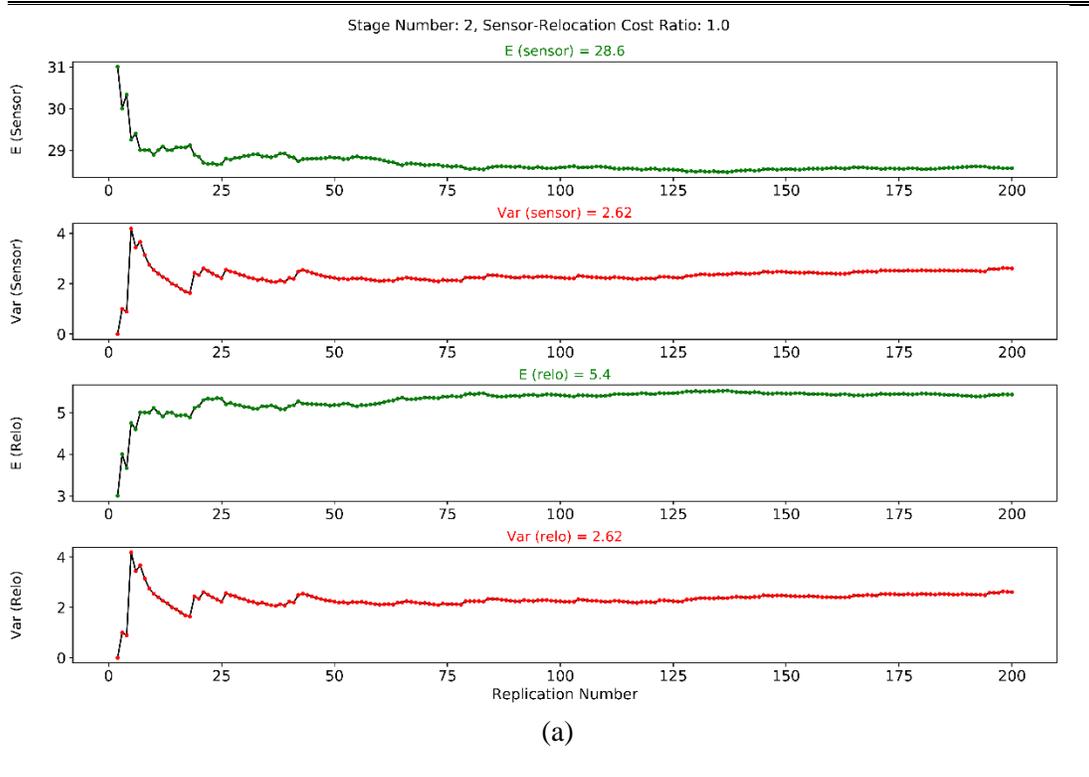


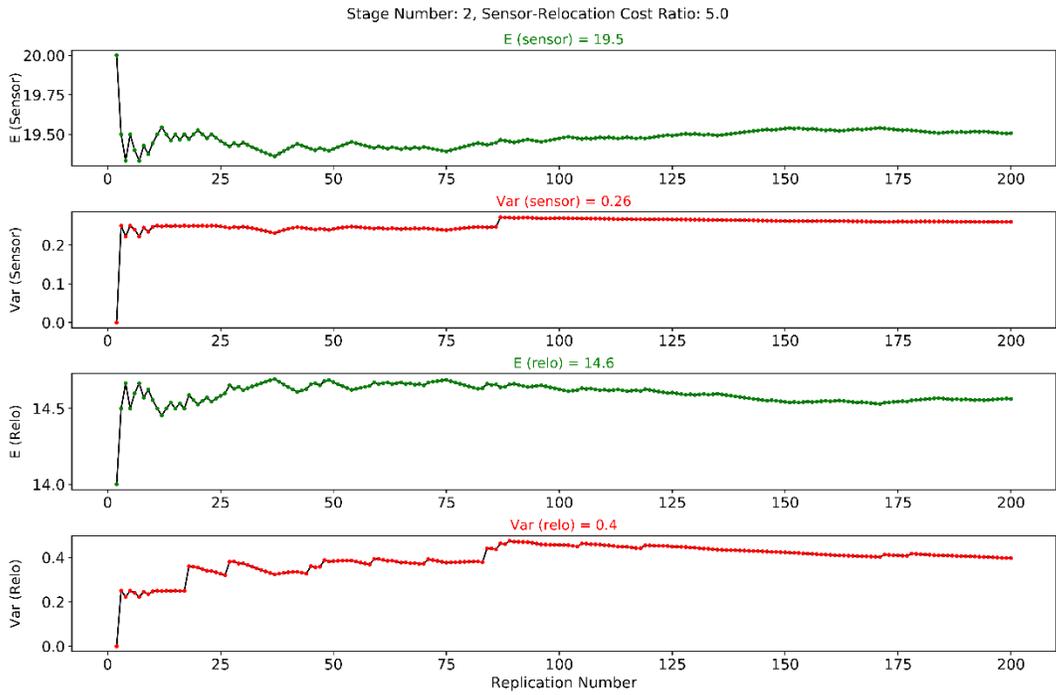
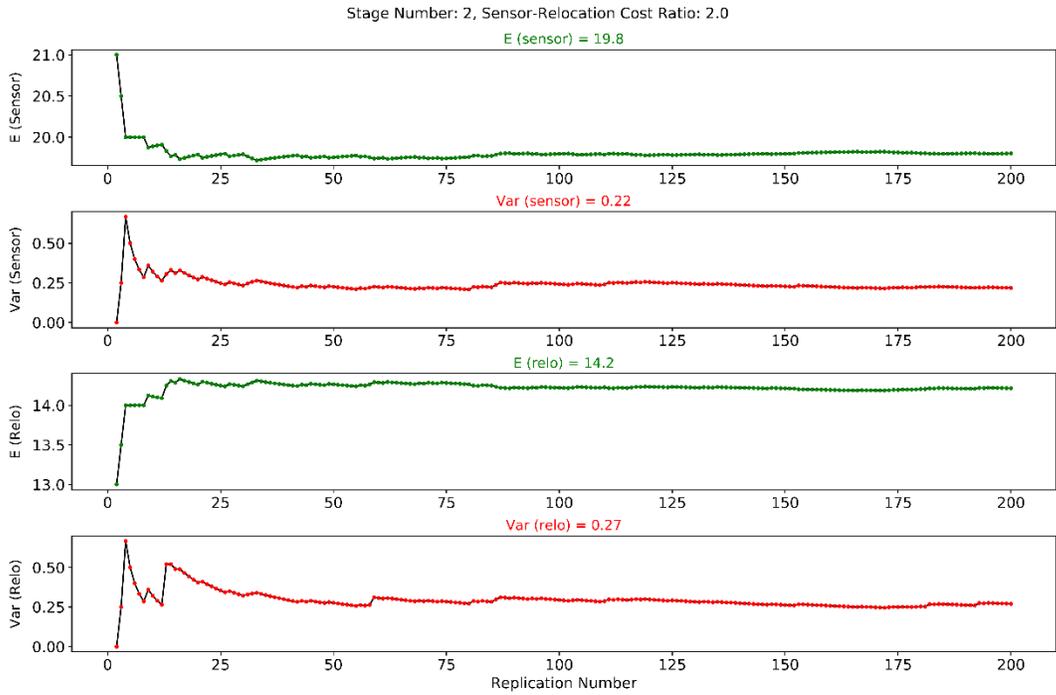
Figure 56: Solutions for 4-Stage Optimization by MR-SAA-N2 Algorithm: (a) Sensor/Relocation Cost Ratio=1.0; (b) Sensor/Relocation Cost Ratio=2.0; (c) Sensor/Relocation Cost Ratio=5.0; and (d) Sensor/Relocation Cost Ratio=10.0.

Further, we calculate the optimal solutions through an alternative configuration of the MR-SAA algorithm with the same amount of random samples, i.e., MR-SAA-N1. The concept of MR-SAA-N1 can be interpreted as aggregating the optimal solutions from each sample separately. Specifically, each replication only takes one random sample into account. Thus using the same random sample set, the MR-SAA-N1 will take two times replications of that by MR-SAA-N2 (i.e., 200 replications here). This is equivalent to the way that running the scenario-aggregation expanded optimization model (i.e., Equations (5-2-1) to (5-2-28)) with multiple scenarios but ignoring the non-anticipativity constraints. The final solutions Obtained by the MR-SAA-N1 approach for the 2-stage, 3-stage, and 4-stage optimization problems are displayed in Figure 57, Figure 58, and Figure 59, respectively.

It is clear to see, the solutions given by the MR-SAA-N1 approach are highly similar to those given by the MR-SAA-N2 approach. Specifically, the dispersions of both the sensor number decisions and the expected relocation frequencies under the same stage-cost

configuration are nearly within  $[-1, +1]$  interval. Also, for optimization problems with nonequivalent sensor-relocation cost, the variance of the initial-stage and sub-stage decisions are still very small. This indicates the robustness of the solutions obtained by the MR-SAA-N1 approach.





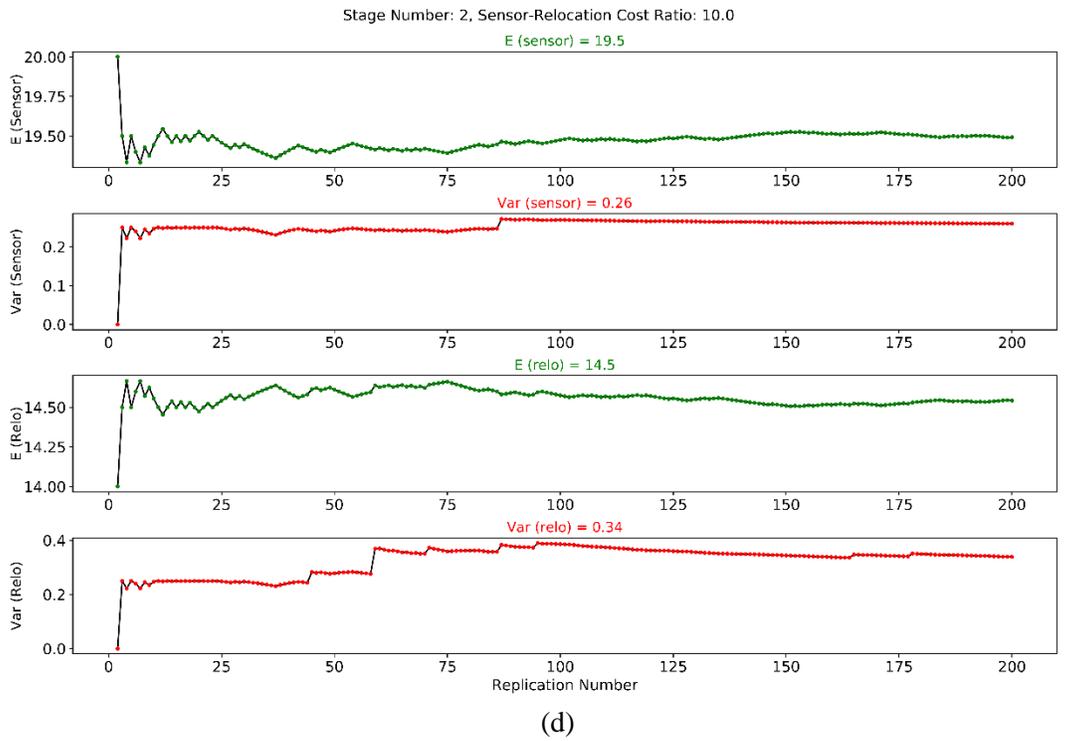
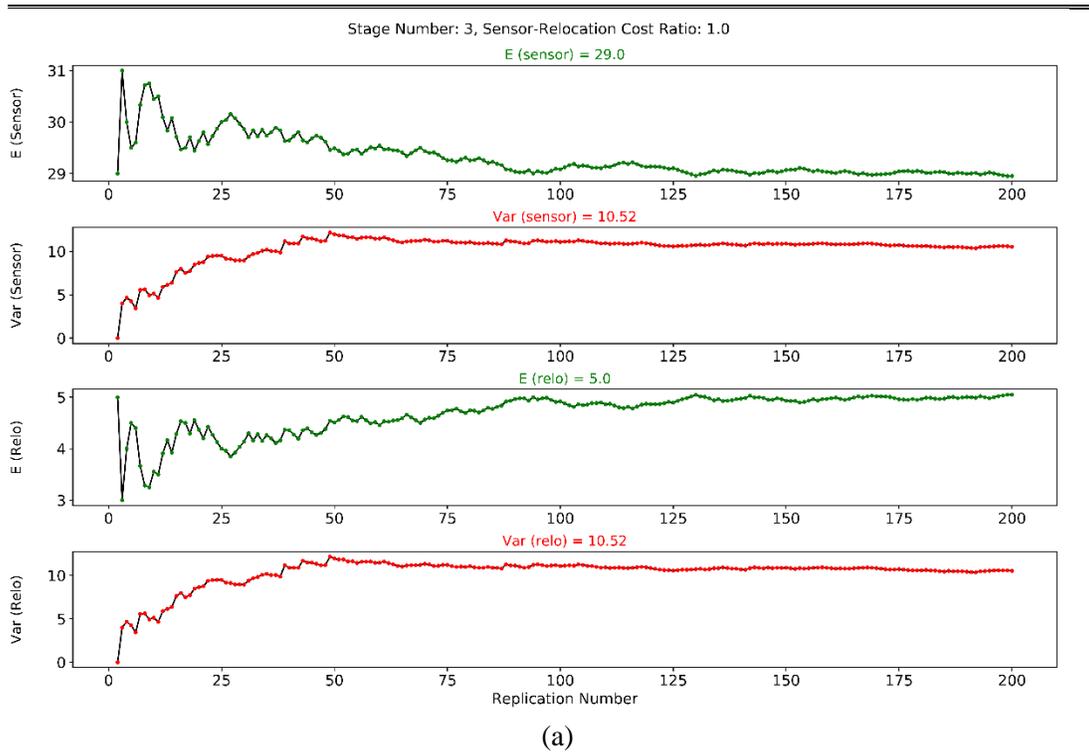
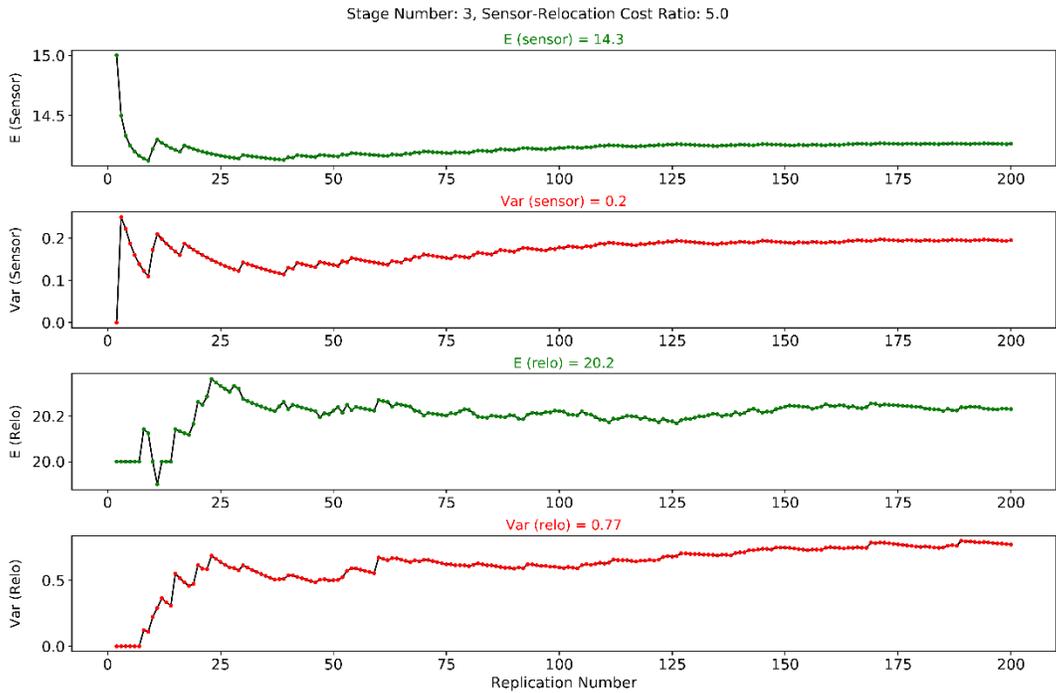
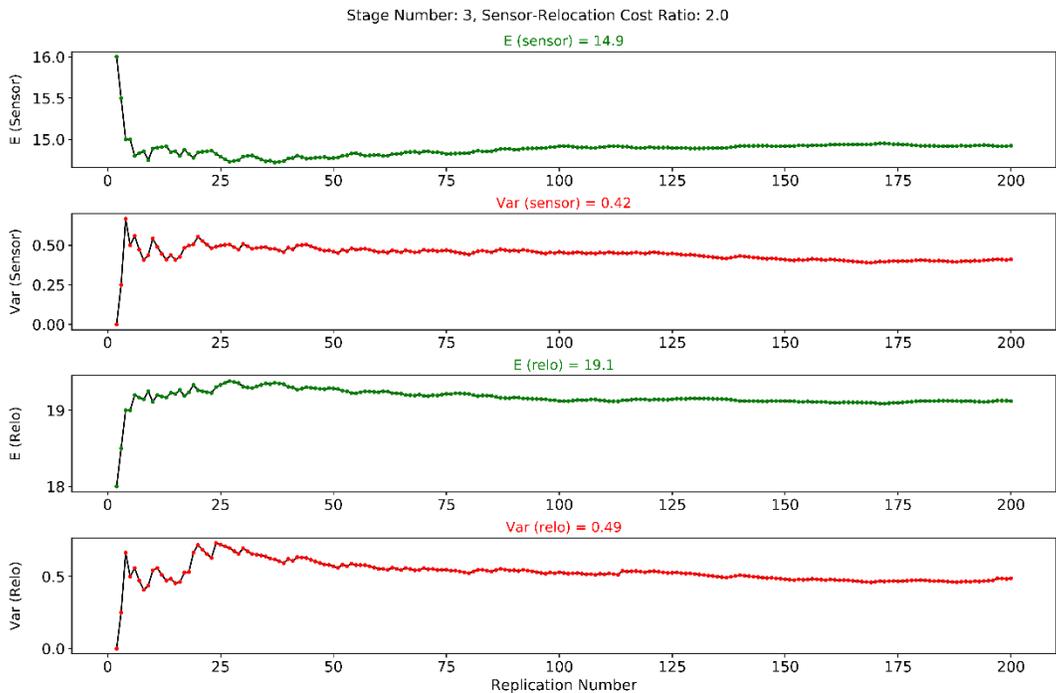


Figure 57: Solutions for 2-Stage Optimization by MR-SAA-N1 Algorithm: (a) Sensor/Relocation Cost Ratio=1.0; (b) Sensor/Relocation Cost Ratio=2.0; (c) Sensor/Relocation Cost Ratio=5.0; and (d) Sensor/Relocation Cost Ratio=10.0.





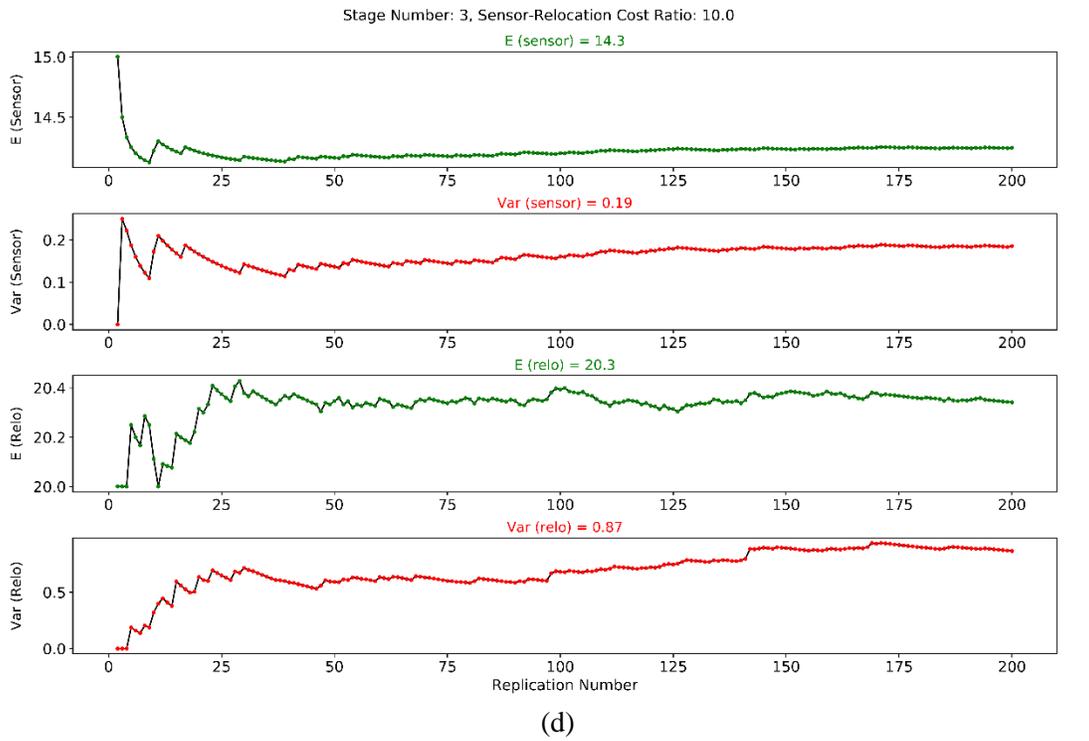
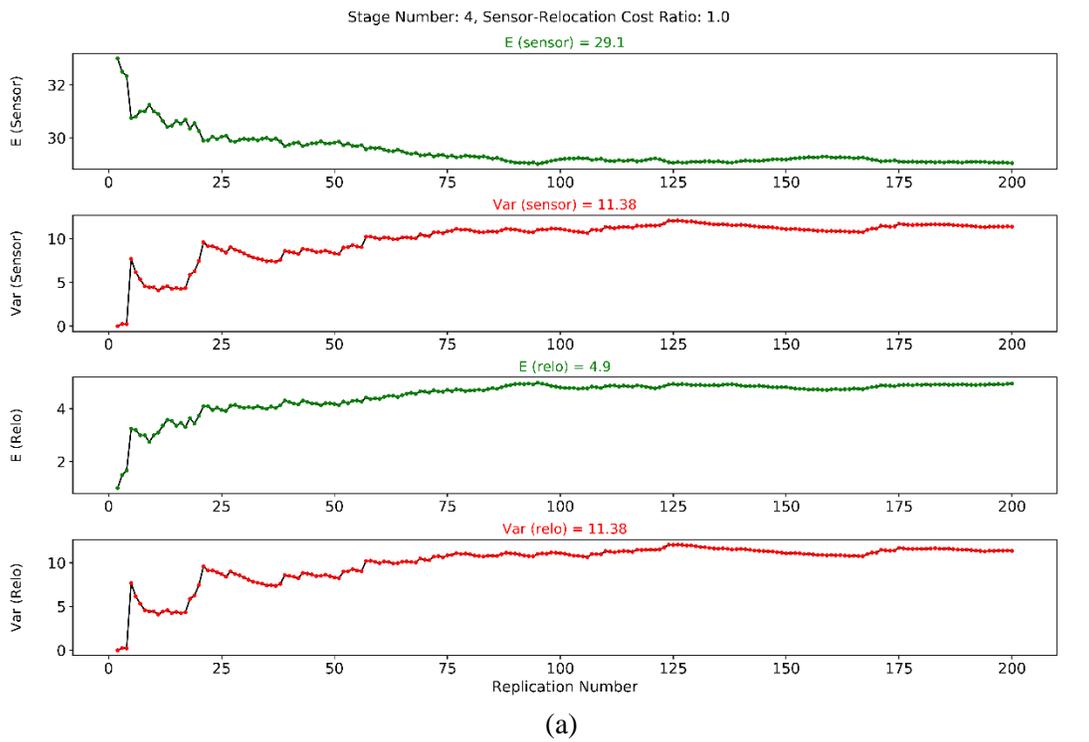
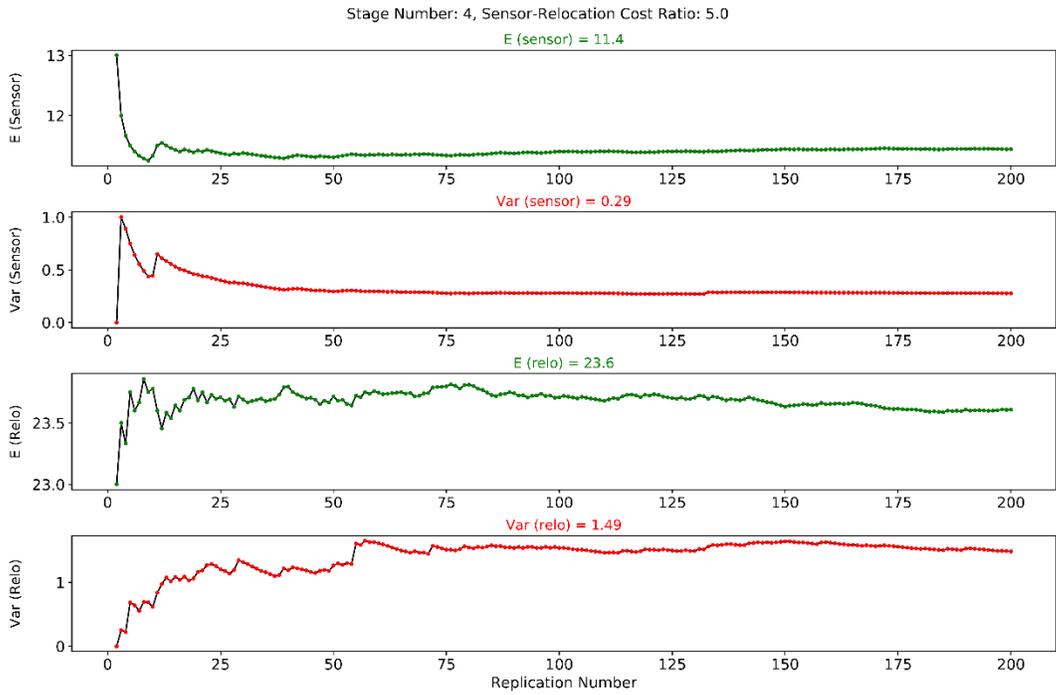
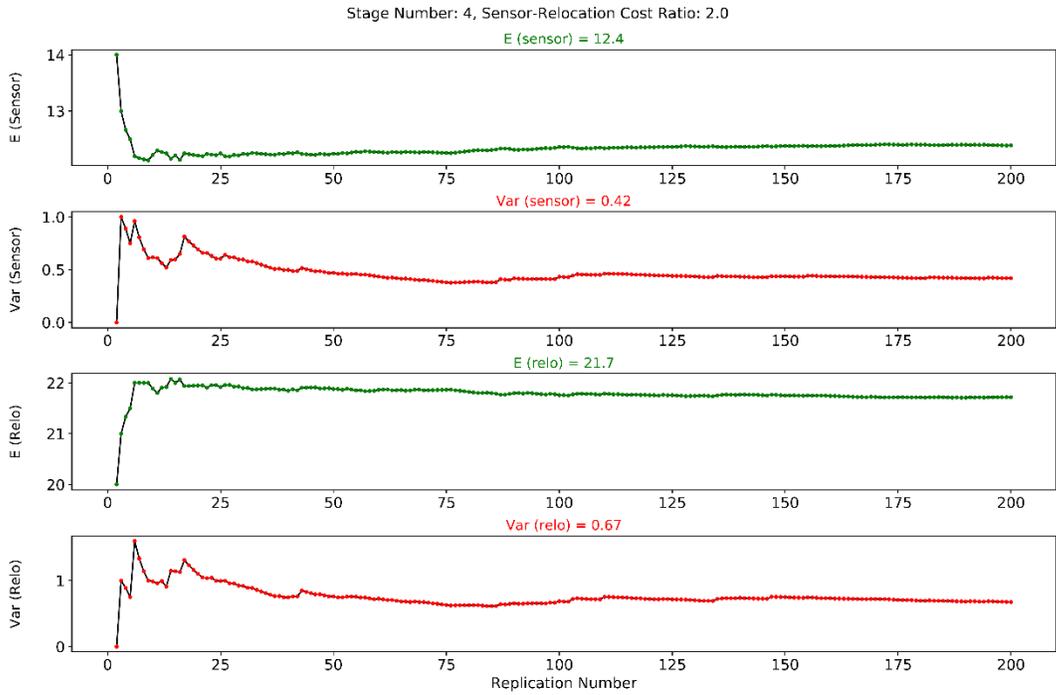


Figure 58: Solutions for 3-Stage Optimization by MR-SAA-N1 Algorithm: (a) Sensor/Relocation Cost Ratio=1.0; (b) Sensor/Relocation Cost Ratio=2.0; (c) Sensor/Relocation Cost Ratio=5.0; and (d) Sensor/Relocation Cost Ratio=10.0.





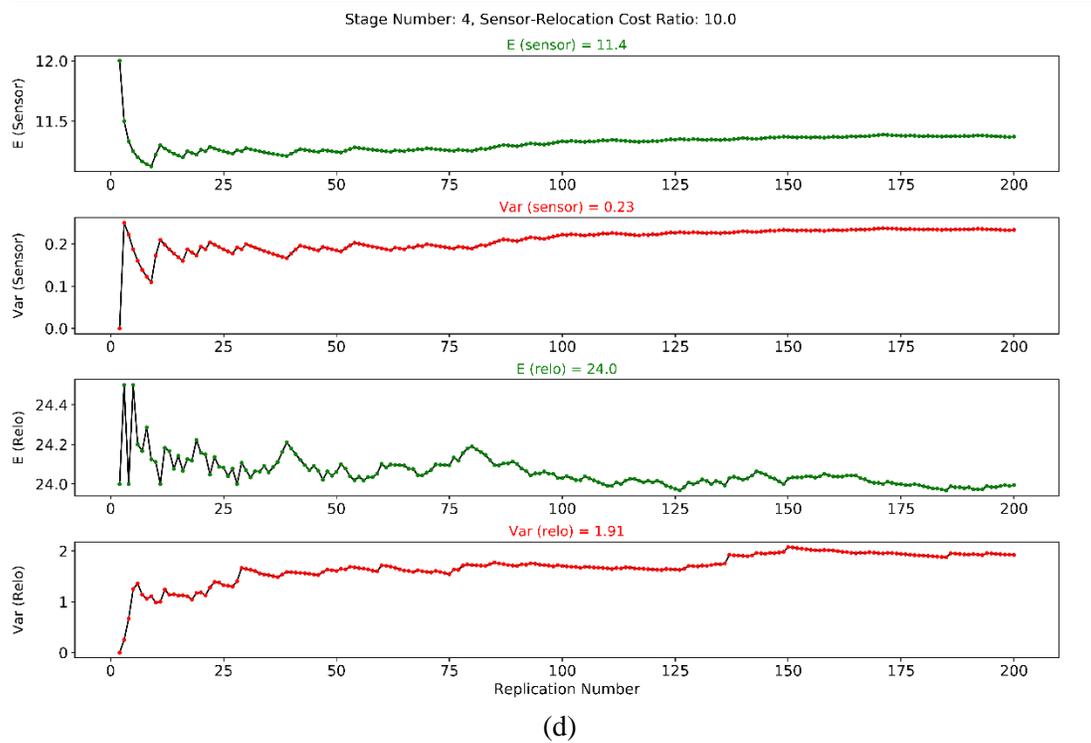


Figure 59: Solutions for 4-Stage Optimization by MR-SAA-N1 Algorithm: (a) Sensor/Relocation Cost Ratio=1.0; (b) Sensor/Relocation Cost Ratio=2.0; (c) Sensor/Relocation Cost Ratio=5.0; and (d) Sensor/Relocation Cost Ratio=10.0.

The overall summary of the final solutions for problems under various model parameters configurations is provided by Table 16. In the table, the first and second column denote the stage number and sensor/relocation (S/R) cost ratio, respectively. The third column indicates the solution method for each solution, i.e., MR-SAA-N1 or MR-SAA-N2. For each specific optimization case, final decisions of the sensor number determined in the initial stage and the expected sensor relocation number are given in the eight and nine column, respectively. The value of the objective function calculated based on the sensor number and the total relocation frequency is given in the 10<sup>th</sup> column in the table. As is shown, the objective value is denoted in the way of  $x \cdot v$ , where  $v$  represents the unit cost of one relocation operation. Considering the linear format of the objective function and the S/R cost ratio explicitly indicated, the objective function can be generally obtained as the function of the unit cost  $v$ , and the number  $x$  is obtained by combining the sensor number, the relocation number and the S/R cost ratio.

Another numerical indicator is given in the 11<sup>th</sup> column, i.e., the sum of the sensor number and the expected total relocation number. An interesting fact can be seen with respect to this indicator that it does not differ a lot for all optimization cases. Specifically, this value is pretty stable around 35. This can be explained by the intrinsic goal of the optimization problem. The optimization problem can be viewed as a network coverage problem, in which the links and each pair of upstream-downstream links are required to be covered for at least one stage. Therefore, to meet this requirement, the initial number of sensors plus the relocation numbers within a given time horizon should be around a constant value given the network is the same for each optimization case. In other words, adding one sensor and inducing one relocation have the same benefit to meet the goal of covering the network. Specifically, adding one sensor can cover one uncovered network node, and inducing one relocation can achieve the same thing. Therefore, those sensor-plus-relocation indicators are the same for all optimization cases with the same network and the same coverage requirement.

Table 16: Solutions Comparisons of Problems under Different Model Parameters Configurations.

Stage #	S/R Cost Ratio	Sol Method	Sensor # (L)	Sensor # (U)	Relo # (L)	Relo # (U)	Optimal Sensor #	Optimal Relo #	Objective Value (\$)	(Sensor #) + (Relo #)	Sol Time (sec)
2	1	MR-SAA-N1	28	29	5	6	29	5	34*v	34	25
	1	MR-SAA-N2					28	6	34*v	34	474
	2	MR-SAA-N1	19	21	13	15	20	14	54*v	34	126
	2	MR-SAA-N2					20	14	54*v	34	761
	5	MR-SAA-N1	19	21	13	15	20	15	112*v	34	216
	5	MR-SAA-N2					20	14	114*v	34	666
	10	MR-SAA-N1	19	21	13	15	19	15	209*v	34	256
	10	MR-SAA-N2					20	14	213*v	34	604
3	1	MR-SAA-N1	28	30	4	6	29	5	34*v	34	1356
	1	MR-SAA-N2					29	5	34*v	34	1827
	2	MR-SAA-N1	14	16	18	20	15	19	49*v	34	7595
	2	MR-SAA-N2					15	19	49*v	34	14132
	5	MR-SAA-N1	14	15	20	21	14	20	92*v	34	8196
	5	MR-SAA-N2					15	20	93*v	35	24026
	10	MR-SAA-N1	14	15	20	21	14	20	163*v	35	8927
	10	MR-SAA-N2					15	20	166*v	35	24512
4	1	MR-SAA-N1	29	30	4	5	29	5	34*v	34	1287
	1	MR-SAA-N2					29	5	34*v	34	5266
	2	MR-SAA-N1	11	14	20	23	12	22	47*v	34	22484
	2	MR-SAA-N2					13	21	48*v	34	58832
	5	MR-SAA-N1	11	12	23	26	11	24	81*v	35	82913
	5	MR-SAA-N2					12	24	85*v	36	60087
	10	MR-SAA-N1	11	12	23	26	11	24	138*v	35	60012
	10	MR-SAA-N2					12	24	145*v	36	60088 <sup>1</sup>

The solution time by the MR-SAA approach for each optimization case is given in the last column of the table. As indicated, even though the final decisions by MR-SAA-N1 and MR-SAA-N2 are very similar, the solution time by MR-SAA-N2 is significantly larger than that by

<sup>1</sup>  $v$  denotes the unit cost of one relocation operation.

MR-SAA-N1. This is because the MR-SAA-N2 incorporates a large number of non-anticipativity constraints in the optimization process. Moreover, another factor that dramatically affects the solution time is the stage number. Specifically, for the proposed stochastic optimization problem with more stages, the solution time tends to be dramatically larger.

To demonstrate the impact of the S/R cost ratio on the optimal sensor-relocation decisions, we graphically plot the sensor-relocation decisions with respect to different S/R cost ratios under a specific planning horizon. Figure 54, Figure 55, and Figure 56 graphically illustrate such cost-decision relationships for 2-stage, 3-stage, and 4-stage optimization cases, respectively. Based on the trends plots, there are two major findings. First, as the sensor-relocation cost ratio increase (i.e., purchasing one sensor is more expensive than conducting one relocation), the optimal decision suggests conducting more relocations instead of purchasing additional sensors. However, when the cost ratio is greater than a threshold (i.e., 5.0), the final decisions cannot be affected too much.

The second finding is that the planning stage number did affect the final decisions of the sensor number and expected total relocation number. As is demonstrated in the three figures with different stage numbers, if the stage number is large, the optimal decision suggests purchasing fewer sensors but conducting more relocation operations to meet the network coverage requirements. If the stage number is limited, such as the 2-stage optimization case with S/R cost ratio equal to 10.0, the least number of sensors required to finish the data collection and validation duties is 20, while these numbers are 15 and 12 for the 3-stage and 4-stage optimization cases. These results are in agreement with a practical sense that when the data validation work is time sensitive, one should apply more sensors for the validation process. Instead, one may consider spreading the relocation operations to more stages to achieve the goal with a few sensors.

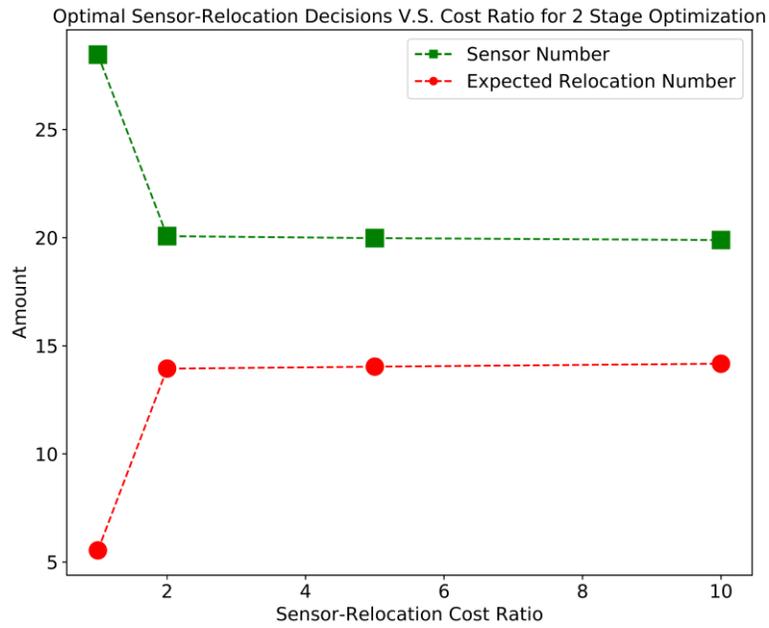


Figure 60: Optimal Sensor & Relocation Numbers with respect to Different S/R Cost Ratios for 2-Stage Optimization.

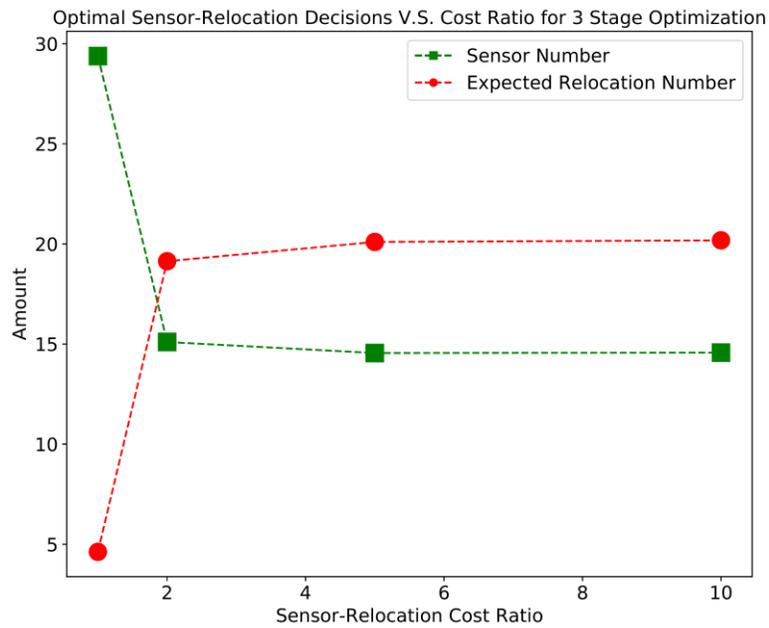


Figure 61: Optimal Sensor & Relocation Numbers with respect to Different S/R Cost Ratios for 3-Stage Optimization.

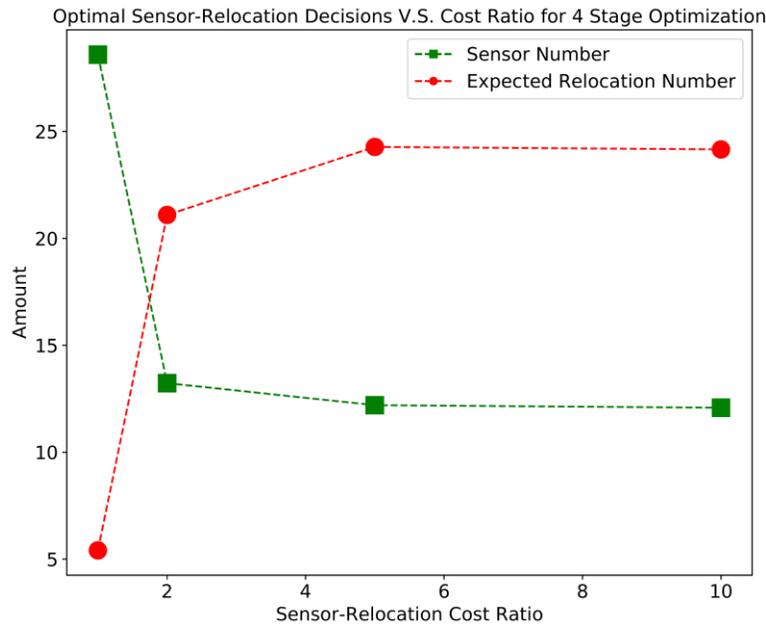


Figure 62: Optimal Sensor & Relocation Numbers with respect to Different S/R Cost Ratios for 4-Stage Optimization.

Figure 63 gives the comparisons of total sensor cost and total relocation cost from different optimization cases with the unit relocation cost set as 100 dollars. The comparison is given by three categories which are classified by the stage number, i.e., 2-stage, 3-stage, and 4-stage. In each category, the costs of total sensors and relocations are indicated and compared with respect to various S/R cost ratio. The black dashed lines indicate the total cost based on the total sensor number and the expected relocation number. For example, for 2-stage optimization with S/R cost ratio equal to 5.0, the cost for purchasing sensors is \$9,990, the cost for conducting relocations is \$1,404, and the total cost is \$11,394.

As is displayed, the percentage of relocation cost increases as the unit sensor cost increases. However, if more stages are assigned in the entire deployment process, the total cost can be decreased. For instance, when the unit sensor cost is ten times the unit relocation cost, using two stages to conduct the data validation process requires \$21,307 while using four stages to achieve the same goal only requires \$14,496. This affirms the fact that the data collection and validation is a time-sensitive process. If the time is not strictly limited, extending the total time

horizon can dramatically decrease the total cost, especially when the sensors are more expensive than the relocation operation.

Moreover, as the cost trend patterns shown by the three black dashed lines indicate, increasing the time horizon for data validation can release the total cost requirement as the unit sensor cost increases to a large extent. In other words, the total cost slope with respect to the S/R cost ratio decreases as the stage number increases (i.e., this indicates a more flexible time horizon). This cost-benefit effect can also be seen in Figure 64. When the S/R cost ratio is around 1.0 (i.e., the sensor is not very expensive compared to relocation cost), assigning more stages does not decrease the total cost to achieve the data collection and validation goals (i.e., the red and green dashed lines). But when the S/R cost ratio is high, assigning more stages can significantly decrease the total cost. However, the trade-off is whether one is willing to wait more time to have the data collection goal achieved.



Figure 63: Costs of Sensor Purchase and Relocation Operations of Different Optimization Cases.

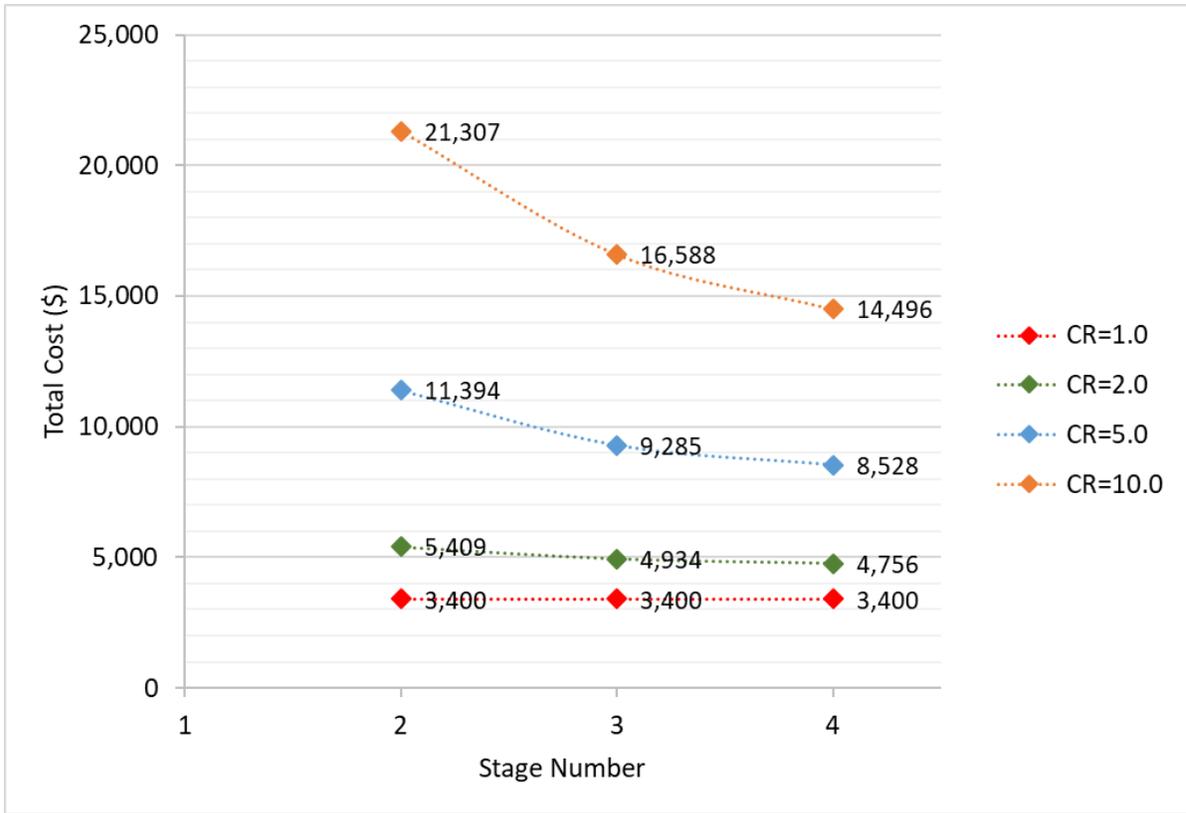


Figure 64: Total Cost w.r.t. Optimal Decisions of Different Cases.

### 5.6 Summary and Conclusions

In this chapter, an optimization model for deploying traffic sensors to collect and validate traffic data was proposed. The proposed optimization model considers the existence of GPS probe data within the highway network. The existence of GPS probe data provides an alternative data source for measuring and reporting the traffic state on a particular highway segment. Therefore, collecting and investigating the temporal-spatial traffic states within a network can not only be achieved through the roadside traffic detectors but can also be assisted with the reliable GPS data reports. However, the reliability of GPS data as an alternative on each highway segment is considered unknown at the beginning of the data collection and validation process. Because of this uncertainty, the proposed optimization problem was developed and formulated as a stochastic optimization problem.

Three goals are set in the optimization problem. They are, (1) using roadside traffic sensors to reveal the GPS data source reliability on each segment, (2) collecting and validating traffic state evolution patterns for each in-network segment, and (3) collecting and validating traffic state correlation patterns between different highway segments of interests. Once reliability of the GPS data source on a particular segment is revealed (i.e., the first goal is met at a location), it can serve as an additional data source to assist in achieving the goals defined in (1) and (2). In summary, the optimization problem itself can be viewed as a network coverage optimization problem with different coverage goals.

The proposed optimization model provides flexibility to relocate and reuse sensors once parts of the network have been validated considering the cost-benefit issue of purchasing traffic sensors. The optimization problem is formulated as a multi-stage optimization model. The initial-stage decision is about how many sensors should be purchased, and the subsequent-stage decisions are about how to relocate and reuse the sensors from the previous stages. Overall, the entire issue becomes a dynamic network coverage problem with a given time horizon that is discretized into several stages. The objective function is set as minimizing the total cost of sensors purchase and relocations. In other words, the proposed model aims to search for an optimal strategy to determine the minimum expense induced by purchasing sensors and conducting relocation operations to meet all of the imposed data collection and validation goals.

By analyzing the stochastic property of the multi-stage optimization model, a Monte Carlo simulation-based solution approach was designed to evaluate and search for the optimal solution specifically. The proposed optimization model was applied to a real-world highway network in Washington D.C.-Baltimore region for testing purposes. Optimal sensor allocation strategies were obtained and compared for different model parameters. Specifically, both the impacts of time horizon length and sensor-relocation cost ratio on the final decisions were investigated and analyzed. The major findings based on our numerical experiments are summarized below:

- In terms of the formulated optimization model, the proposed Monte Carlo simulation-based solution approach can assist in finding optimal solutions robustly;
- Adding more stages in the optimization model dramatically increases the solution time since it introduces more non-anticipativity constraints for the scenario-aggregation based formulation;
- The relationship between the sensor cost and relocation cost has a significant impact on the final decisions when the stage number is small;
- With the same data collection and validation goals, increasing the stage number (i.e., expanding the time horizon) which means increasing the time flexibility, can significantly decrease the total cost spent for the entire process when the sensors are more expensive than the relocation operation.
- The proposed model can be applied to any highway network when one wants to plan and evaluate the cost of traffic data collection and validation, such as the project reported in Haghani, Hamed, and Sadabadi (2009).

## Chapter 6 Summary and Future Research

Highway system plays a fundamental role in people's daily commute. Understanding real-time traffic state and congestion level helps travelers to plan their routes better and schedule their departure times. This dissertation deals with three major issues about the highway system. They are, (1) highway system state evolution uncertainty modeling and estimation, (2) real-time traffic state prediction-based sensor network optimization and dynamic planning, and (3) data validation based multistage traffic sensor network planning.

### 6.1 *Contributions*

This dissertation contributes three advanced models to the field of highway traffic information system. Chapter 3 presents a general probabilistic model to estimate temporal-spatial system state evolution uncertainty. The proposed model can be specified and applied to any stochastic system to evaluate the surveillance (i.e., measurement) effectiveness for prediction of a particular system state. The concept of conditional entropy is adopted to model the system state evolution uncertainty. The advantage of the proposed model is that the prediction uncertainty evaluation process does not require one to specify the prediction model structure. In other words, it models the system state evolution pattern only based on the measurement space and target state space. Further, we applied the model to a real-world highway network to evaluate the temporal-spatial travel time prediction uncertainty for 88 different corridor segments. In the real-world case study, a three-component characteristic vector representing highway travel time patterns is carefully designed and applied to the travel time uncertainty estimation process.

In Chapter 4, we presented both a static and a dynamic optimization model to plan traffic sensor placement strategies to improve real-time network surveillance. The developed optimization model has two highlights. First, the possibility of spatial information-based network state prediction is considered in the planning stage. Conventionally, people only focus on the

surveillance benefit at the location where sensors are placed while ignoring the surveillance benefit improvement resulting from spatial information-based inference. Second, the proposed dynamic network optimization model provides one with the possibility to come up with optimal sensor relocation strategies. Specifically, when traffic demand and travel time uncertainty are heterogeneously distributed in a highway network in a given time period, considering to relocate sensors can fully make use of the surveillance resources and enhance the network surveillance.

Lastly, we presented a sensor placement optimization model with the goal of efficiently collecting and validating traffic information. The purpose of this optimization model is to determine the stage-wise traffic sensor placement and relocation operations to collect and validate temporal-spatial traffic state patterns cost-effectively. The basic concept of the optimization model is a multi-stage network link coverage model. Moreover, the existence of independent data sources providing the same type of traffic state information is considered in the planning model. Data reliability is considered as a stochastic variable and can only be revealed after data validation process. A Monte Carlo simulation-based scenario decomposition algorithm is designed to solve the optimization model with endogenous uncertainty.

## 6.2 *Key Findings*

All models developed in this dissertation were implemented and applied to real-world case studies. Based on the numerical experiments, we summarize the following key findings:

- a. Travel time prediction uncertainty analysis on 88 highway corridors indicates real-time travel time surveillance does improve the travel time predictability compared to the historical inference;
- b. For the transportation network in study region (i.e., Washington D.C. – Baltimore area), implementing real-time travel time surveillance system can bring significant benefits in saving travelers' travel time planning error, specifically during the PM peak period;

- c. Travel time pattern uncertainty reduction, measured in entropy by the proposed model, does indicate the prediction error reduction level by a specific prediction model. Specifically, the prediction error reduction level has an increasing polynomial relationship with the uncertainty reduction level. Therefore, the uncertainty model can serve as a new measurement index evaluating the state predictability of any stochastic system without specifying the prediction model;
- d. To provide real-time travel time information in a given highway network, considering sensor relocation operations can further enhance total surveillance effectiveness. The cost-benefit ratio of conducting sensor relocations is dependent on the network size, spatial-temporal traffic demand distributions and traffic state predictability distributions in the network;
- e. For the Washington D.C.-Baltimore commute network, conducting relocation operations is more cost-benefit effective when the sensor fleet size is around 20 to 24, and improving the sensor fleet size is more cost-benefit effective when the fleet size is under 16;
- f. Whether choosing to add more sensors or conducting relocations of existing sensors is scenario dependent. Problem parameters such as sensor-relocation cost ratio, network size, and traffic demand distribution can affect the optimal sensor placement strategies. The proposed dynamic sensor network optimization model provides one with the opportunity to fully explore marginal surveillance benefit improvement with respect to adding more sensors and conducting relocations;
- g. For sensor placement with the purpose of collecting and validating temporal-spatial traffic information, sensor-relocation cost ratio and preferred time horizon (i.e., stage number) are two key parameters to affect the optimal placement decisions. Specifically, for scenarios where sensor-relocation cost ratio is high, or time constraint is not strict,

one should consider purchasing fewer sensors while conducting more relocations to meet the data validation requirements.

### 6.3 *Practical Concerns and Application*

Real-time data feeds describing traffic congestion level from various aspects play key roles in the transportation system. Different approaches are invented, and various data agencies are established to collect real-time traffic information such as travel time and traffic volume data. This research gave a comprehensive study on investigating the value of real-time travel time information on a complex highway system consisting of multiple corridors and developed two optimization models to practically determine the optimal traffic sensor deployment strategies to collect and validate such real-time data feeds in a real-world system. Most importantly, the value of the real-time surveillance at specific locations should be scientifically recognized to optimally plan a traffic sensor network.

The proposed models in this research consider the phrase “sensor” as a general sensing technique to collect traffic state information dynamically. The type of traffic sensor is not strictly limited if one considers applying the proposed model to plan a sensor-based surveillance network in a given highway system. However, before applying the optimization model, an important issue must be considered, i.e., accuracy and performance of the selected sensor. As is introduced in the literature review part of this dissertation, there are different types of sensors adapted in current transportation systems. Each type of sensor has its advantages and disadvantages.

Performance of a specific type of sensor has a significant impact on the surveillance effectiveness. Moreover, the joint data collection mechanism of an integrated sensing system also affects the data collection effectiveness. For example, the space between two adjacent identification and re-identification sensors (e.g., Bluetooth sensors) should be carefully selected based on the performance of the two sensors. Shorter distance might induce larger measurement errors, while overlong distance might induce inadequate probe samples and larger data latency.

Therefore, the sensor deployment configurations such as sensor space and detection radius should be carefully recognized and taken into account when evaluating the real-time surveillance effectiveness (e.g., using the evaluations models of Chapter 3) before the optimization process.

To give an overall view of the application of the developed models of this dissertation, we provided a flowchart shown in Figure 65. As is illustrated at the beginning of the flowchart, one should comprehensively understand the performance and associated configurations of the sensors he/she selects for building the surveillance system. Issues and concerns regarding this part have been mentioned in the last paragraph. Otherwise, it is inaccurate to specify and evaluate the surveillance benefits by inducing a particular sensing system. Detailed efforts regarding this part are out of the scope of this dissertation and should not be ignored in any practical traffic surveillance development projects.

Another interesting aspect of developing a traffic sensing system is about how to utilize the collected real-time information to improve the mobility of the transportation system. This issue talks more about the application of the proposed traffic surveillance system and is indicated at the end of the flowchart given in Figure 65. The purpose of deploying a real-time traffic state surveillance system proposed in this research is to provide timely and accurate traffic information to both travelers and transportation system operators. The underlying reason for traffic congestion frequently occurring in urban areas is the traffic demand that is consistently increasing. On one side, accurate knowledge of travel time or speed information across the entire highway network can help travelers to better plan their upcoming trips through adjusting departure times and routes. On the other hand, by obtaining network-level real-time traffic state information, transportation operators to further improve the mobility of the highway system, such as bottleneck identification, dynamic patrolling fleet optimization, and broadcasting smart traveling guidance. Consequently, traffic congestion can be mitigated.

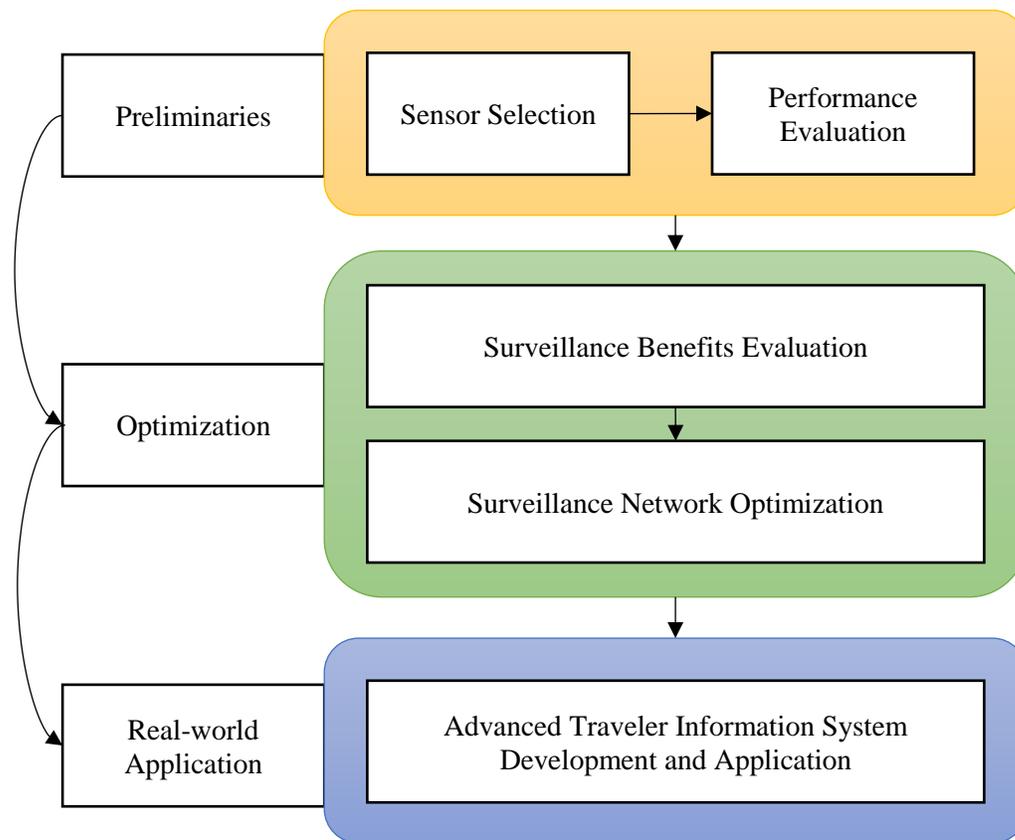


Figure 65: Flow Chart Demonstrating the Application of the Developed Models in this Dissertation.

#### 6.4 Future Research

Promising research in the future can follow the following directions:

- a. Underlying reasons why travel time predictability differs at different locations are not investigated and answered by this research. It is an interesting research direction to figure out the factors affecting travel time prediction uncertainty empirically. For example, how traffic incident occurrence rate or roadway geometric characteristics affect the travel time evolution patterns at a particular corridor?
- b. As is indicated by the numerical experiments in Chapter 3, the RF model cannot always improve the real-time travel time prediction performance for all locations in the network. However, the uncertainty evaluation model indicates that introducing real-time surveillance can enhance the prediction performance for each location compared to the

predictions based on historical inference. Investigating the reason why RF model fails at some of the locations is an interesting task.

- c. In future works, it is useful to develop sensor network optimization models with different surveillance objectives, such as minimizing traffic incident detection time and maximizing time-dependent O-D volume inference accuracy. Further, evaluating the marginal surveillance benefit induced by relocation operation with such objectives can provide useful guidance for real-world operations.

## References

- Abbott-Jard, Michael, Harpal Shah, and Ashish Bhaskar. 2013. "Empirical Evaluation of Bluetooth and Wifi Scanning for Road Transport." In *Australasian Transport Research Forum (ATRF), 36th, 2013, Brisbane, Queensland, Australia*, 14. [http://atrf.info/papers/2013/2013\\_abbott-jard\\_shah\\_bhaskar.pdf](http://atrf.info/papers/2013/2013_abbott-jard_shah_bhaskar.pdf).
- Abedi, Naeim, Ashish Bhaskar, and Edward Chung. 2013. "Bluetooth and Wi-Fi MAC Address Based Crowd Data Collection and Monitoring: Benefits, Challenges and Enhancement." <http://eprints.qut.edu.au/71808/>.
- Abolghasemi, Vahid, and Alireza Ahmadyfard. 2009. "An Edge-Based Color-Aided Method for License Plate Detection." *Image and Vision Computing* 27 (8): 1134–1142.
- Ahmad, I., and Pi-Erh Lin. 1976. "A Nonparametric Estimation of the Entropy for Absolutely Continuous Distributions (Corresp.)." *IEEE Transactions on Information Theory* 22 (3): 372–75. <https://doi.org/10.1109/TIT.1976.1055550>.
- Ahmed, Samir A., T. M. Hussain, and Tarek N. Saadawi. 1994. "Active and Passive Infrared Sensors for Vehicular Traffic Control." In *Vehicular Technology Conference, 1994 IEEE 44th*, 1393–1397. IEEE. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=345323](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=345323).
- Almorox-Gonzalez, Pablo, Jose-Tomas Gonzalez-Partida, Mateo Burgos-Garcia, Cristina de la Morena-Alvarez-Palencia, Lara Arche-Andradas, and B. Pablo Dorta-Naranjo. 2007. "Portable High Resolution LFM-CW Radar Sensor in Millimeter-Wave Band." In *Sensor Technologies and Applications, 2007. SensorComm 2007. International Conference On*, 5–9. IEEE. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=4394888](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4394888).
- Anagnostopoulos, Christos Nikolaos E., Ioannis E. Anagnostopoulos, Vassilis Loumos, and Eleftherios Kayafas. 2006. "A License Plate-Recognition Algorithm for Intelligent Transportation System Applications." *IEEE Transactions on Intelligent Transportation Systems* 7 (3): 377–392.
- Asudegi, Mona, and Ali Haghani. 2013. "Optimal Number and Location of Node-Based Sensors for Collection of Travel Time Data in Networks." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2338: 35–43.
- "Automatic Number Plate Recognition." 2016. *Wikipedia, the Free Encyclopedia*. [https://en.wikipedia.org/w/index.php?title=Automatic\\_number\\_plate\\_recognition&oldid=736496401](https://en.wikipedia.org/w/index.php?title=Automatic_number_plate_recognition&oldid=736496401).
- Balid, Walid, Hasan Tafish, and Hazem H. Refai. 2015. "Development of Portable Wireless Sensor Network System for Real-Time Traffic Surveillance." In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 1630–1637. IEEE. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=7313357](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7313357).
- Ban, Xuegang Jeff, Ryan Herring, J. D. Margulici, and Alexandre M. Bayen. 2009. "Optimal Sensor Placement for Freeway Travel Time Estimation." In *Transportation and Traffic Theory 2009: Golden Jubilee*, 697–721. Springer. [http://link.springer.com/chapter/10.1007/978-1-4419-0820-9\\_34](http://link.springer.com/chapter/10.1007/978-1-4419-0820-9_34).
- Bartin, Bekir, Kaan Ozbay, and Cem Iyigun. 2007. "Clustering-Based Methodology for Determining Optimal Roadway Configuration of Detectors for Travel Time Estimation." *Transportation Research Record: Journal of the Transportation Research Board*. <http://trrjournalonline.trb.org/doi/abs/10.3141/2000-12>.
- Beirlant, Jan, Edward J. Dudewicz, László Györfi, and Edward C. Van der Meulen. 1997. "Nonparametric Entropy Estimation: An Overview." *International Journal of Mathematical and Statistical Sciences* 6 (1): 17–39.

- Ben-Akiva, Moshe, Michel Bierlaire, Haris Koutsopoulos, and Rabi Mishalani. 1998. "DynaMIT: A Simulation-Based System for Traffic Prediction." In *DACCORD Short Term Forecasting Workshop*, 1–12.
- Bertini, Robert, and Sutti Tantiyanugulchai. 2004. "Transit Buses as Traffic Probes: Use of Geolocation Data for Empirical Evaluation." *Transportation Research Record: Journal of the Transportation Research Board*, no. 1870: 35–45.
- Bhaskar, Ashish, and Edward Chung. 2013. "Fundamental Understanding on the Use of Bluetooth Scanner as a Complementary Transport Data." *Transportation Research Part C: Emerging Technologies* 37: 42–72.
- Bhaskar, Ashish, Edward Chung, and André-Gilles Dumont. 2011. "Fusing Loop Detector and Probe Vehicle Data to Estimate Travel Time Statistics on Signalized Urban Networks." *Computer-Aided Civil and Infrastructure Engineering* 26 (6): 433–450.
- Bianco, Lucio, Giuseppe Confessore, and Pierfrancesco Reverberi. 2001. "A Network Based Model for Traffic Sensor Location with Implications on O/D Matrix Estimates." *Transportation Science* 35 (1): 50–60.
- Brennan, Thomas M., Joseph M. Ernst, Christopher M. Day, Darcy M. Bullock, James V. Krogmeier, and Mary Martchouk. 2010. "Influence of Vertical Sensor Placement on Data Collection Efficiency from Bluetooth MAC Address Collection Devices." *Journal of Transportation Engineering* 136 (12): 1104–1109.
- Caruso, Michael J., and Lucky S. Withanawasam. 1999. "Vehicle Detection and Compass Applications Using AMR Magnetic Sensors." In *Sensors Expo Proceedings*, 477:39. <http://saba.kntu.ac.ir/EECD/ECOURSES/INSTRUMENTATION/PROJECTS/REPORTS/MAGNETIC%20POSITION%20SENSORS/Special%20application/rrefrences/VEHICLE.PDF>.
- Castillo, E., F. Jubete, R. E. Pruneda, and C. Solares. 2002. "Obtaining Simultaneous Solutions of Linear Subsystems of Inequalities and Duals." *Linear Algebra and Its Applications* 346 (1): 131–154.
- Castillo, Enrique, Inmaculada Gallego, José María Menéndez, and Ana Rivas. 2010. "Optimal Use of Plate-Scanning Resources for Route Flow Estimation in Traffic Networks." *IEEE Transactions on Intelligent Transportation Systems* 11 (2): 380–391.
- Castro-Neto, Manoel, Young-Seon Jeong, Myong-Kee Jeong, and Lee D. Han. 2009. "Online-SVR for Short-Term Traffic Flow Prediction under Typical and Atypical Traffic Conditions." *Expert Systems with Applications* 36 (3): 6164–6173.
- Chan, K. S., and William HK Lam. 2002. "Optimal Speed Detector Density for the Network with Travel Time Information." *Transportation Research Part A: Policy and Practice* 36 (3): 203–223.
- Chang, Shyang-Lih, Li-Shien Chen, Yun-Chung Chung, and Sei-Wan Chen. 2004. "Automatic License Plate Recognition." *IEEE Transactions on Intelligent Transportation Systems* 5 (1): 42–53.
- Chaudhuri, Piyali, Peter T. Martin, Aleksandar Z. Stevanovic, and Chongkai Zhu. 2010. "The Effects of Detector Spacing on Travel Time Prediction on Freeways." *World Academy of Science, Engineering and Technology* 42: 1–10.
- Chen, Anthony, Piya Chootinan, and Surachet Pravinvongvuth. 2004. "Multiobjective Model for Locating Automatic Vehicle Identification Readers." *Transportation Research Record: Journal of the Transportation Research Board*, no. 1886: 49–58.
- Chen, Mei, and Steven Chien. 2000. "Determining the Number of Probe Vehicles for Freeway Travel Time Estimation by Microscopic Simulation." *Transportation Research Record: Journal of the Transportation Research Board*, no. 1719: 61–68.
- Cheung, Sing, Sinem Coleri, Baris Dundar, Sumitra Ganesh, Chin-Woo Tan, and Pravin Varaiya. 2005. "Traffic Measurement and Vehicle Classification with Single Magnetic Sensor."

- Transportation Research Record: Journal of the Transportation Research Board*, no. 1917: 173–181.
- Chung, Il Ho. 2001. *An Optimim Sampling Framework for Estimating Trip Matrices from Day-to-Day Traffic Counts*.
- Connolly, Robert, Chris Stivers, and Licheng Sun. 2005. “Stock Market Uncertainty and the Stock-Bond Return Relation.” *Journal of Financial and Quantitative Analysis* 40 (1): 161–94. <https://doi.org/10.1017/S0022109000001782>.
- Danczyk, Adam, Xuan Di, and Henry X. Liu. 2016. “A Probabilistic Optimization Model for Allocating Freeway Sensors.” *Transportation Research Part C: Emerging Technologies* 67: 378–398.
- Danczyk, Adam, and Henry X. Liu. 2011. “A Mixed-Integer Linear Program for Optimizing Sensor Locations along Freeway Corridors.” *Transportation Research Part B: Methodological* 45 (1): 208–217.
- Dezani, Henrique, Furio Damiani, Norian Marranghello, Ulysses Viudes, and Ícaro Augusto Parra. 2012. “Mobile Application as a Tool for Urban Traffic Data Collection and Generation to Advanced Traveler Information Systems Using Wi-Fi Networks Available in Urban Centers.” In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, 288–292. IEEE. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6232135](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6232135).
- Du, Rong, Cailian Chen, Bo Yang, Ning Lu, Xinping Guan, and Xuemin Shen. 2015. “Effective Urban Traffic Monitoring by Vehicular Sensor Networks.” *IEEE Transactions on Vehicular Technology* 64 (1): 273–286.
- Du, Shan, Mahmoud Ibrahim, Mohamed Shehata, and Wael Badawy. 2013. “Automatic License Plate Recognition (ALPR): A State-of-the-Art Review.” *IEEE Transactions on Circuits and Systems for Video Technology* 23 (2): 311–325.
- Edara, Praveen K., Jianhua Guo, Brian Lee Smith, and Catherine C. McGhee. 2008. “Optimal Placement of Point Detectors on Virginia’s Freeways: Case Studies of Northern Virginia and Richmond.” <https://trid.trb.org/view.aspx?id=850807>.
- Ehlert, Anett, Michael GH Bell, and Sergio Grosso. 2006. “The Optimisation of Traffic Count Locations in Road Networks.” *Transportation Research Part B: Methodological* 40 (6): 460–479.
- Farokhi Sadabadi, Kaveh. 2014. “Vehicular Traffic Modelling, Data Assimilation, Estimation and Short Term Travel Time Prediction.” <http://drum.lib.umd.edu/handle/1903/16088>.
- Fei, Xiang, Chung-Cheng Lu, and Ke Liu. 2011. “A Bayesian Dynamic Linear Model Approach for Real-Time Short-Term Freeway Travel Time Prediction.” *Transportation Research Part C: Emerging Technologies* 19 (6): 1306–1318.
- Fei, Xiang, Hani S. Mahmassani, and Pamela Murray-Tuite. 2013. “Vehicular Network Sensor Placement Optimization under Uncertainty.” *Transportation Research Part C: Emerging Technologies* 29: 14–31.
- Foote, Robert S. 1974. “Automatic Vehicle Identification.” *Transportation Research News*, no. 56. <https://trid.trb.org/view.aspx?id=139981>.
- Friswell, Michael, and John E. Mottershead. 1995. *Finite Element Model Updating in Structural Dynamics*. Vol. 38. Springer Science & Business Media. [https://books.google.com/books?hl=en&lr=&id=KyMVKqrEEZIC&oi=fnd&pg=PR11&dq=finite+element+model+updating+in+structural+dynamics&ots=6Sd0Zdm\\_y&sig=51uBi7dKI9z4LjTKkpfPpNKLOW](https://books.google.com/books?hl=en&lr=&id=KyMVKqrEEZIC&oi=fnd&pg=PR11&dq=finite+element+model+updating+in+structural+dynamics&ots=6Sd0Zdm_y&sig=51uBi7dKI9z4LjTKkpfPpNKLOW).
- Gentili, M., and P. B. Mirchandani. 2012. “Locating Sensors on Traffic Networks: Models, Challenges and Research Opportunities.” *Transportation Research Part C: Emerging Technologies* 24: 227–255.
- Gilly, Divya, and Kumudha Raimond. 2013. “A Survey on License Plate Recognition Systems.” *International Journal of Computer Applications* 61 (6).

- <http://search.proquest.com/openview/3d2bb674b61ba1a2bf7058802697cd87/1?pq-origsite=gscholar>.
- Guo, Jing-Ming, and Yun-Fu Liu. 2008. "License Plate Localization and Character Segmentation with Feedback Self-Learning and Hybrid Binarization Techniques." *IEEE Transactions on Vehicular Technology* 57 (3): 1417–1424.
- Haghani, Ali, Masoud Hamed, and Kaveh Farokhi Sadabadi. 2009. *I-95 Corridor Coalition Vehicle Probe Project: Validation of INRIX Data*. I-95 Corridor Coalition.  
<http://www.i95coalition.org/wp-content/uploads/2015/02/I-95-CC-Estimation-of-Travel-Times-for-Multiple-TMC-Segments-FINAL2.pdf>.
- Haghani, Ali, Masoud Hamed, Kaveh Sadabadi, Stanley Young, and Philip Tarnoff. 2010. "Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2160: 60–68.
- Hainen, Alexander, Jason Wasson, Sarah Hubbard, Stephen Remias, Grant Farnsworth, and Darcy Bullock. 2011. "Estimating Route Choice and Travel Time Reliability with Field Observations of Bluetooth Probe Vehicles." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2256: 43–50.
- Haoui, Amine, Robert Kavalier, and Pravin Varaiya. 2008. "Wireless Magnetic Sensors for Traffic Surveillance." *Transportation Research Part C: Emerging Technologies* 16 (3): 294–306.
- Haseman, Ross, Jason Wasson, and Darcy Bullock. 2010. "Real-Time Measurement of Travel Time Delay in Work Zones and Evaluation Metrics Using Bluetooth Probe Tracking." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2169: 40–53.
- Hassett, John J. 1998. *Electronic Vehicle Toll Collection System and Method*. Google Patents.  
<https://www.google.com/patents/US5805082>.
- Herrera, Juan C., Daniel B. Work, Ryan Herring, Xuegang Jeff Ban, Quinn Jacobson, and Alexandre M. Bayen. 2010. "Evaluation of Traffic Data Obtained via GPS-Enabled Mobile Phones: The Mobile Century Field Experiment." *Transportation Research Part C: Emerging Technologies* 18 (4): 568–583.
- Herring, Ryan, Aude Hofleitner, Pieter Abbeel, and Alexandre Bayen. 2010. "Estimating Arterial Traffic Conditions Using Sparse Probe Data." In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference On*, 929–936. IEEE.  
[http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5624994](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5624994).
- Hinsbergen, CP IJ van, J. W. C. Van Lint, and H. J. Van Zuylen. 2009. "Bayesian Committee of Neural Networks to Predict Travel Times with Confidence Intervals." *Transportation Research Part C: Emerging Technologies* 17 (5): 498–509.
- Hofleitner, Aude, Ryan Herring, Pieter Abbeel, and Alexandre Bayen. 2012. "Learning the Dynamics of Arterial Traffic from Probe Data Using a Dynamic Bayesian Network." *IEEE Transactions on Intelligent Transportation Systems* 13 (4): 1679–1693.
- Hu, Ta-Yin, Chee-Chung Tong, Tsai-Yun Liao, and Wei-Ming Ho. 2012. "Simulation-Assignment-Based Travel Time Prediction Model for Traffic Corridors." *IEEE Transactions on Intelligent Transportation Systems* 13 (3): 1277–1286.
- Hung, Ming-Lung, and Hwong-wen Ma. 2009. "Quantifying System Uncertainty of Life Cycle Assessment Based on Monte Carlo Simulation." *The International Journal of Life Cycle Assessment* 14 (1): 19–27.
- Hunter, Timothy, Ryan Herring, Pieter Abbeel, and Alexandre Bayen. 2009. "Path and Travel Time Inference from GPS Probe Vehicle Data." *NIPS Analyzing Networks and Learning with Graphs* 12 (1).  
<http://bayen.eecs.berkeley.edu/sites/default/files/conferences/nips09.pdf>.

- INRIX. n.d. "Methodology: INRIX 2015 Traffic Scorecard." *INRIX* (blog). Accessed January 13, 2017. <http://inrix.com/scorecard/methodology-en/>.
- Jenelius, Erik, and Haris N. Koutsopoulos. 2013. "Travel Time Estimation for Urban Road Networks Using Low Frequency Probe Vehicle Data." *Transportation Research Part B: Methodological* 53: 64–81.
- . 2015. "Probe Vehicle Data Sampled by Time or Space: Consistent Travel Time Allocation and Estimation." *Transportation Research Part B: Methodological* 71: 120–137.
- Khosravi, Abbas, Ehsan Mazloumi, Saeid Nahavandi, Doug Creighton, and J. W. C. Van Lint. 2011. "Prediction Intervals to Account for Uncertainties in Travel Time Prediction." *IEEE Transactions on Intelligent Transportation Systems* 12 (2): 537–547.
- Kianfar, Jalil, and Praveen Edara. 2010. "Optimizing Freeway Traffic Sensor Locations by Clustering Global-Positioning-System-Derived Speed Patterns." *IEEE Transactions on Intelligent Transportation Systems* 11 (3): 738–747.
- Kim, Joonhyo, Byungkyu Brian Park, Joyoung Lee, and Jongsun Won. 2011. "Determining Optimal Sensor Locations in Freeway Using Genetic Algorithm-Based Optimization." *Engineering Applications of Artificial Intelligence* 24 (2): 318–324.
- Klein, L. A., M. K. Milton, and R. P. Gibson. 2006. "Traffic Detector Handbook: –Volume IUS Department of Transportation, Publication No." FHWAHRT-06-108.
- Kleywegt, Anton J., Alexander Shapiro, and Tito Homem-de-Mello. 2002. "The Sample Average Approximation Method for Stochastic Discrete Optimization." *SIAM Journal on Optimization* 12 (2): 479–502.
- Koerner, Ralph J. 1976. *Inductive Loop Vehicle Detector*. Google Patents. <https://www.google.com/patents/US3989932>.
- Komguem, Rodrigue Domga, Razvan Stanica, Maurice Tchuenta, and Fabrice Valois. 2014. "WARIM: Wireless Sensor Networks Architecture for a Reliable Intersection Monitoring." In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 1226–1231. IEEE. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6957855](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6957855).
- Kotzenmacher, Jerry, Erik D. Minge, and Bingwen Hao. 2005. "Evaluation of Portable Non-Intrusive Traffic Detection System." Minnesota Department of Transportation, Research Services Section.
- Kuhn, John Patrick, Binh C. Bui, and Gregory J. Pieper. 1998. *Acoustic Sensor System for Vehicle Detection and Multi-Lane Highway Monitoring*. Google Patents. <https://www.google.com/patents/US5798983>.
- Lad, Ms Ankita, and Mr Dhaval Patel. 2015. "A SURVEY ON LICENSE PLATE RECOGNITION SYSTEM." <http://www.academia.edu/download/37720245/IJARIE1135.pdf>.
- Leduc, Guillaume. 2008. "Road Traffic Data: Collection Methods and Applications." *Working Papers on Energy, Transport and Climate Change* 1 (55). <ftp://139.191.159.82/pub/EURdoc/EURdoc/JRC47967.TN.pdf>.
- Li, Jing-Quan, Kun Zhou, Steven Shladover, and Alexander Skabardonis. 2013. "Estimating Queue Length under Connected Vehicle Technology: Using Probe Vehicle, Loop Detector, and Fused Data." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2356: 17–22.
- Li, Xiaopeng, and Yanfeng Ouyang. 2011. "Reliable Sensor Deployment for Network Traffic Surveillance." *Transportation Research Part B: Methodological* 45 (1): 218–231.
- . 2012. "Reliable Traffic Sensor Deployment under Probabilistic Disruptions and Generalized Surveillance Effectiveness Measures." *Operations Research* 60 (5): 1183–1198.

- Lindveld, Charles D., Remmelt Thijs, Piet Bovy, and Nanne Van der Zijpp. 2000. "Evaluation of Online Travel Time Estimators and Predictors." *Transportation Research Record: Journal of the Transportation Research Board*, no. 1719: 45–53.
- Liu, Hao, Henk van Zuylen, Hans van Lint, and Maria Salomons. 2006. "Predicting Urban Arterial Travel Time with State-Space Neural Networks and Kalman Filters." *Transportation Research Record: Journal of the Transportation Research Board*, no. 1968: 99–108.
- Liu, Tai K. 1994. "Field Tests of Travel Time Survey Methodologies and Development of a Standardized Data Processing and Reporting System." In *National Traffic Data Acquisition Conference (NATDAC '94) Proceedings*, 2:383–426.
- Losilla, Fernando, Antonio-Javier Garcia-Sanchez, Felipe Garcia-Sanchez, Joan Garcia-Haro, and Zygmunt J. Haas. 2011. "A Comprehensive Approach to WSN-Based ITS Applications: A Survey." *Sensors* 11 (11): 10220–10265.
- Lu, Yang. 2013. "Study of Real-Time Traffic State Estimation and Short-Term Prediction of Signalized Arterial Network Considering Heterogeneous Information Sources." <http://drum.lib.umd.edu/handle/1903/14227>.
- Luber, Andreas, Marek Junghans, Sascha Bauer, and Jan Schulz. 2011. "On Measuring Traffic with Wi-Fi and Bluetooth." In *18th ITS World Congress*. <https://trid.trb.org/view.aspx?id=1214841>.
- Lv, Yisheng, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang. 2015. "Traffic Flow Prediction with Big Data: A Deep Learning Approach." *IEEE Transactions on Intelligent Transportation Systems* 16 (2): 865–873.
- Malinovskiy, Yegor, Un-Kun Lee, Yao-Jan Wu, and Yinhai Wang. 2011. "Investigation of Bluetooth-Based Travel Time Estimation Error on a Short Corridor." In *Transportation Research Board 90th Annual Meeting*. <https://trid.trb.org/view.aspx?id=1092820>.
- Malinovskiy, Yegor, and Yinhai Wang. 2012. "Pedestrian Travel Pattern Discovery Using Mobile Bluetooth Sensors." In *Transportation Research Board 91st Annual Meeting*. <https://trid.trb.org/view.aspx?id=1129741>.
- Malinovskiy, Yegor, Yao-Jan Wu, Yinhai Wang, and Un Kun Lee. 2010. "Field Experiments on Bluetooth-Based Travel Time Data Collection." In *Transportation Research Board 89th Annual Meeting*. <https://trid.trb.org/view.aspx?id=910907>.
- Martchouk, Maria, Fred Mannering, and Darcy Bullock. 2010. "Analysis of Freeway Travel Time Variability Using Bluetooth Detection." *Journal of Transportation Engineering* 137 (10): 697–704.
- Matsuo, T., Y. Kaneko, and M. Matano. 1999. "Introduction of Intelligent Vehicle Detection Sensors." In *Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEJ/JSAI International Conference On*, 709–713. IEEE. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=821148](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=821148).
- Mei, Zhenyu, Dianhai Wang, and Jun Chen. 2012. "Investigation with Bluetooth Sensors of Bicycle Travel Time Estimation on a Short Corridor." *International Journal of Distributed Sensor Networks* 2012. <http://www.hindawi.com/journals/ijdsn/2012/303521/abs/>.
- Michalopoulos, Panos G. 1991. "Vehicle Detection Video through Image Processing: The Autoscope System." *IEEE Transactions on Vehicular Technology* 40 (1): 21–29.
- Mimbela, Luz Elena Y., and Lawrence A. Klein. 2000. "Summary of Vehicle Detection and Surveillance Technologies Used in Intelligent Transportation Systems." <https://trid.trb.org/view.aspx?id=681316>.
- Mirchandani, Pitu B., M. Gentili, and Yang He. 2009. "Location of Vehicle Identification Sensors to Monitor Travel-Time Performance." *IET Intelligent Transport Systems* 3 (3): 289–303.
- Nelson, Edward, and E. Nelson. 1985. *Quantum Fluctuations*. Princeton University Press Princeton. <http://press.princeton.edu/titles/2357.html>.

- Ozbay, Kaan, Bekir Bartin, and Steven Chien. 2004. "South Jersey Real-Time Motorist Information System: Technology and Practice." *Transportation Research Record: Journal of the Transportation Research Board*, no. 1886: 68–75.
- Papadimitriou, Costas, James L. Beck, and Siu-Kui Au. 2000. "Entropy-Based Optimal Sensor Location for Structural Model Updating." *Journal of Vibration and Control* 6 (5): 781–800.
- Park, Dongjoo, Laurence R. Rilett, Byron J. Gajewski, Clifford H. Spiegelman, and Changho Choi. 2009. "Identifying Optimal Data Aggregation Interval Sizes for Link and Corridor Travel Time Estimation and Forecasting." *Transportation* 36 (1): 77–95.
- Park, Hyoshin, and Ali Haghani. 2015. "Optimal Number and Location of Bluetooth Sensors Considering Stochastic Travel Time Prediction." *Transportation Research Part C: Emerging Technologies* 55: 203–216.
- Plesník, Ján. 1999. "Constrained Weighted Matchings and Edge Coverings in Graphs." *Discrete Applied Mathematics* 92 (2): 229–241.
- Pu, Wenjing, Jie Lin, and Liang Long. 2009. "Real-Time Estimation of Urban Street Segment Travel Time Using Buses as Speed Probes." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2129: 81–89.
- Rice, John, and Erik Van Zwet. 2004. "A Simple and Effective Method for Predicting Travel Times on Freeways." *IEEE Transactions on Intelligent Transportation Systems* 5 (3): 200–207.
- Robert-Nicoud, Y., B. Raphael, and I. F. C. Smith. 2005. "Configuration of Measurement Systems Using Shannon's Entropy Function." *Computers & Structures* 83 (8): 599–612.
- Rrnyi, ALFRPED. 1961. "On Measures of Entropy and Information." In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1:547–561.  
[http://biocomparison.ucoz.ru/\\_ld/0/37\\_SjS.pdf](http://biocomparison.ucoz.ru/_ld/0/37_SjS.pdf).
- Schauer, Lorenz, Martin Werner, and Philipp Marcus. 2014. "Estimating Crowd Densities and Pedestrian Flows Using Wi-Fi and Bluetooth." In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 171–177. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). <http://dl.acm.org/citation.cfm?id=2693006>.
- Schrank, David, Bill Eisele, Tim Lomax, and Jim Bak. 2015. "Urban Mobility Scorecard." *College Station: Texas A&M Transportation Institute and INRIX*.
- Sen, Rijurekha, Pankaj Siriah, and Bhaskaran Raman. 2011. "Roadsoundsense: Acoustic Sensing Based Road Congestion Monitoring in Developing Regions." In *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2011 8th Annual IEEE Communications Society Conference On*, 125–133. IEEE.  
[http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5984883](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5984883).
- Sergent, Edward W. 1981. *Traffic Radar Unit*. Google Patents.  
<https://www.google.com/patents/US4293859>.
- Shannon, Claude E., and Warren Weaver. 1949. *The Mathematical Theory of Communication*. University of Illinois press.  
<https://books.google.com/books?hl=en&lr=&id=IZ77BwAAQBAJ&oi=fnd&pg=PP1&dq=The+mathematical+theory+of+communication,+1949&ots=hkLxdUsN5C&sig=eQmDGSNToiCPKhzoIFcD-bB1VIs>.
- Sherali, Hanif D., Jitamitra Desai, and Hesham Rakha. 2006. "A Discrete Optimization Approach for Locating Automatic Vehicle Identification Readers for the Provision of Roadway Travel Times." *Transportation Research Part B: Methodological* 40 (10): 857–871.
- Shiller, Robert J. 1999. "Chapter 20 Human Behavior and the Efficiency of the Financial System." In , edited by BT - Handbook of Macroeconomics, 1, Part C:1305–40. Elsevier.  
<http://www.sciencedirect.com/science/article/pii/S1574004899100338>.

- Smith, Brian L., and Michael J. Demetsky. 1994. "Short-Term Traffic Flow Prediction Models-a Comparison of Neural Network and Nonparametric Regression Approaches." In *Systems, Man, and Cybernetics, 1994. Humans, Information and Technology., 1994 IEEE International Conference On*, 2:1706–1709. IEEE.
- Solak, Senay, John-Paul B. Clarke, Ellis L. Johnson, and Earl R. Barnes. 2010. "Optimization of R&D Project Portfolios under Endogenous Uncertainty." *European Journal of Operational Research* 207 (1): 420–433.
- Soriguera, Francesc, Leif Thorson, and Francesc Robusté. 2007. "Travel Time Measurement Using Toll Infrastructure." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2027: 99–107.
- Taghvayeean, Saber, and Rajesh Rajamani. 2014. "Portable Roadside Sensors for Vehicle Counting, Classification, and Speed Measurement." *IEEE Transactions on Intelligent Transportation Systems* 15 (1): 73–83.
- Tanaka, Yoshimi. 1992. "Travel-Time Data Provision System Using Vehicle License Number Recognition Devices." In *Intelligent Vehicles '92 Symposium., Proceedings of The*, 353–358. IEEE. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=252285](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=252285).
- Teodorovic, Dusan, Michel Van Aerde, Fulin Zhu, and Francois Dion. 2002. "Genetic Algorithms Approach to the Problem of the Automated Vehicle Identification Equipment Locations." *Journal of Advanced Transportation* 36 (1): 1–21.
- Toppen, Alan, and Karl Wunderlich. 2003. *Travel Time Data Collection for Measurement of Advanced Traveler Information Systems Accuracy*. Mitretek Systems. [http://ntl.bts.gov/lib/jpodocs/repts\\_te/13867\\_files/13867.pdf](http://ntl.bts.gov/lib/jpodocs/repts_te/13867_files/13867.pdf).
- Turner, Shawn M., William L. Eisele, Robert J. Benz, and Douglas J. Holdener. 1998. "Travel Time Data Collection Handbook." <https://trid.trb.org/view.aspx?id=497690>.
- Uno, Nobuhiro, Fumitaka Kurauchi, Hiroshi Tamura, and Yasunori Iida. 2009. "Using Bus Probe Data for Analysis of Travel Time Variability." *Journal of Intelligent Transportation Systems* 13 (1): 2–15.
- Vanajakshi, L., S. C. Subramanian, and R. Sivanandan. 2009. "Travel Time Prediction under Heterogeneous Traffic Conditions Using Global Positioning System Data from Buses." *IET Intelligent Transport Systems* 3 (1): 1–9.
- Veronesi, Pietro. 1999. "Stock Market Overreactions to Bad News in Good Times: A Rational Expectations Equilibrium Model." *Review of Financial Studies* 12 (5): 975–1007. <https://doi.org/10.1093/rfs/12.5.975>.
- Vlahogianni, Eleni I., Matthew G. Karlaftis, and John C. Golias. 2014. "Short-Term Traffic Forecasting: Where We Are and Where We're Going." *Transportation Research Part C: Emerging Technologies* 43: 3–19.
- Von Tomkewitsch, Romuald. 1982. *Method for Traffic Determination in a Routing and Information System for Individual Motor Vehicle Traffic*. Google Patents. <https://www.google.com/patents/US4350970>.
- Wahlström, Niklas, Roland Hostettler, Fredrik Gustafsson, and Wolfgang Birk. 2014. "Classification of Driving Direction in Traffic Surveillance Using Magnetometers." *IEEE Transactions on Intelligent Transportation Systems* 15 (4): 1405–1418.
- Walton, Charles A. 1983. *Electronic Identification and Recognition System with Code Changeable Reactance*. Google Patents. <https://www.google.com/patents/US4388524>.
- Washburn, Scott S., and Nancy L. Nihan. 1999. "Estimating Link Travel Time with the Mobilizer Video Image Tracking System." *Journal of Transportation Engineering* 125 (1): 15–20.
- Wen, Ying, Yue Lu, Jingqi Yan, Zhenyu Zhou, Karen M. von Deneen, and Pengfei Shi. 2011. "An Algorithm for License Plate Recognition Applied to Intelligent Transportation System." *IEEE Transactions on Intelligent Transportation Systems* 12 (3): 830–845.

- Westgate, Bradford S., Dawn B. Woodard, David S. Matteson, Shane G. Henderson, and others. 2013. "Travel Time Estimation for Ambulances Using Bayesian Data Augmentation." *The Annals of Applied Statistics* 7 (2): 1139–1161.
- Wu, Chun-Hsin, Jan-Ming Ho, and Der-Tsai Lee. 2004. "Travel-Time Prediction with Support Vector Regression." *IEEE Transactions on Intelligent Transportation Systems* 5 (4): 276–281.
- Xing, Tao. 2012. "Information-Theoretic Sensor Network Design and Reliable Route Guidance: Planning and Enriching Advanced Traveler Information Systems." The University of Utah. <http://content.lib.utah.edu/utils/getfile/collection/etd3/id/697/filename/637.pdf>.
- Xing, Tao, Xuesong Zhou, and Jeffrey Taylor. 2013. "Designing Heterogeneous Sensor Networks for Estimating and Predicting Path Travel Time Dynamics: An Information-Theoretic Modeling Approach." *Transportation Research Part B: Methodological* 57: 66–90.
- Yang, Baiyu, and Elise Miller-Hooks. 2002. "Determining Critical Arcs for Collecting Real-Time Travel Information." *Transportation Research Record: Journal of the Transportation Research Board*, no. 1783: 34–41.
- Yang, Hong, Kaan Ozbay, and Kun Xie. 2015. "Improved Travel Time Estimation for Reliable Performance Measure Development for Closed Highways." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2526: 29–38.
- Ye, Qixiang, and David Doermann. 2015. "Text Detection and Recognition in Imagery: A Survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37 (7): 1480–1500.
- Zeng, Xiaosi, and Yunlong Zhang. 2013. "Development of Recurrent Neural Network Considering Temporal-Spatial Input Dynamics for Freeway Travel Time Modeling." *Computer-Aided Civil and Infrastructure Engineering* 28 (5): 359–371.
- Zhang, Xuechi, Masoud Hamed, and Ali Haghani. 2015. "Arterial Travel Time Validation and Augmentation with Two Independent Data Sources." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2526: 79–89.
- Zhang, Yanru. 2015. "UNCERTAINTY ASSOCIATED WITH TRAVEL TIME PREDICTION: ADVANCED VOLATILITY APPROACHES AND ENSEMBLE METHODS." <http://drum.lib.umd.edu/handle/1903/16570>.
- Zhang, Yanru, and Ali Haghani. 2015. "A Gradient Boosting Method to Improve Travel Time Prediction." *Transportation Research Part C: Emerging Technologies* 58: 308–324.
- Zhang, Yanru, Ali Haghani, and Xiaosi Zeng. 2015. "Component GARCH Models to Account for Seasonal Patterns and Uncertainties in Travel-Time Prediction." *IEEE Transactions on Intelligent Transportation Systems* 16 (2): 719–729.
- Zhao, Wenjuan, Edward McCormack, Daniel J. Dailey, and Eric Scharnhorst. 2012. "Using Truck Probe GPS Data to Identify and Rank Roadway Bottlenecks." *Journal of Transportation Engineering* 139 (1): 1–7.
- Zheng, Weizhong, Der-Horng Lee, and Qixin Shi. 2006. "Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach." *Journal of Transportation Engineering* 132 (2): 114–121.
- Zhou, Xuesong, and George F. List. 2010. "An Information-Theoretic Sensor Location Model for Traffic Origin-Destination Demand Estimation Applications." *Transportation Science* 44 (2): 254–273.
- Zhu, Ning, Yang Liu, Shoufeng Ma, and Zhengbing He. 2014. "Mobile Traffic Sensor Routing in Dynamic Transportation Systems." *IEEE Transactions on Intelligent Transportation Systems* 15 (5): 2273–2285.