Locating Counting Sensors in Traffic Network to

Estimate Origin-Destination Volumes

by

Ning Wang

A Dissertation Presented in Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy

Approved April 2013 by the
Graduate Supervisory Committee:

Pitu Mirchandani, Chair
Alan Murray
Ram Pendyala
George Runger
Muhong Zhang

ARIZONA STATE UNIVERSITY

May 2013

ABSTRACT

Improving the quality of Origin-Destination (OD) demand estimates increases the effectiveness of design, evaluation and implementation of traffic planning and management systems. The associated bilevel Sensor Location Flow-Estimation problem considers two important research questions: (1) how to compute the best estimates of the flows of interest by using anticipated data from given candidate sensors location; and (2) how to decide on the optimum subset of links where sensors should be located. In this dissertation, a decision framework is developed to optimally locate and obtain high quality OD volume estimates in vehicular traffic networks. The framework includes a traffic assignment model to load the OD traffic volumes on routes in a known choice set, a sensor location model to decide on which subset of links to locate counting sensors to observe traffic volumes, and an estimation model to obtain best estimates of OD or route flow volumes.

The dissertation first addresses the deterministic route flow estimation problem given apriori knowledge of route flows and their uncertainties. Two procedures are developed to locate "perfect" and "noisy" sensors respectively. Next, it addresses a stochastic route flow estimation problem. A hierarchical linear Bayesian model is developed, where the real route flows are assumed to be generated from a Multivariate Normal distribution with two parameters: "mean" and "variance-covariance matrix". The prior knowledge for the "mean" parameter is described by a probability distribution. When assuming the "variance-covariance matrix" parameter is known, a Bayesian *A*-optimal design is developed. When the "variance-covariance matrix" parameter is unknown, Markov Chain Monte Carlo approach is used to estimate the aposteriori quantities. In all the sensor location model the objective is the maximization of the reduction in the variances of the distribution of the estimates of the OD volume. Developed models are compared with other available models in the literature. The comparison showed that the models developed performed better than available models.

I dedicate this dissertation to my lovely dad, mom and Siyao.

ACKNOWLEDGMENTS

I want to thank my advisor – Dr. Pitu Mirchandani for his mentoring, guidance and support through the course of my Ph.D. study. My career has been and will be continuously influenced by his dedication to research, his passion for the field of operations research, and his caring for students.

Thanks go as well to members of my dissertation committee, Dr. Alan Murray, Dr. Ram Pendyala, Dr. George Runger and Dr. Muhong Zhang. I want to thank them for their participation, supporting and constructive comments for my research proposal. I would also like to thank Dr. Monica Gentili. I have benefited from the constructive interactions with her on the research related to sensor location.

Very special thanks go to my colleague and friend, Zhuoyang Zhou, who was supportive through my Ph.D. study. I was always amazed with her talents and kindness.

I would like to express my deep appreciation to my beloved husband, Siyao Xu for his unconditional support throughout my studies. I must thank my parents for their long time encouragement and support.

TABLE OF CONTENTS

Page

LIST OF FIGURES viii

LIST OF TABLES x

CHAPTER

1 INTRODUCTION 1

1.1 Motivation 1

1.2 Research Overview 4

1.3 Research Contributions 13

1.4 Dissertation Organization 15

2 LITERATURE REVIEW 17

2.1 Type of Sensors 17

2.2 Traffic Assignment Models 20

2.3 OD Demand Estimation Models 23

2.4 Sensor Location Models 29

3 LOCATING SENSORS TO ESTIMATE ROUTE FLOWS USING DATA FROM NOISE-FREE SENSORS AND ASSUMING AN UNCERTAINTY INTERVAL FOR PRIOR MEANS 40

3.1 Problem Description 40

3.2 Notation 40

3.3 Route Flow Estimation Model 43

3.4 New Location Model Formulation 44

3.5 Algorithm 46

CHAPTER Page

3.6 Experimental Results ... 47

3.7 Chapter Conclusions ... 54

4 LOCATING SENSORS TO ESTIMATE ROUTE FLOWS USING DATA FROM NOISY SENSORS AND ASSUMING NORMAL DISTRIBUTIONS FOR PRIOR UNCERTAINTIES ... 56

4.1 Introduction and Problem Description ... 56

4.2 Bayesian Linear Model of Deterministic Route Flows ... 64

4.3 Locating Sensors to Maximize Variance Reduction ... 69

4.4 Algorithm ... 70

4.5 Experimental Result ... 73

4.6 Chapter Conclusions ... 84

5 LOCATING SENSORS TO ESTIMATE STOCHASTIC ROUTE FLOWS WITH KNOWN VARIANCE AND ASSUMING NORMAL DISTRIBUTIONS FOR PRIOR UNCERTAINTIES ... 86

5.1 Introduction and Problem Description ... 86

5.2 Hierarchical Linear Bayesian Model of Stochastic Route Flows ... 92

5.3 Sensor Location Model for Stochastic Route Flow Mean Estimation ... 97

5.4 Algorithm ... 101

5.5 Experimental Result ... 106

5.6 Chapter Conclusions ... 122

CHAPTER Page

6 LOCATING SENSORS TO ESTIMATE STOCHASTIC ROUTE FLOWS WITH UNKNOWN VARIANCE AND ASSUMING NORMAL DISTRIBUTION FOR PRIOR UNCERTAINTIES 124

6.1 Introduction... 124

6.2 Hierarchical Linear Model for Stochastic Route Flow Estimation 128

6.3 Bayesian Computation for Posterior Route Flow Mean 129

6.4 Sensor Location Modeling for Stochastic Route Flow Estimation 135

6.5 Experimental Result.. 138

6.6 Chapter Conclusions 145

7 SUMMARY AND FUTURE RESEARCH... 148

7.1 Summary of Research Results 148

7.2 Directions of Future Research 151

REFERENCES 154

LIST OF FIGURES

Figure Page

1 Relationship between traffic assignment and OD estimation models 5

2 A network example with 6 nodes and 6 links 5

3 Illustration of Sensor Location Flow-Estimation Problem 12

4 Dissertation Organization 15

5 Relationship among the seven categories of existing sensor location models 38

6 Procedure for the computational experiments for RVR-perfect 48

7 Topologies of experiment networks 50

8 Procedure for the computational experiments for RVR-noisy 75

9 Route flow posterior variance using RVR-noisy in network 1 80

10 Route flow posterior variance using RVR-noisy in network 2 80

11 Procedure for the computational experiments for Chapter 5 108

12 Relative Posterior Variances from Bayesian Estimation when Link Measurements are Independent (Network 1) 113

13 Relative Posterior Variances from Bayesian Estimation when Link Measurements are Independent (Network 2) 114

14 Relative SSE from Bayesian Estimation when Link Measurements are Independent (Network 1) 116

15 SSE from Bayesian Estimation when Link Measurements are Independent (Network 2) 117

16 Relative Posterior SSE from Bayesian Estimation when Link Measurements are Independent (Houston Network) 118

Figure Page

17 Posterior SSE from Bayesian Estimation when Link Measurements are Independent and Observe a Few Links (Houston Network) 119

18 Relative Posterior SSE Relative SSE from Bayesian Estimation when Link Measurements are Dependent (Network 1) 121

19 Relative Posterior SSE Relative SSE from Bayesian Estimation when Link Measurements are Dependent (Network 2) 121

20 Posterior SSEs from Bayesian Estimation when Link Measurements are Dependent and Observe a Few Links (Houston Network) 122

21 Procedure for the computational experiments for Chapter 6 141

22 Posterior SSEs Estimated using MCMC Method with Good Prior for Φ 144

23 Posterior SSEs Estimated using MCMC Method with Bad Prior for Φ 145

LIST OF TABLES

Table Page

1 OD pairs and corresponding routes for the network of Figure 2 6

2 RMAE Results using RVR-perfect for Grid Network 1...................................... 51

3 RMAE Results using RVR-perfect for Grid Network 2...................................... 52

4 RMAE Results using RVR-perfect for Problem Instances for *N*=1 to *N*=1/2|*A*|. 53

5 *P*-values for Paired-*t* Tests Comparison of RMSE of RVR-perfect.................... 54

6 SSE Results using RVR-noisy for Grid Network1 78

7 SSE Results using RVR-noisy for Grid Network 2 79

8 SSE Results using RVR-noisy for Problem Instances for *N*=1 to *N*=1/2|*A*| 82

9 SSE Results using RVR-noisy for Problem Instances for *N*=1 to *N*=1/4|*A*| 83

10 *P*-values for Paired-*t* Tests Comparison of SSE of RVR-noisy 84

11 Posterior Variances from Bayesian Estimation using Different Sensors Location. 111

12 Posterior SSE for Five Sensor Location Models when Link Measurements are Independent... 115

13 Posterior variances and bias from Bayesian Estimation when Link Measurements are Dependent 120

Chapter 1

INTRODUCTION

1.1 Motivation

According to the U.S. Department of Transportation (FHWA, 2011), the number of registered vehicles in the US has increased steadily since 1960 and reaches 246 million in 2009. The total vehicle miles traveled in 2009 reached 3 trillion and consumed 172 billion gallons of fuels. In most recent two decades, transportation averagely consumed 9~10 percent of annual Gross Domestic Product (GDP), over 70% of which are related to the highway and transit. The rapid growth in travel demand and the slow growth in supply of roads and public transportation cause large increases in congestion on capacity-limited transportation networks. The annual cost of traffic congestion is more than $100 billion and $750 for every U.S. commuter. According to Urban Mobility Report (2011), traffic congestion threatens the economic competitiveness and productivity of the nation and is currently becoming a global issue.

Origin-Destination (OD) trip demands, which specify the amount of trips between each pair of origin and destination nodes, are required by many traffic planning and management applications. Examples of transportation planning and management decisions are: how links in the traffic network should be constructed; how to evaluate the effects for speed limit and number of lanes; how to introduce road tolls, etc. (Peterson, 2007). Mathematical models can represent a system of traffic flows and its observed travel patterns (Ortúzar and Willumsen, 1994; Oppenheim, 1995). The predicted travel patterns then provide useful information to support the decision-making in planning and managing transportation systems. In practice, travel demand models are tools to predict travel patterns under various conditions. However, complex traffic demand is difficult to estimate because of its dynamic and stochastic structure. For example, stochastic dynamic traffic demand varies significantly by the time of day and the day of week. Lacking the ability of providing high-quality OD demand estimates limits the effectiveness of the evaluation and implementation of traffic planning and management decisions, and restricts the potential of technology deployments to control traffic congestion and enhance the mobility in traffic networks.

Traditional methods of obtaining an OD demand utilize home interviews, roadside surveys (or direct sampling estimation) or physical and/or behavioral models (for example, the gravity-type trip distribution model). These procedures are usually costly and time-consuming and they have seldom been researched or applied frequently. The OD demands are often estimated from link flow volumes which are measured using traffic sensors or detectors. Much research has been conducted to study the relationship between measured link traffic counts and the corresponding traffic demand estimates.

The quality of estimated OD demands from link counts depends on several factors, such as (1) the route-choice and traffic loading assumptions, (2) the quality of observed data from sensors, (3) the dependencies between link flows due to network topology and traffic loading, (4) the choice of OD estimation methods, and (5) where the sensors are located (Larsson et al., 2010). The first factor is paramount because different traffic loading assumptions, such as equilibrium or near-equilibrium assumption, are involved in each estimation model, implicitly or explicitly. The observed data from sensors is the direct input of OD demand estimation model, and the reliability of the counting devices and the accuracy of data can be improved through technologies and data collecting methods respectively. The third factor comes from the complexity of the problem scenario. The last two factors reveal two important research questions:

- How to compute the best estimates of the flows of interest by using anticipated data from given candidate sensors location;
- How to decide on the optimum subset of links where counting sensors are to be located.

These two research questions constitute the bilevel *Sensor Location Flow-Estimation* (*SLFE*) problem defined by Gentili and Mirchandani (2011, 2012). The upper level is an optimization model that selects the best location set based on lower level solutions for each candidate set, while the lower level is an optimization model that computes estimates to minimize the expected estimation errors using the anticipated data from sensors.

To date, the potential benefits of enhancing travel demand modeling capability in modeling both levels of *SLFE* problem together have not been adequately addressed. The theoretical and algorithmic aspects of the stochastic OD demand estimation problem and the corresponding sensor location problem are still relatively undeveloped. In order to enhance the methodological capabilities required for traffic planning and management decision, the following challenging questions need to be addressed:

(1) How to effectively amalgamate information from different sources, e.g. a-priori knowledge of modelers and observations from sensors, especially when there are uncertainties in the prior knowledge?

(2) What is the optimal strategy to locate the sensors in a network in order to optimize the OD estimates?

(3) How to consistently handle the implicit or explicit traffic assignment or traffic loading assumptions at both levels of *SLFE* problem?

(4) What are the decision strategies for *SLFE* problem for deterministic and stochastic traffic demands?

1.2 Research Overview

1.2.1 Traffic Loading Assumption

Traffic assignment model, or traffic loading model, aims to determine the number of trips on different links of the network given the travel demands between different OD pairs from the mathematical description of the route choice behaviors. Suppose $\mathbf{q}$ is the OD flows in a network; $\mathbf{x}$ and $\mathbf{v}$ are the corresponding route and link flows loaded by the actual/assumed methods. Figure 1 illustrates the relationship between traffic assignment (flow direction with solid line) and OD estimation (flow direction with dotted line). "Traffic assignment model" splits the OD volumes $\mathbf{q}$ into route volumes according to specific traffic assignment rules as shown in Eq. (1.1). The link flows $\mathbf{v}$ are simultaneously obtained from the network topology as Eq. (1.2) — link flow on a particular link is the summation of flows on routes that pass through it. Eq. (1.3) expresses the relationship between OD flows $\mathbf{q}$ and link flows $\mathbf{v}$ when both traffic loading method and network topology are known. As an inverse approach, "OD estimation model" uses the link flows $\mathbf{v}$ as input and allocates such link flows into OD pairs based on the measurement model in Eq. (1.3).

$$\mathbf{x} = assign(\mathbf{q}) \tag{1.1}$$

$$\mathbf{v} = top(\mathbf{x}) \tag{1.2}$$

$$\mathbf{v} = assign\_top(\mathbf{q}) \tag{1.3}$$

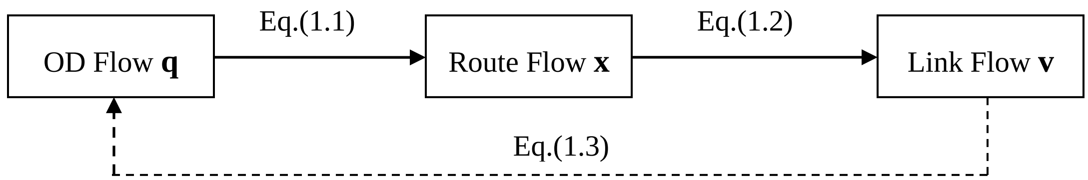

Figure 1 Relationship between traffic assignment and OD estimation models

Usually, the relationship expressed by Eq. (1.2) depends on a defined parameter ρ_a^i, where $\rho_a^i = 1$ if link *a* is a part of route *i*, and $\rho_a^i = 0$ otherwise.

As an example of above relationships, consider the simple network (from Yang et. al, 1991) shown in Figure 2. It includes 4 OD pairs: 1-5, 1-6, 2-5 and 2-6. Each OD pair is connected by two different routes: one using link a_3, the other using link a_4.

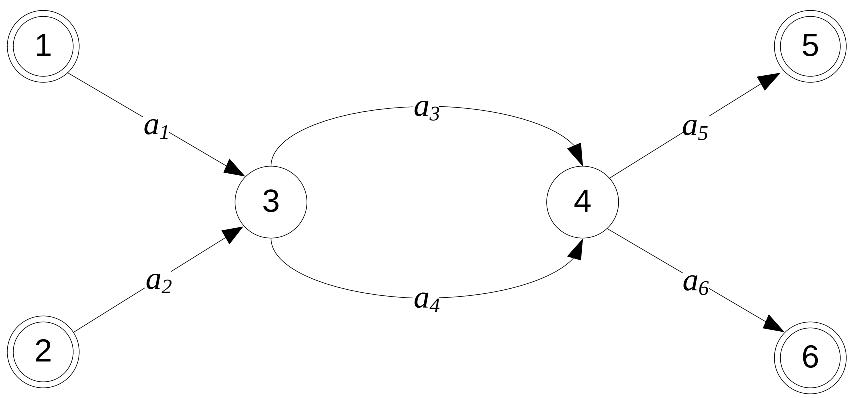

Figure 2 A network example with 6 nodes and 6 links

Table 1 OD pairs and corresponding routes for the network of Figure 2

OD	Route	Links	OD	Route	Links
1-5	R1	a_1-a_3-a_5	2-5	R5	a_2-a_3-a_5
	R2	a_1-a_4-a_5		R6	a_2-a_4-a_5
1-6	R3	a_1-a_3-a_6	2-6	R7	a_2-a_3-a_6
	R4	a_1-a_4-a_6		R8	a_2-a_4-a_6

Table 1 contains the information about OD pairs and the corresponding routes in the network of Figure 2. For example, OD pair 1-5 contains two routes R1 and R2. When x_i is flow on route R*i*, the relationships between OD flow vector **q** and route flow vector **x** describing Eq. (1.1), for the network of Figure 2 are:

$$x_1 + x_2 = q_{1-5} \quad (1.4)$$

$$x_3 + x_4 = q_{1-6} \quad (1.5)$$

`

$$x_5 + x_6 = q_{2-5} \quad (1.6)$$

$$x_7 + x_8 = q_{2-6} \quad (1.7)$$

For example, $\rho_3^1 = 1$(since link a_3 is on route R1), but $\rho_4^1 = 0$(since link a_4 is not on route R1). The relationships in Eq. (1.2) between connect route flows **x** and link flows **v** on links a_3 an a_4 are:

$$x_1 + x_3 + x_5 + x_7 = v_3 \quad (1.8)$$

$$x_2 + x_4 + x_6 + x_8 = v_4 \quad (1.9)$$

The coefficients for route flows x_i in the set of linear equations above are the link route parameters ρ_a^i. All these coefficients define the link-route topology through the incidence matrix **H**. The incidence matrix **H** for link flows v_3 and v_4 due to the eight route flows in the network can be written as:

Link \ Route	1	2	3	4	5	6	7	8
3	1	0	1	0	1	0	1	0
4	0	1	0	1	0	1	0	1

Often, a major underlying traffic loading assumption is used in estimating static OD trips. The assumption is that the link flows observed are from an equilibrated flow pattern where Wardrop's First Principle (Wardrop, 1952) holds (that is, all used routes between an origin and a destination have equal and minimum travel times, while other routes between that OD pair have no flow from that OD demand). In the corresponding mathematical network equilibrium models, a very large number of routes (some could be quite unusual) may result for an OD pair. Furthermore, it is generally believed that not all travelers perceive cost in the same way so that a perceived travel time may differ from actual travel time. Therefore the observed OD flow patterns are unlikely to be in a deterministic equilibrium where every traveler minimizes his/her actual travel time (e.g., Daganzo and Sheffi, 1977; Mirchandani and Soroush, 1987; Ortúzar and Willumsen, 1994). This dissertation assumes that a *route choice set* (or simply a *choice set* when no confusion may arise) is associated with each OD pair and includes all the routes that may be used for that OD pair. This is a generalization of the equilibrium assumption. Indeed, it includes equilibrium flows when the choice set is obtained from a traffic equilibrium model. However, this model also allows the case when the observed link flows come either from nearly equilibrated or other OD flow patterns that only use OD choice sets. The only assumption the developed models use is that the route choice set for each OD pair is known and its cardinality is not large. Thus, in the models developed in this dissertation, the observed data need not be restricted to an equilibrated pattern but, due to sensors errors, traveler perceptions, and modeling approximations, need only fit a traffic loading model with finite sets of OD routes. Hence, the assumption of known route choice set in this dissertation is a more general model since these route sets could be loaded so that equilibrium or near-equilibrium traffic pattern results.

These OD routes may have been empirically observed or may be from census data; they may have come from a traffic loading model (possibly from a traffic equilibrium model where only routes with significant flows are identified as important); they may be from a minimum total cost flow model; they may be a set of “efficient” routes (e.g., Dial, 1997), or they may have been developed using some other reasonable procedure.

Furthermore, by assuming knowledge of the route choice sets, the typical OD estimation problem can be converted to a route flow estimation problem. Thus, this dissertation focuses on estimating the route flows when the route choice set for each OD pair is known.

1.2.2 Demand Modeling

The traditional traffic models used in traffic planning assume steady-state flows and try to model the total or average behavior of traffic flow in given periods of time. These models input the deterministic OD demands and output total link or path flows in selected periods (hour, day, week, etc.). The deterministic OD demands are usually referred to as OD matrices, the element of which represents the number of trips moving regularly from one zone to another at a specific time. In the deterministic OD estimation problem, the traffic demand to be inferred is the OD matrix that represents the mean or expected number of trips for each OD pair.

However, in reality, the static traffic demand has a stochastic nature that needs to be considered. In addition to the steady-state flows described by the deterministic demands, the normal periodic variations could be included in demand models. Daily variations may be caused from some frequently-occurring events, such as varying trip-taking behavior, minor accidents, weather-related traveler decision making, road maintenance and traffic signal failures. The resulting stochastic OD demands can be described by two parameters for the steady-state flow pattern and periodic variations respectively. The stochastic OD demand estimation problem is to infer one or both of these parameters from sensor data.

In this dissertation, both deterministic and stochastic traffic demand patterns are addressed.

1.2.3 Types of Information

In the problem space, assume there are two roles taking actions and each role has the control of certain types of information. The first role is the personification of “Nature” who is aware of the “actual” OD demands in the network, as well the “real” traffic assignment rules and the corresponding “actual” route flows and link flows.

The second role in the problem space is the “modeler”, who does not have the same privilege of knowing the “real” traffic demand but is interested in inferring it from the sources on hand. Two types of information are usually available to the modeler — observations and a-priori information.

In this dissertation, observations refer to as the link flow volumes from counting sensors or detectors. The traffic sensors located on a lane or a set of lanes of the road count the number of vehicles passing over them per hour. The traffic "counts" are then translated to link flow volumes, which are used as observation information by the route flow estimation models. In addition, when the link data is collected from noise-free or "perfect" sensors, then no measurement errors are assumed and the observation equals to the real link flow. If "noisy" sensors are used (as most cases in practice), then a random measurement error may be added to the real link flows to model observed flows. The measurement error ε is usually modeled having Gaussian distribution with mean zero and a known variance. Error variance describes the reliability of a particular sensor or detector and can be estimated from the same type of sensors at similar locations or from studies related to sensor errors.

A-priori information represents the prior estimates of traffic volumes without the awareness of any observations. These prior route flows can be obtained from survey data or historical data. They could also be derived from a calibrated static traffic assignment models for planning purposes, or the priors on the routes could be such that the resulting flows are not too far from the mathematically developed flows using an equilibrium model since it is known that travel time is the predominant factor that travelers use in selecting routes from their origins to their destinations. In practice, it is common that the prior information may come from an out-of-date study, naturally with perception errors.

The simplest version of apriori information is a point value or a mean value, which represents the expected value of flow volumes. The modeler can represent the uncertainty of information by a confidence interval or a parameter denoting the degree of belief. In general, this uncertainty is based on one or more of these factors: (a) quality of historical data (empirical uncertainty); (b) our subjective estimation (e.g., the subjective probability that a given team will win, the kind of subjectivity gamblers use), and (c) physical models, (e.g., the probability a six will come up with a roll of a die is 1/6). One may imagine that the uncertainties will be based on the planners' subjective feeling of the numbers they have based the combination of empirical evidences and anecdotal observations. The prior information can usually be expressed as a probability distribution, with a specified mean and variance. A prior distribution has a variance of zero if the mean is known precisely, while it has a very large variance if it is assumed that very little is known about the mean value.

1.2.4 Problem Description

The problem addressed in this dissertation is the *Sensor Location Flow-Estimation* (SLFE) problem, where the upper level decides on the optimum subset of links where sensors are to be located, and the lower level computes the best estimates of the flows of interest by using anticipated data from given candidate sensors location. In general, the upper level is an optimization model that selects the best location set based on lower level solutions for each candidate set, while the lower level is an estimation model, also an optimization model that calculates the best estimates by minimizing the expected estimation errors using the anticipated data from sensors.

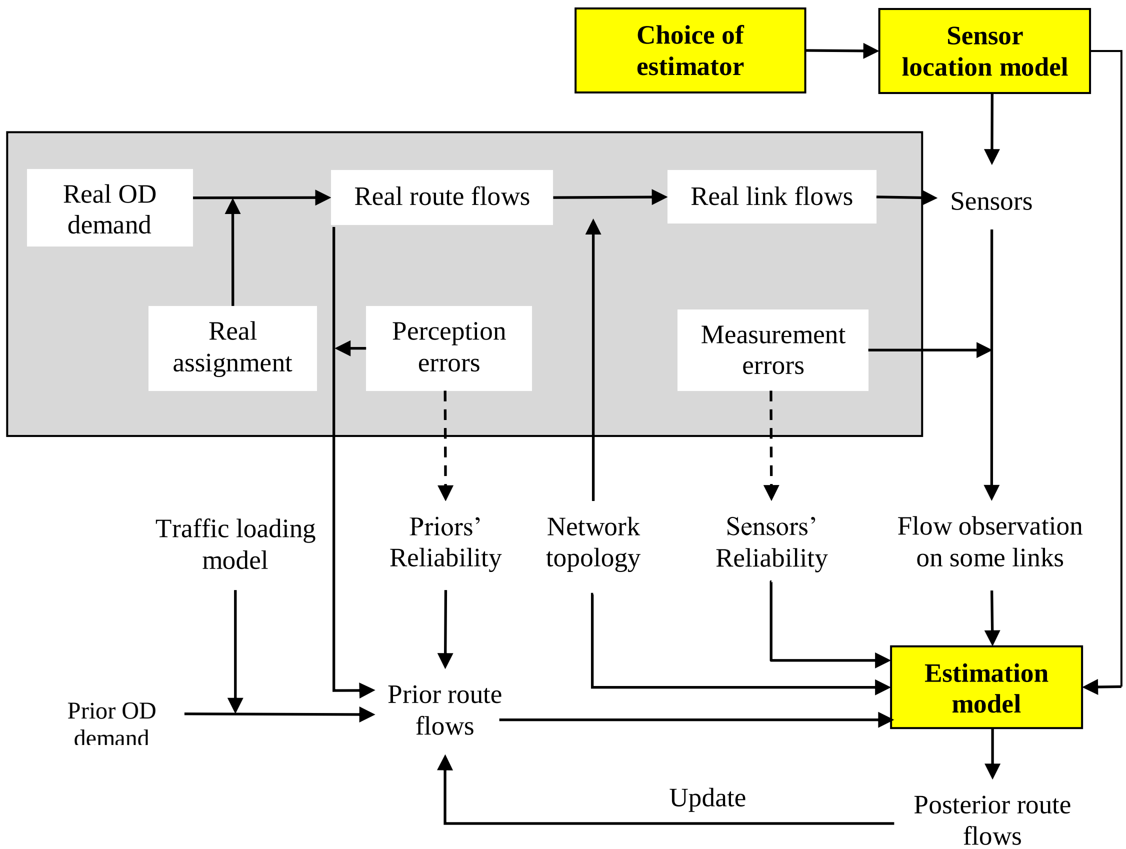

Figure 3 Illustration of Sensor Location Flow-Estimation Problem

The *SLFE* problem is illustrated in Figure 3. Consider estimating OD demands as a statistical decision problem. The estimation problem can be considered as a statistical two-person zero-sum game of the "modeler" against "Nature". Nature controls the real OD demand for a network and assigns them onto routes and obtains the corresponding links flows. Nature also chooses a probability distribution for the random measurement error and a probability distribution for the random perception error. The prior route flows result by adding the perception errors to the real route flows. With given sensors' location (the choice by modeler), the measurement errors are added to the "selected" real link flows if the detectors are not accurate. Nature's actions are depicted in the large gray box with dashed line in Figure 3.

The modeler chooses an estimation method (an estimator) for traffic demand and the sensors' location (an experimental design). Considering traffic data collection as an experiment, locating the sensors in the traffic network is an experimental design problem. For a given design, the route flows will be estimated according to the estimation method, and the goodness of the estimation is valued by a loss function, e.g. the quadratic loss. The modeler intends to make several decisions based on the information out of the large gray box of Figure 3. The three tasks of the modeler in the problem space (shown in three small gray boxes in Figure 3) are:

(1) Select an the estimation method (estimator)

(2) Select the sensors' location (design)

(3) Estimate route flows according to the estimator and design

For a given estimator, tasks 2 and 3 define the both levels of *SLFE* problem. Tasks 1 and 2 together can provide an optimal statistical decision strategy. Bayes strategy is one type of optimal statistical decision strategy that is related to the Bayesian estimator, which provides a natural and mathematically convenient way of combining prior knowledge with the observations. The Bayes strategy provides the guideline or objective about how to decide the sensors' location when Bayesian approach is applied to estimate the traffic demand.

1.3 Research Contributions

The aim of this dissertation is to develop a decision framework for locating sensors to obtain the high-quality OD volume estimates in traffic networks. The major contribution of this dissertation is that it develops the decision framework for the recently defined *Sensor Location Flow-Estimation* problem. The developed framework is an integration of several well-defined problems in traffic modeling, such as a traffic assignment model to load the OD volumes on routes and corresponding links, a sensor location model to make the decision about which subset of links to observe, and an estimation model to obtain best estimates of OD or route flow volumes. Each sub-problem is an optimization problem with specific objective. The advantage of integrating the individual problems is to improve the quality of traffic demand estimates. Within the decision framework, the sensor location problem is the focus of this dissertation.

The second major contribution of this dissertation is that the proposed decision framework is compatible with both deterministic and stochastic demand estimation problems in traffic networks. The traditional traffic demand modeling problem focuses on deterministic demands, or deterministic OD matrix estimation. This dissertation expands the capability of traffic demand estimation to stochastic demands. Specifically, four types of location models are developed for scenarios (1) when demand is deterministic and data is from noise-free or “perfect” counting sensors, (2) for deterministic demand and noisy sensors, (3) stochastic demand with known variance and noisy sensors, and (4) stochastic demand with unknown variance and noisy sensors, respectively.

As the third contribution, this dissertation develops an experimental environment which can handle the evaluation and comparison of different sensor location models in terms of OD estimation qualities. The experimental environment has the following features:

(1) The sensor location method being used is independent of the environment;

(2) The relationship between the choice of sensor location method and the quality of the estimated OD demand can be isolated and evaluated;

(3) The environment is capable of handling both deterministic and stochastic OD demand situations;

(4) The environment can provide the caparison of sensor location methods with minimum effect of traffic assignment assumptions which are coupled implicitly or explicitly with OD demand estimation.

1.4 Dissertation Organization

This dissertation includes seven chapters. The interconnection among the remaining chapters, and assumptions of the corresponding models addressed are shown in Figure 4.

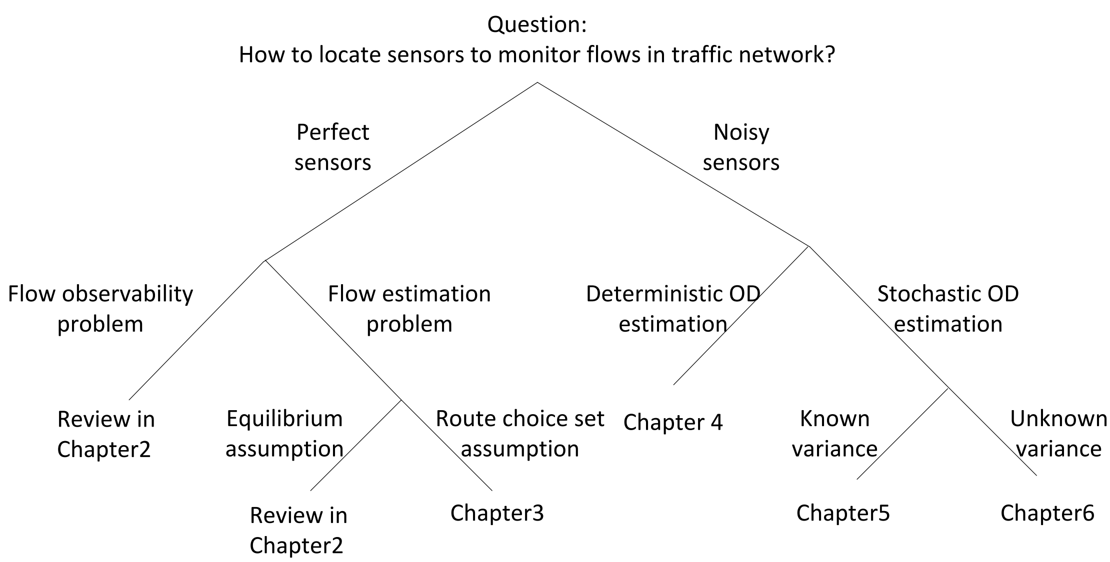

Figure 4 Dissertation Organization

The remainder of this dissertation is organized as follows:

Chapter 2 provides a comprehensive review and discussions on the sub-problems in *SLFE*.

Chapters 3 and 4 focus on deterministic route flow estimations. Prior information for the deterministic route flow is assumed as a prior distribution or a confidence interval. In Chapter 3, a new linear integer programming model is presented to locate "perfect" sensors in order to maximize the reduction in uncertainties of route flows estimates from a Generalized Least Squares (GLS) estimator. A greedy algorithm is developed to solve the model. In Chapter 4, a new model is presented to locate "noisy" sensors to minimize the variances in route flow posteriors using a linear Bayesian estimator. A sequential algorithm is proposed to obtain the optimal sensor locations.

Chapters 5 and 6 address the stochastic route flows and estimate them by hierarchical linear Bayes estimator. In Chapter 5, the route flow variances are assumed to be known. The posterior distribution for the route flow mean is estimated by a Bayesian approach. In Chapter 6, the route flow variances are unknown. In the model developed the route flow means are estimated by a Markov Chain Monte Carlo algorithm. The sensor location objective is formulated to minimize the approximation of expected posterior variances.

Chapter 7 summarizes the research works in this dissertation and points out some future directions in both the application and algorithm development research.

Chapter 2

LITERATURE REVIEW

This chapter reviews topics relevant to *Sensor Location Flow-Estimation* problem. Section 2.1 introduces the traffic detection techniques and the potentials and challenges when applying to the traffic demand modeling. Traffic assignment problem, which is the inverse of the OD estimation problem, is reviewed in section 2.2. Section 2.3 reviews OD demand estimation models from the aspect of different available information. The sensor location models for OD demand estimation is reviewed in section 2.4.

2.1 Type of Sensors

Sensors are widely used in all aspects of engineering and science, such as automotive and transportation, physics, chemistry and biology, etc. A traffic sensor is a device that indicates the presence or passage of vehicles and provides data or information to support traffic management applications, such as signal control, freeway mainline and ramp control, incident detection, and data gathering of vehicle volume and classifications to meet State and Federal reporting requirements (Klein et al., 2006). Intelligent Transportation Systems (ITS) are highly dependent on traffic sensors, such as automatic traveler surveillance, real-time traffic adaptive signal control, and emergency information services, etc.

Traffic detectors can be categorized into “in-roadway” sensors and “over-roadway” sensors (Klein et al., 2006). In-roadway sensors are embedded in the pavement or subgrade of the roads, taped or attached to the surface of the road. Over-roadway sensors are located above the road or alongside the road.

Sensors can also be categorized as "passive" sensors and "active" sensors. For example, the inductive loop detector embedded on the lane of a road is a *passive sensor* that the vehicle does not actively generate a signal on its own; when the vehicle passes over the loop, the change in the magnetic field sends an electrical signal to indicate the passage of the vehicle. Generally, such loop detectors are used to count the number of vehicles passing over them in a period of time and the vehicles movement monitored by the passive sensors is anonymous. Other examples of passive sensors include passive acoustic, passive infrared and microwave radar detectors. In addition, images of the moving flows can be obtained with cameras installed on links and nodes and the image information can be processed for flow volumes, speeds, travel times, queues and turning ratios. Usually passive sensors provide measurements at point.

Active sensors include Vehicle-ID sensors and Path-ID sensors. Vehicle-ID sensors can identify vehicles on the network with an identification number. The examples of vehicle-ID sensors are license plate readers that use camera images, and Automatic Vehicle Identification (AVI) readers that use RFID tags or bar-codes. A purpose of license plate readers could be to monitor travel times, while AVI readers are normally used to collect tolls on equipped vehicles. Vehicle-ID sensors are able to provide point-to-point measurement data. Path-ID sensors are located to measure the flow volumes on each planned route for some special vehicles with electronic tags containing identification information, such as commercial trucks, buses, emergency vehicles, and trucks carrying hazardous material, and they can monitor paths and their flows in the network. With advances in Geographic Information System (GIS) and telecommunication, Automatic Vehicle Location (AVL) technologies, such as Global Positioning System

(GPS), electronic distance measuring instruments (DMI's), cellular telephone and smartphone tracking, have been widely use in recent years to dramatically increase the quality and quantity of traffic data. AVL can provide semi-continuous path trajectory for individual equipped vehicles (Zhou, 2004).

Above traffic devices are available to collect and process different types of traffic data, such as point, point-to-point and path measurements, which formulate different measurement models for traffic monitoring. The relationship between unknown route demands and the measurements for both point and point-to-point type is:

$$\hat{\mathbf{v}} = \mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon}, \text{ where } \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \Sigma) \quad (2.1)$$

where $\hat{\mathbf{v}}$ includes both link count and vehicle identification counts; $\mathbf{H}$ is the parameter to indicate linear the route-link choice relationship; $\boldsymbol{\varepsilon}$ is the measurement error with known variance-covariance matrix $\boldsymbol{\Sigma}$. Flow measurements at points from counting sensors directly apply to Eq. (2.1). Point measurements from counting sensors are the only type of observation data considered in this dissertation. The reasons include:

(1) The counting sensor is the most predominant device in traffic planning practice with low installation cost and high accuracy (Klein et al., 2006);

(2) Although advanced sensor devices provide a data rich environment, the derivation from data to demand estimates usually relies on other parameters which are difficult to estimate (such as the vehicle-ID market penetration rate);

(3) The linear measurement model from counting sensors (see Eq. (2.1)) is the foundation of OD estimation models and provides the general form of the relationship between observations and flows of interest to be estimated.

2.2 Traffic Assignment Models

In traffic planning, there is a traditional four-step process to model traffic demands. This approach was originally developed in the 1950s and still widely used for traffic planning. The first phase, “trip generation”, is designed to estimate the number of trips originating in, and/or ending in given zones (demands on nodes). The second phase, “trip distribution”, forms the travel demands (origin-destination demands or OD demands) by connecting the node demands in the network to each other. “Mode split” procedure partitions the OD demands into different travel modes. Finally, in phase “trip assignment”, the OD demand for each mode is assigned on the traffic network based on reasonable assumptions and hence route flows or link flows can be calculated.

Because there are many possible routes from each origin to destination, the assignment procedure must use some assumptions on how the routes are chosen. The most common assumption is that each traveler chooses the route with least instantaneous generalized cost. The generalized cost for a route is assumed to be the summation of the travel costs on all included links. The generalized link cost is usually a function of free flow travel time (the constant link characteristic) and the congestion level on the links. In most models, the link cost function is an exponential or a higher order polynomial function which is monotonically non-decreasing with link flow. The link cost grows rapidly when its flow approaches the maximum capacity.

The most widely used assumption to assign the OD travel demand over alternative routes is based on Wardrop’s First Principle (Wardrop, 1952), which states:

“The journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route.”

Any traffic assignment which can satisfy this criterion is called a user-equilibrium. Define x_i as the route flow on route *i* between OD pair *j* ($\overline{OD}$ is the set containing all OD pairs in a network; R_j is the route choice set for OD pair *j*); v_a is the link flow variable for link *a* ($\overline{A}$ is the set of links in a network) and c_a is the cost on link *a* depending on the link volume v_a, the user-equilibrium (UE) assignment of OD demand q_j onto links in the network can be obtained by optimally solving the following mathematical program:

$$\min_v f(v) = \sum_{a \in A} \int_0^{v_a} c_a(s) ds \qquad (2.2)$$

$$s.t. \quad \sum_{i \in R_j} x_i = q_j, \forall j \in \overline{OD} \qquad (2.3)$$

$$\sum_{j \in OD} \sum_{i \in R_j} \rho_a^i x_i = v_a, \forall a \in \overline{A} \qquad (2.4)$$

$$x_i \geq 0, \forall i \in R_j, \forall j \in \overline{OD} \qquad (2.5)$$

The objective function Eq. (2.2) minimizes a convex function so that user-equilibrium results given the linear constraints (Beckmann el al., 1956). Eq. (2.3) represents a set of flow conservation constraints. These constraints ensure that the OD demand for each OD pair *j* equals the summation of all alternative routes connecting each O and D. A set of definitional constraints Eq. (2.4) express the link flows v_a in terms of the route flows x_i (link flow volume on any link *a* equals to the sum of route flows on it). The link-route parameter ρ_a^i is defined to assist the formulation of Eq. (2.4), where ρ_a^i is 1 means route *i* include link *a*.

This formulation was first developed by Beckmann et al. (1956), who also proved the existence and uniqueness of the solution when the link cost is a monotonically increasing function of link flow volume. The application of Frank and Wolfe's algorithm to the solution of the mathematical formulation of UE was first suggested by Bruynooghe et al. (1968) and applied by Murchland (1969). Incremental assignment techniques and capacity restraint methods are two heuristics widely used in practice to find the UE flow pattern over a network (Sheffi, 1985).

In the deterministic UE assignment described above, each traveler chooses his/her path with least generalized cost. This assumption does not consider the variation in the travelers' different perception of travel time (cost). Daganzo and Sheffi (1977) extended the Wardrop's user equilibrium condition to the principle of stochastic user equilibrium (SUE) as:

"In a stochastic user equilibrium network no user believes he/she can improve the travel time by unilaterally changing routes."

The definition of SUE assumes the link cost to be random and flow dependent. Mathematically this is modeled by adding an error term to the generalized cost of each route. The randomness in the cost comes from the variability in travelers' perception.

The problem to find a SUE flow pattern can be formulated by an unconstrained minimization program (Daganzo, 1979; Sheffi and Powell, 1982; Daganzo, 1982). The method of successive averages (MSA) is an algorithm that has been applied in practice to obtain a SUE solution (Sheffi and Powell, 1982).

A time-dependent (dynamic) traffic model considers the influence of traffic conditions in a certain time period on any succeeding time period. In a time-dependent traffic assignment model, the interaction between time and vehicle volumes needs to be additionally described. Dynamic traffic assignment (DTA) models describe the dynamics of network flow propagation and travelers' behavior in response to past and/or current time dependent information. Peeta and Ziliaskopoulos (2001) provided a state-of-art review and detailed discussion of DTA formulation approaches, model objectives, assumptions, solution methodologies, traffic flow modeling strategies, operational requirements and capability, etc.

In this dissertation, the traffic assignment assumption is that a known route choice set is associated with each OD pair and includes all the routes that may be used for that OD pair. This assumption allows the equilibrated pattern when the choice set is obtained from a traffic equilibrium model. However, it also allows the case when the observed link flows come either from nearly-equilibrated or other OD flow patterns that only use OD choice sets. The only assumption the developed models use is that the route choice set for each OD pair is known and its cardinality is not large. Based on this assumption, the OD estimation problem is converted to the route flow estimate problem since route flows give easily and directly both OD volumes and link flows.

2.3 OD Demand Estimation Models

Generally, monitoring flows $\mathbf{x}$ from flow volumes observed by sensors $\hat{\mathbf{v}}$ depend on a system of linear equations as

$$\hat{\mathbf{v}} = \mathbf{H}\mathbf{x} \tag{2.6}$$

where matrix **H** defines the route-link relationship. The row amounts in Eq. (2.6) come from the measurement data by sensors. If the data-to-flow matrix **H** is full rank, the unique solution to the system of linear equations can be determined by matrix inversion operations, and this formulates the "flow-observability problem" (Gentili, 2002).

Otherwise, additional information, such as the a-priori route flow information, is required to formulate the "flow-estimation problem" (Gentili and Mirchandani, 2011, 2012).

In traffic networks, the number of counting sensors (the number of rows in Eq. (2.6)) is generally less than the total number of ODs (the number of columns in Eq. (2.6)). Hence, in practice, there are usually not enough linear independent equations of measurement to result in a unique solution of OD estimates without using additional information. The OD estimation problem attempts to find OD demand estimates which reproduce the observed link flow counts when the demands are assigned on the network.

According to Abrahamsson (1998), an estimated OD matrix is produced by combining the prior information $\hat{\mathbf{q}}$ and observations $\hat{\mathbf{v}}$ using the general formulation as follows:

$$\min F(\mathbf{q}, \mathbf{v}) = \gamma_1 F_1(\mathbf{q}, \hat{\mathbf{q}}) + \gamma_2 F_2(\mathbf{v}, \hat{\mathbf{v}}) \tag{2.7}$$

$$s.t \qquad \mathbf{v} = assign\_top(\mathbf{q}) \tag{2.8}$$

In this model, OD volumes are estimated by minimizing two distances: the distance between estimated OD matrices $\mathbf{q}$ and prior OD $\hat{\mathbf{q}}$ (measured by distance function $F_1(\mathbf{q}, \hat{\mathbf{q}})$) and the distance between estimated link flows $\mathbf{v}$ and observed $\hat{\mathbf{v}}$ (measured by the distance function $F_2(\mathbf{v}, \hat{\mathbf{v}})$). Usually the distance measure F_1 is mainly the minimum-information metric (such as maximum entropy) and F_2 is often a Euclidean metric. In objective function Eq. (2.7), different weights γ_1 and γ_2 are assigned to the distance measures F_1 and F_2 respectively and the values of the weights depend on the reliability of the corresponding means. If prior OD matrix $\hat{\mathbf{q}}$ is more reliable than observations $\hat{\mathbf{v}}$, then the weight γ_1 should be larger than γ_2. On the other hand, γ_2 is greater than γ_1 if observations $\hat{\mathbf{v}}$ are more reliable than prior OD matrix $\hat{\mathbf{q}}$. Constraint defined by Eq. (2.8) describes the general relationship between link flow estimates **v** and OD volume estimates **q**, in particular the relationship defined by traffic assignment and network topology.

The deterministic OD estimation models can be categorized into:

(1) Approaches based on traffic modeling concept, such as the gravity distribution model.

(2) Statistical inference approaches, includes the Maximum Likelihood, Generalized Least Squares and Bayesian Inference approaches. In this category, the traffic volumes and the prior OD matrix are assumed to be generated by some probability distributions, and the OD estimates are obtained by estimating the parameters of the probability distributions.

In the first category, Van Zuylen and Willumsen (1980) first proposed two gravity type models based on entropy maximization and information minimization principles to find an OD matrix that reproduces the observed link flows. Fisk (1988) extended the entropy model to the congested networks by introducing the constraints of user-equilibrium conditions. The proposed model has two levels; it maximizes the entropy at the upper level and solves the user-equilibrium problem at the lower level. Fisk (1989) showed that the extended entropy model has the same solution as a combined trip distribution and assignment model when the observed flow pattern is a user-equilibrium flow. For the combined model, the number of trips originating in and ending in each zone (O and D respectively) may be expressed as constraints. The observed traffic counts are reproduced by a combined model if observed counts for all links are available and if they are consistent with user-equilibrium. Fisk and Boyce (1983) proposed a method to estimate the weighted average link cost by the importance of various link types when the observations are only available on a subset of links in the network. Kawakami et al. (1992) extended Fisk and Boyce's combined model to include two modes of travel - large size trucks and cars. Tamin and Willumsen (1989) presented both a gravity-opportunity model and an intervening opportunity model and applied the gravity model to a small test problem without congestion. Nguyen (1977) presented a formulation of the equilibrium based OD matrix estimation problem for congested network and analyzed the properties of the solution. Jörnsten and Nguyen (1979) and LeBlanc and Farhangian (1982) formulated the entropy maximizing and minimum least squares models with the equilibrium assignments; while Jörnsten Nguyen's model does not require a prior OD matrix, LeBlanc and Farhangian's method does.

Maximum Likelihood (ML) is the first type of statistical inference approach reviewed. ML approach maximizes the likelihood of observing the prior OD matrix $\hat{\mathbf{q}}$ and the observed link traffic counts $\hat{\mathbf{v}}$ conditional on the estimated OD matrix $\mathbf{q}$. With the assumption that the prior OD $\hat{\mathbf{q}}$ is statistically independent of link flow observation $\hat{\mathbf{v}}$, the likelihood of $\hat{\mathbf{q}}$ and $\hat{\mathbf{v}}$ is:

$$L(\hat{\mathbf{q}}, \hat{\mathbf{v}} \mid \mathbf{q}) = L(\hat{\mathbf{q}} \mid \mathbf{q}) \cdot L(\hat{\mathbf{v}} \mid \mathbf{q}) \qquad (2.9)$$

By assuming the Poisson probability distribution for the OD priors and traffic count observations, Spiess (1987) provided a formulation using ML approach to maximize the likelihood $L(\hat{\mathbf{q}}, \hat{\mathbf{v}} \mid \mathbf{q})$ with the assumption that the proportion of flow on each link that is associated with each OD is known. The proposed optimization problem was solved by a cyclic coordinate ascent algorithm and tested on small examples.

Generalized-least-squares (GLS) is another type of statistical estimator. The prior OD $\hat{\mathbf{q}}$ is assumed to have a random perception error $\boldsymbol{\eta}$ with respect to the true OD trips $\mathbf{q}$ as defined in Eq. (2.10), and traffic count observation $\hat{\mathbf{v}}$ is assumed including a random measurement error $\boldsymbol{\varepsilon}$ as in Eq. (2.11).

$$\hat{\mathbf{q}} = \mathbf{q} + \boldsymbol{\eta} \qquad (2.10)$$

$$\hat{\mathbf{v}} = \mathbf{v}(\mathbf{q}) + \boldsymbol{\varepsilon} \qquad (2.11)$$

One advantage of GLS approach is that no distributional assumptions need to be made for the error terms $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$, but assumes zero means and known dispersion matrices $\mathbf{Z}$ and $\mathbf{W}$ respectively. Assuming $\hat{\mathbf{q}}$ and $\hat{\mathbf{v}}$ are mutually independent, the GLS estimator can be obtained by solving the following optimization problem:

$$\min \frac{1}{2}(\hat{\mathbf{q}} - \mathbf{q})'\mathbf{Z}^{-1}(\hat{\mathbf{q}} - \mathbf{q}) + \frac{1}{2}(\hat{\mathbf{v}} - \mathbf{v}(\mathbf{q}))'\mathbf{W}^{-1}(\hat{\mathbf{v}} - \mathbf{v}(\mathbf{q})) \qquad (2.12)$$

$$s.t. \ \mathbf{q} \geq 0 \qquad (2.13)$$

Cascetta (1984) developed expressions for the mean and variance of the GLS estimator when non-negativity constraints in Eq. (2.13) are not active. GLS approach combines the two sources of information from observation and priors through the dispersion matrices $\mathbf{Z}$ and $\mathbf{W}$. If either dispersion matrix is close to zero, this reflects a great belief in this part of the information because the matrix inverse ($\mathbf{Z}^{-1}$ or $\mathbf{W}^{-1}$) wights this part highly in objective function. Bell (1991) derived the solution to GLS estimator by considering non-negative constraints in Eq. (2.13). In both approaches by Cascetta (1984) and Bell (1991), proportional assignment is assumed for the link flow patterns.

Yang et al. (1992) extended the GLS model by formulating a bilevel programming to integrate the GLS model on the upper level and the equilibrium assignment problem on the lower level. A heuristic algorithm was applied to the small problems for testing.

Bayesian inference approach is the statistical inference approach that considers the prior OD knowledge as a prior probability function $f(\mathbf{q})$, and the observed link flow counts given OD estimates as a likelihood $L(\hat{\mathbf{v}} | \mathbf{q})$. Then the posterior probability of OD matrices conditional on observed link flows can be derived by combining the prior distribution of OD volumes and the observations using Bayes theorem:

$$f(\mathbf{q} | \hat{\mathbf{v}}) \propto L(\hat{\mathbf{v}} | \mathbf{q}) \cdot f(\mathbf{q}) \quad (2.14)$$

Maher (1983) assumed that the link flow observations follow the Multivariate Normal (*MVN*) distribution. He also assumed the prior distribution $f(\mathbf{q})$ as *MVN* with the OD trips/links proportions known. Hence the posterior OD information is *MVN* distributed and the updating equations for posterior mean and variance can be derived from Bayes' Theorem.

Lo et al. (1996) formulated the statistical models and developed the corresponding Bayesian estimator by assuming that the link flow observations are independent Poisson random variables and the OD trips/links proportions are random variables as well. With the fixed routing assumption (there is only one given route between each OD pair), Tebaldi and West (1998) presented a Bayesian approach to infer the independent-Poisson-distributed OD flows based on observed counts on all links in the network. The posterior distributions for OD flows were obtained by an iterative simulation method, Markov Chain Monte Carlo (MCMC).

Hazelton (2000) derived the joint distribution of the link flows and the full likelihood distribution under the standard assumption of Poisson distributed number of OD flows. After approximating this likelihood function to *MVN*, which is more mathematically tractable, the posterior OD flows and route choice probabilities were updated by Bayesian estimator.

If the travel demand to be estimated is assumed varying over time, generic formulation of estimating the time-dependent OD demand is an extended model from the general formulation in Eq. (2.7) and Eq. (2.8), where the DTA assumption is formulated in constraint Eq. (2.8). Zhou (2004) provided a thorough review for the state-of-the-art research on dynamic OD demand estimation.

2.4 Sensor Location Models

As discussed in section 2.3, if the number of noisy-free sensors to be used is sufficient and the data-to-flow matrix **H** is full rank, the flow monitoring problem is a flow-observability problem; otherwise a flow-estimation problem must be solved to estimate route flows. In particular, following two classes of sensor location problems arise (Gentilli and Mirchandani, 2012):

(1) *The Sensor Location Flow-Observability Problems*: identify the optimum location of sensors on the network that allows the unique determination of the solution of the linear system of equations associated with the located sensors.

(2) *The Sensor Location Flow-Estimation Problems (SLFE)*: identify the optimum location of sensors on the network to best improve the quality of the related estimates (OD trips estimates, link flows estimates, route flows estimates, etc.)

that can be obtained using the system of linear equations associated with the located sensors.

With the rationalization of good flow estimation, most sensor location models that have been reported in the literature are related to some type of covering problems, which are simply proxies for estimation optimization. These will be reviewed next. The notation used in these formulations is described below:

Parameters:

R: Route-choice set for a network

A: Set of all the links in the network

OD: Set of ODs in a network

N: Number of sensors to locate

$\rho_a^i \in (0,1)$: Link-route parameter. $\rho_a^i = 1$ if route *i* uses link *a*, otherwise, $\rho_a^i = 0$

$\pi_a^j \in (0,1)$: Link-OD parameter. $\pi_a^j = 1$ if the OD pair *j* uses link *a*, otherwise, $\pi_a^j = 0$

$\hat{q}^j$: Prior mean of flow on OD pair *j*

$\xi_j^{(0)}$: Prior variance of flow on OD pair *j*

$\xi_j^{(1)}(\mathbf{y})$: Posterior variance of flow on OD pair *j* using the link count observations from sensors' location **y**

$\mu_i^{(0)}$: Prior mean of flow on route *i*

$\eta_i^{(0)}$: Prior variance of flow on route *i*

$\eta_i^{(1)}(\mathbf{y})$: Posterior variance of flow on route *i* using the link count observations from sensor allocation **y**

$\tilde{v}_a$: Prior mean of flow on link *a;* $\tilde{v}_a = \sum_{i \in R} \rho_a^i \mu_i^{(0)}$

Decision Variables:

$y_a \in (0,1)$: $y_a = 1$ if locating sensor on link *a*, otherwise $y_a = 0$

$\psi^i \in (0,1)$: $\psi^i = 1$ if that route *i* is covered by a located sensor, otherwise $\psi^i = 0$

$o^j \in (0,1)$: $o^j = 1$ if that OD pair *j* is covered by a located sensor, otherwise $o^j = 0$

(1) Link flow coverage model

The link flow coverage (LFC) method intends to locate sensors so that link flow volume covered (or intercepted) is maximized (Lam and Lo, 1990). By using this method, the link can be ranked in descending order of the prior link flows $\tilde{v}_a$ on each link *a*. The location model can be formulated by the following integer programming program:

[LFC]

$$\max \sum_{a \in A} \tilde{v}_a y_a \quad (2.15)$$

$$s.t. \quad \sum_{a \in A} y_a = N \quad (2.16)$$

$$y_a \in \{0,1\}, \forall a \in A \quad (2.17)$$

The objective function Eq. (2.15) maximizes the total link flow volumes that will be observed by sensors. The constraint Eq. (2.16) is the budget constraint. The direct objective of LFC method is to cover links with large traffic flows, but not the quality of OD estimates.

(2) OD-pair coverage model

OD-pair coverage (ODC) method tries to locate sensors to maximize the number of OD pairs covered, where an OD pair is "covered" if at least one route in its choice set is observed. The model can be formulated as the following integer programming program:

[ODC_1] $$\max \sum_{j \in OD} o^j \quad (2.18)$$

$$s.t. \quad \sum_{a \in A} \pi_a^j y_a \geq o^j, \forall j \in OD \quad (2.19)$$

$$\sum_{a \in A} y_a = N \quad (2.20)$$

$$y_a \in \{0,1\}, \forall a \in A \quad (2.21)$$

$$o^j \in \{0,1\}, \forall j \in OD \quad (2.22)$$

Given the budget constraint Eq. (2.20), the objective function Eq. (2.18) and constraint Eq. (2.19) together maximize the total number of OD pairs covered by sensors. An inverse formulation [ODC_2] is to minimize the number of sensors used in Eq. (2.23) given that each OD pair has to be covered by at least one sensor, as formulated by constraint Eq. (2.24):

[ODC_2] $$\min \sum_{a \in A} y_a \quad (2.23)$$

$$s.t. \quad \sum_{a \in A} \pi_a^j y_a \geq 1, \forall j \in OD \quad (2.24)$$

$$y_a \in \{0,1\}, \forall a \in A \quad (2.25)$$

The first model belonging to this category was proposed by Lam and Lo (1990). Yang et al. (1991) extended the model by defining the number of available sensors N in budget constraint Eq. (2.20) as a variable, instead of a parameter. Yang and Zhou (1998) proposed an ODC model where only the links carrying maximum fraction of OD demands are eligible in the coverage constraint Eq. (2.24). Ehlert et al. (2006) modified the formulation [ODC_2] by assigning costs to candidate links. Chootinan et al. (2005) proposed a two-objective formulation for the tradeoff between the coverage of OD-pairs (quality) and the number of link flow detectors being used (cost). A genetic algorithm was used to generate non-dominated solutions. Gan and Yang (2001) and Gan et al. (2005) provided the screen-line based traffic counting location models to select the optimal locations for a given number of traffic counting stations to separate as many O-D pairs as possible. Chen et al. (2007) investigated the effect of locating additional sensors to augment OD coverage.

(3) Route cardinality coverage model

Route cardinality coverage (RCC) model locates sensors to maximize the number of routes covered (or intercepted). The model can be formulated as the following integer programming program:

[RCC_1] $$\max \sum_{i\in R} \psi^i \tag{2.26}$$

$$s.t.\quad \sum_{a\in A} \rho_a^i y_a \geq \psi^i, \forall i \in R \tag{2.27}$$

$$\sum_{a\in A} y_a = N \tag{2.28}$$

$$y_a \in \{0,1\}, \forall a \in A \tag{2.29}$$

$$\psi^i \in \{0,1\}, \forall i \in R \tag{2.30}$$

In formulation [RCC_1], the objective function Eq. (2.26) and constraint Eq. (2.27) together maximize the total number of routes covered by sensors given the budget constraint Eq. (2.28). An inverse formulation [RCC_2] is to minimize the number of sensors used given that each route has to be covered by at least one sensor:

[RCC_2] $$\min \sum_{a\in A} y_a \tag{2.31}$$

$$s.t.\ \sum_{a\in A}\rho_a^i y_a \geq 1, \forall i \in R \quad (2.32)$$

$$y_a \in \{0,1\}, \forall a \in A \quad (2.33)$$

Gentili and Mirchandani (2005) and Castillo et al. (2008) both considered the case that all routes should be covered by using a minimum number of sensors similar to formulation [ODC_2] regardless of flows on the routes.

(4) OD-demand coverage model

OD-demand coverage (ODDC) method locates sensors to maximize total OD demands coverage. The basic formulation of [ODDC] is shown as:

[ODDC]
$$\max \sum_{j\in OD}\hat{q}^j o^j \quad (2.34)$$

$$s.t.\ \sum_{a\in A}\pi_a^j y_a \geq o^j, \forall j \in OD \quad (2.35)$$

$$\sum_{a\in A} y_a = N \quad (2.36)$$

$$y_a \in \{0,1\}, \forall a \in A \quad (2.37)$$

$$o^j \in \{0,1\}, \forall j \in OD \quad (2.38)$$

The objective function Eq. (2.34), together with constraint Eq. (2.35), maximizes the total demand captured in the network given the budget constraint Eq. (2.36). [ODDC] is a weighted version of [ODC_1]—the weights are the prior OD flow volumes.

Hodgson (1990) proposed the first model of this type to locate the sensors on the nodes, instead of links. Yim and Lam (1998) proposed the model to select links according to the contribution of covering an OD demand.

(5) Route flow coverage model

Route flow coverage (RFC) method intends to locate sensors to maximize the total route flow intercepted. Similar to ODDC, the RFC formulation is a weighted version of RCC, while the weights are the prior route flows. The formulation is:

[RFC] $$\max \sum_{i \in R} \mu_i^{(0)} \psi^i \quad (2.39)$$

$$s.t. \quad \sum_{a \in A} \rho_a^i y_a \geq \psi^i, \forall i \in R \quad (2.40)$$

$$\sum_{a \in A} y_a = N \quad (2.41)$$

$$y_a \in \{0,1\}, \forall a \in A \quad (2.42)$$

$$\psi^i \in \{0,1\}, \forall i \in R \quad (2.43)$$

The maximal flow-intercepting rule proposed by Yang and Zhou (1998) belongs to this category because it requires that the links should be chosen as to intercept as many route flows as possible. Ehlert et al. (2006) extended the formulation by additional constraint that requires that all OD-pairs should be covered.

(6) OD demand variance reduction model

OD demand variance reduction (ODVR) method tries to obtain the sensor allocation **y** by maximizing the variance reduction for the OD demand estimates. The general formulation is:

[ODVR] $$\max \sum_{j \in OD} (\xi_j^{(1)}(\mathbf{y}) - \xi_j^{(0)}) \quad (2.44)$$

$$s.t. \quad \sum_{a \in A} y_a = N \quad (2.45)$$

$$y_a \in \{0,1\}, \forall a \in A \quad (2.46)$$

The objective function Eq. (2.44) maximizes the total variance reduction in OD estimates by using different sensors allocation **y**. This objective function is equivalent as minimizing the total variances in OD estimates $\sum_{j \in OD} (\xi_j^{(1)}(\mathbf{y}))$. The posterior variance $\xi_j^{(1)}(\mathbf{y})$ can be obtained by the OD estimation models using sensor location **y** as the input.

Zhou and List (2010) proposed a sensor location model to determine the location of both counting sensors and AVI sensors, the objective of which is to minimize the posterior variance of OD demands. In their model, $\xi_j^{(1)}(\mathbf{y})$ is obtained from a Kalman filter approach using the observations from sensors' location **y**. A scenario-based stochastic optimization procedure and a beam search algorithm were developed to find the solution.

(7) Route flow variance reduction model

Route flow variance reduction (RVR) method tries to obtain the sensor allocation **y** by maximizing the variance reduction for the OD demand estimates. The general formulation is:

[RVR]
$$\max \sum_{i \in R} (\eta_i^{(1)}(\mathbf{y}) - \eta_i^{(0)}) \quad (2.47)$$

$$s.t. \quad \sum_{a \in A} y_a = N \quad (2.48)$$

$$y_a \in \{0,1\}, \forall a \in A \quad (2.49)$$

RVR model uses the quality of flow estimate as the direct objective in the formulation. Both RVR and ODVR models can be used in the upper level of *SLFE* problem and the posterior variance $\eta_i^{(1)}(\mathbf{y})$ is calculated from a specific estimator in the lower level of *SLFE*.

Wang et al. (2012) and Wang and Mirchandani (2013) proposed the first two models in this category. The first model assumed a special situation that noise-free or “perfect” sensors are used, and the second model assumed the counting devices are “noisy” sensors introducing random measurement errors. These are described in detail in Chapters 3 and 4 to follow.

The seven types of models described above rely on different sources of information. The information includes the network structures and assignment (network topology, link choice set, etc.), average values (point values) of the flows of interest, and the reliability (variance) of the information. ODC and RCC models depend on the information of network structure and link choice set only; LFC model employs the point value of link flow priors; ODDC and RFC models require both network structure and the point value (average) of prior flow volumes for the flow of interest; ODVC and RVD models require the reliabilities of the prior information in addition. The relationship between the seven categories, in terms of the information resources by each formulation, is illustrated in Figure 5. The outer corner defines the boundary of all types of available information. Each corner represents one type of flows in priors: the upper corner is the link flows, the lower left corner is the OD flows and the lower right corner is the route flows. The inner circle indicates the prior information of different type. Hence, ODC and RCC should be placed in Figure 5 in the sets of difference between the outer triangle and inner circle (this area includes network topologies and link choice sets only). The inner triangle specifies the usage of reliabilities as available information. Therefore, LFC, ODDC and RFC models fill the areas within the inner circle but out of inner triangle in Figure 5 because they do not rely on the reliability of information. ODVR and RVR models require all information in the problem space and fill the areas within the inner triangle in Figure 5. The difference between ODVR and RVR is the traffic assignment assumption for the demand pattern. ODVR usually requires the equilibrium assignment between each OD, while RVR allow more generalized flow patterns in the network. This dissertation focuses on RVR model.

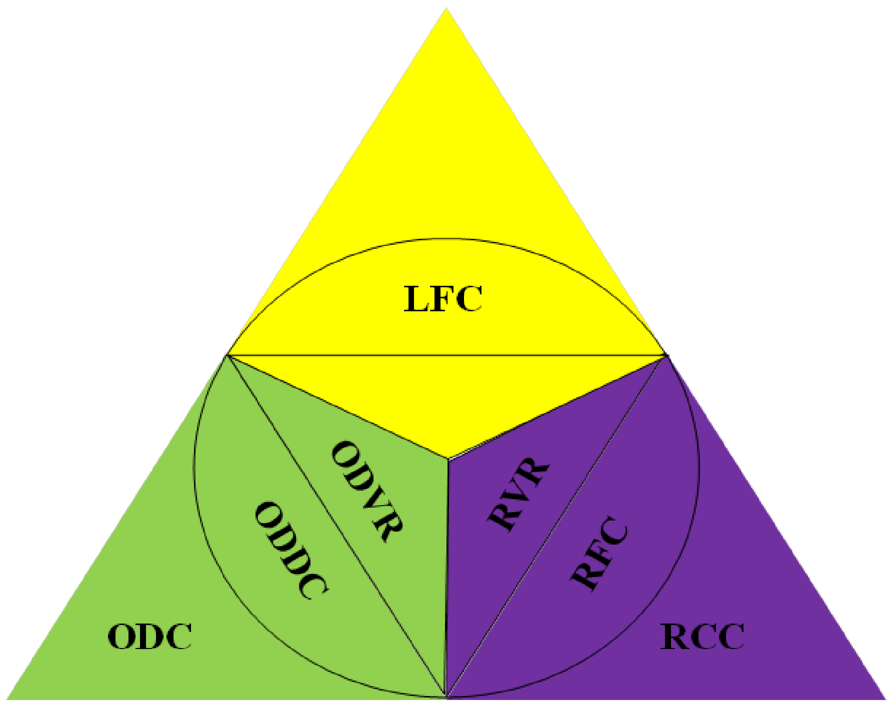

Figure 5 Relationship among the seven categories of existing sensor location models

Gentili and Mirchandani (2011, 2012) first define the *SLFE* problem as a bilevel problem, where the upper level decides the optimum subset of links where sensors are located, and the lower level computes the best estimates of the flows of interest by using anticipated data from given candidate sensors location. They also have surveyed sensor location models and discussed "rules" or objectives for location models that appeared in the literature. They observed that most existing location models (other than ODVR and RVR) focus on the location decisions that use proxies for OD estimation quality but do not explicitly attempt to minimize an estimation error. The proxies (such as maximizing

OD flow coverage) are heuristic evaluation criteria that capture only indirectly the quality of OD estimates.

Larsson et al. (2010) also reviewed the existing models for sensor location problem. They conducted computational experiments to compare the performances of sensor location models in terms of the quality of OD estimates using different link allocation by assuming the equilibrium demand patterns.

Chapter 3

LOCATING SENSORS TO ESTIMATE ROUTE FLOWS USING DATA FROM NOISE-FREE SENSORS AND ASSUMING AN UNCERTAINTY INTERVAL FOR PRIOR MEANS

3.1 Problem Description

In this chapter, the flow of interest to be estimated is the deterministic OD demand with the traffic assignment assumption of a known route choice set. This chapter first introduces a route flow estimation model that will be used to obtain a good point estimate of flows on all the routes in a traffic network. The route flow estimates represent the mean or average of the flow volumes. In the estimation stage, the route flow priors and observations from link sensors are available. In this chapter, a route flow prior is assumed known as a mean $\tilde{x}^k$ and a confidence level $\tilde{u}^k$ for any route *k*, which represent the magnitude and reliability of the prior information for route *k*. The counting sensors used in this chapter are assumed as "perfect sensors", which means the data collected from the counting sensors are error free. The second stage of this chapter is to derive a model for the location decision of limited number of sensors — the observations from which are used in the estimation model in estimation stage. The objective of the second stage is to locate sensors on a subset of links to maximize the reduction of uncertainties in route flow estimates. A new integer programming model is developed to solve this location problem.

3.2 Notation

Parameters:

R: Route-choice set for a network

A: Set of all the links

N: Number of sensors to locate

M: A large number

$\rho_a^k \in (0,1)$: Link-route parameter. $\rho_a^k = 1$ means that route *k* uses link *a*, otherwise $\rho_a^k = 0$

$\tilde{x}^k$: Route flow prior of k^{th} route

$\tilde{u}^k$: Confidence level (max % error) on route flow prior of k^{th} route

$\tilde{v}_a$: Link flow prior on link *a*

v_a : Link flow observation on link *a*

Decision Variables:

$z_a \in (0,1) : z_a = 1$ represents locating sensor on link *a* otherwise $z_a = 0$.

x^k : Actual route flow of k^{th} route (variable in ESTR model)

$\hat{x}^k$: Flow estimate on k^{th} route (solution of ESTR model)

ε^k : percentage estimation error ($\leq \tilde{u}^k$) on route flow prior of k^{th} route

$\hat{v}_a$: Link flow estimate on link *a*

Among these parameters, ρ_a^k, $\tilde{x}^k$, $\tilde{u}^k$ and $\tilde{v}_a$ relate to prior information. The first parameter ρ_a^k are obtained from definitions of the given route choice sets. The latter three, $\tilde{x}^k$, $\tilde{u}^k$ and $\tilde{v}_a$ are from models' knowledge to the prior route flows. As discussed earlier, prior information can be derived from survey data or historical data, or derived using some reasonable traffic loading models. Such information on flows usually has errors compared to the true information. As in most prior-to-posterior estimation schemes, they are used to obtain an estimate for each flow variable among possibly infinite feasible solutions. In our computational laboratory experiments, errors in $\tilde{x}^k$ are randomly generated. $\tilde{v}_a$ is simply obtained from ρ_a^k and $\tilde{x}^k$. The following expression represents the relationship between route flows and link flows:

$$\sum_{k \in R} \rho_a^k x^k = v_a \tag{3.1}$$

The equation indicates that the flow on a link is equal to the sum of all route flows that use the link.

The crux of the chapter's thesis is the Bayesian thinking that the planners have different confidence levels on mean values of the data-based parameters such as a prior mean on a route flow. In the model to be presented, the level uncertainty of the prior means is represented by a confidence interval $[-\tilde{u}, \tilde{u}]$. If the confidence interval is small then it means that the decision-maker has very high confidence in the given prior mean. Vice versa, if the confidence interval is large then the decision-maker is not too confident about the reliability of the given prior. In general, this uncertainty can be based on one or more of these factors: (a) historical data (empirical uncertainty); (b) planner's subjective reliability, the kind of uncertainties used in decision making when the situation is new and there is no empirical information, and (c) physical reasoning, (e.g., the probability we will get a six with a toss of a fair die is 1/6). In the context of our estimation problem, we anticipate that the uncertainties will correspond to the planners' subjective reliabilities that are based mostly on empirical evidences but tempered by anecdotal observations on the numbers they have before them.

3.3 Route Flow Estimation Model

Once the sensors are located, the procedure of estimating flow variables from observations is analogous to obtaining a posteriori knowledge based on observation evidence and prior knowledge. If the flows of interest are OD volumes, the corresponding problem is to estimate an OD matrix so that the induced link flows are as close as possible to the observed link flows from the counting sensors. With the assumption of given route flow priors, the problem of finding best route flow estimates from link flow sensor observations can be formulated as:

[ESTR] $$\min \sum_{k \in R} \left(\frac{x^k - \tilde{x}^k}{\tilde{x}^k} \right)^2 \quad (3.2)$$

$$s.t \quad \sum_{k \in R} \rho_a^k x^k + M(1 - z_a) \geq v_a, \forall a \in A \quad (3.3)$$

$$\sum_{k \in R} \rho_a^k x^k - M(1 - z_a) \leq v_a, \forall a \in A \quad (3.4)$$

$$x^k \geq 0, \forall k \in R \quad (3.5)$$

The solution of the optimization problem ESTR gives the best route flow estimates $\hat{x}^k$ and the corresponding link flow estimates $\hat{v}_a$.

In this model, the continuous variables are the route flows x^k. The objective function Eq. (3.2) attempts to minimize the total estimation error, where each component error is the relative distance between an estimated route flow to its route flow prior. Objective value in Eq. (3.2) is the sum of squared relative errors, where the relative error is the fraction of deviation of route flow estimate from its prior. When $z_a = 1$ (that is, when a sensor is located on link *a*), constraints Eq. (3.3) and Eq. (3.4) together force Eq. (3.1) which relates observed link flows to estimated route flows. When $z_a = 0$ there is no restriction for link flow estimates on that link, which means this link flow estimate has no information coming from the flows on this link since this link is not directly observed. That is, in this model, when link a is observed then link flow estimate $\hat{v}_a$ is equal to the link flow observed (v_a), while unobserved link flows are computed from the estimated route flows. The idea of ESTR is to estimate route flows close to the priors, under the constraint that route flow estimates are consistent with observed link flows. Note that ESTR assumes location decisions, described by variables z_a, are known; this model is assumed to be used after the locational decisions are made. In subsection 3.4, we will use this ESTR formulation to develop the sensor location model RVR-perfect.

3.4 New Location Model Formulation

Since RVR-perfect aims to locate sensors to obtain the best possible route flows estimates, we try to take into account ESTR to formulate the sensor location model. Because future link flow observations from sensors (v_a) in ESTR cannot be obtained before sensors are located, this observation information is not available for sensor location decisions. However, prior flows are available information for location decision making. In particular, known are route flow priors with their reliabilities, and corresponding link flow priors that are consistent with Eq. (3.1).

We assume that along with the route priors we have an idea of their reliability in terms of a confidence interval. For example one may have a good prior and say that the flow on a route k is $x^k \pm$ 1% error, while a bad prior may be $x^k \pm$ 15% error. This uncertainty will be explicitly taken into account in the model development below.

Suppose prior flow for route k is $\tilde{x}^k$ then we can state $\tilde{x}^k = x^k(1+\varepsilon^k)$ where ε^k comes from a range $[-\tilde{u}^k, \tilde{u}^k]$; that is, $\tilde{u}^k \in (0,1)$ represents our uncertainty in the route prior.

Locating a sensor on a link gives us the actual flow on that link, and hence it reduces the sum of the uncertainties of the route flows that use that link. In particular, we have

$$\tilde{v}_a = \sum_{k\in R} \rho_a^k \tilde{x}^k = \sum_{k\in R} \rho_a^k x^k + \sum_{k\in R} \rho_a^k x^k \varepsilon^k = \sum_{k\in R} \rho_a^k x^k + \sum_{k\in R} \rho_a^k \tilde{x}^k \frac{\varepsilon^k}{1+\varepsilon^k} \quad (3.6)$$

and locating on link a makes the second term on the right hand side of Eq. (3.6) going to zero. Hence, the problem becomes to locate sensors to maximize the reduction of uncertainty and thus RVR-perfect formulates as:

$$\max \sum_{a\in A} \left(\sum_{k\in R} \rho_a^k \tilde{x}^k \left| \frac{\varepsilon^k}{1+\varepsilon^k} \right| \right) z_a \quad (3.7)$$

$$s.t. \quad \sum_{a\in A} z_a = N \quad (3.8)$$

$$-\tilde{u}^k \le \varepsilon^k \le \tilde{u}^k, \forall k \in R \quad (3.9)$$

$$z_a \in \{0,1\}, \forall a \in A \quad (3.10)$$

Since we are maximizing, each term of Eq. (3.7) is maximized by letting $\varepsilon^k = -\tilde{u}^k$ because the value of $\tilde{u}^k$ is bounded by the expression in Eq. (3.9) and $\tilde{u}^k \in (0,1)$ as we assumed. Then we can simplify the formulation to

$$\text{[RVR-perfect]} \quad \max \sum_{a\in A} \left(\sum_{k\in R} \rho_a^k \tilde{x}^k \frac{\tilde{u}^k}{1-\tilde{u}^k} \right) z_a \quad (3.11)$$

$$s.t. \quad \sum_{a\in A} z_a = N \quad (3.12)$$

$$z_a \in \{0,1\}, \forall a \in A \quad (3.13)$$

This formulation is quite similar to LFC except the weight for link a is not $\tilde{v}_a$ but the measures of uncertainties of the route flows that use the link. This is a linear-integer programming formulation for the RVR-perfect problem whose solution gives an optimal location set for each sets of priors $\tilde{x}^k, \tilde{u}^k$. Constraint Eq. (3.12) is the limitation on the number of sensors to be located. Eq. (3.13) is the binary constraint for location variables z_a.

3.5 Algorithm

A greedy algorithm can solve the above model by first calculating $\sum_{k\in R} \rho_a^k \tilde{x}^k \frac{\tilde{u}^k}{1-\tilde{u}^k}$ for each link a, then sorting the calculated values and selecting the N links with largest values.

The algorithm for solving [RVR-perfect] is:

GREEDY ALGORITHM TO SOLVE [RVR_PERFECT]

Input: $\tilde{x}^k, \tilde{u}^k, \rho_a^k$, N

Output: z_a

(1) For each link $a \in A$, calculate $\sum_{k\in R} \rho_a^k \tilde{x}^k \frac{\tilde{u}^k}{1-\tilde{u}^k}$.

(2) Sort all links in set A according to the descending order of the values calculated in step (1).

(3) Select the first N links with largest values calculated in step (1).

The running time for this algorithm is polynomial because step (1) and (3) can be conducted in constant time and the run time of step (2) depends on the sorting algorithm (e.g. Quicksort in $\mathcal{O}(n^2)$ and merge sort in $\mathcal{O}(n \log n)$).

3.6 Experimental Results

3.6.1 Experiment Setup

Figure 6 gives the scheme of the experimental procedure. First (Steps 1-2), problem scenarios are generated by defining the network (network supply), loading the assumed traffic flows (demand) and obtaining the “actual” network traffic. Assumed prior route and link flows are generated by relating them to actual flows plus some perception errors (Step 3). The perception error for a route flow prior is obtained by first generating reliability for each route k denoted by $\tilde{u}^k \in (0, u^{\max})$; and then generating a random error in range $[-\tilde{u}^k, \tilde{u}^k]$ and adding it to the actual route flow. Link flow priors are then computed using Eq. (3.1). In Step 4, RVR-perfect is solved by the algorithm in section 3.5 and other location models are solved using *Cplex* to obtain location decisions z_a. Then, fixing these decisions, estimates for route and link flows are obtained using the estimation model ESTR (Step5). To evaluate each location model, the route estimates are compared with actual route flows (Step 6).

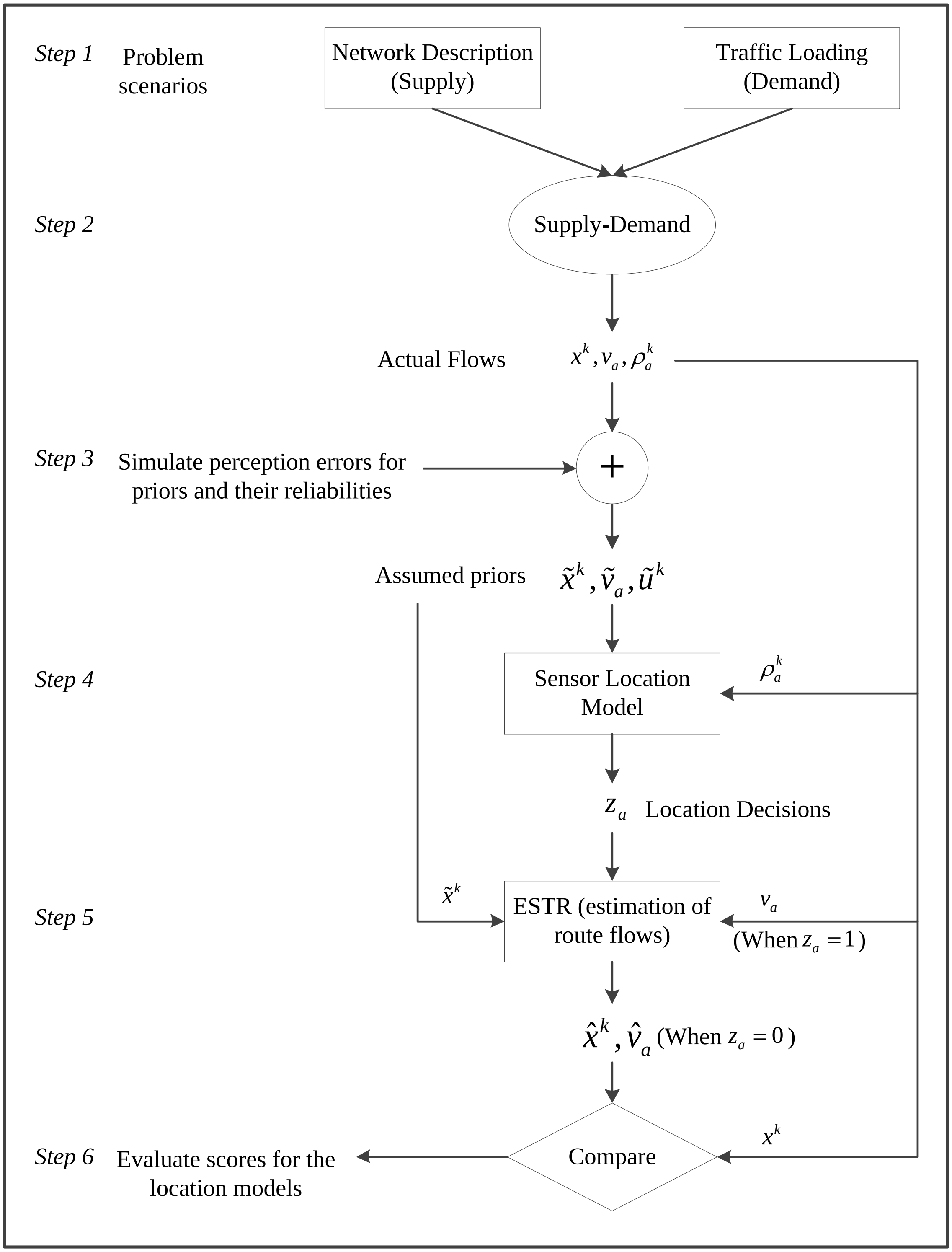

Figure 6 Procedure for the computational experiments for RVR-perfect

When evaluating the quality of route flow estimates, the following Relative Mean Absolute Error (RMAE) criterion was used:

$$RMAE = \frac{1}{|R|}\sum_{k \in R}\left|\frac{\hat{x}^k - x^k}{x^k}\right| \cdot 100\% \qquad (3.14)$$

where $|R|$ is the number of routes in the given scenario. Hence, for a problem instance, four RMAEs are calculated in Step 6 one for each location model: RVR-perfect, LFC, RCC and RFC.

Finally, note that for each network scenario, we can have different problem instances by locating different number of sensors between 1 and (say) *m*, the number of possible sensor sites. Furthermore, we can have different priors, ranging from quite accurate priors with errors less than 5% to less accurate priors with errors ranging between 10% and 15%. Hence, each problem instance in our computational experiments is defined by:

(1) Network supply

(2) Traffic demand

(3) Reliability of priors on traffic flows to be estimated

(4) Number of sensors to be located.

Solving the location model for each problem instance, for the four models being compared, gives us a score for each model, as defined by the RMAE.

Three sets of network scenarios were used in the computational experiments.

(1) A grid network with 16 OD pairs, 43 routes and 48 links (grid network 1).

(2) A grid network with 16 OD pairs, 204 routes and 112 links (grid network 2).

(3) Houston data used by Mirchandani et al. (2009), with 1768 routes and 468 links with flows.

The network topologies for grid network 1 and 2 are displayed in Figure 7.

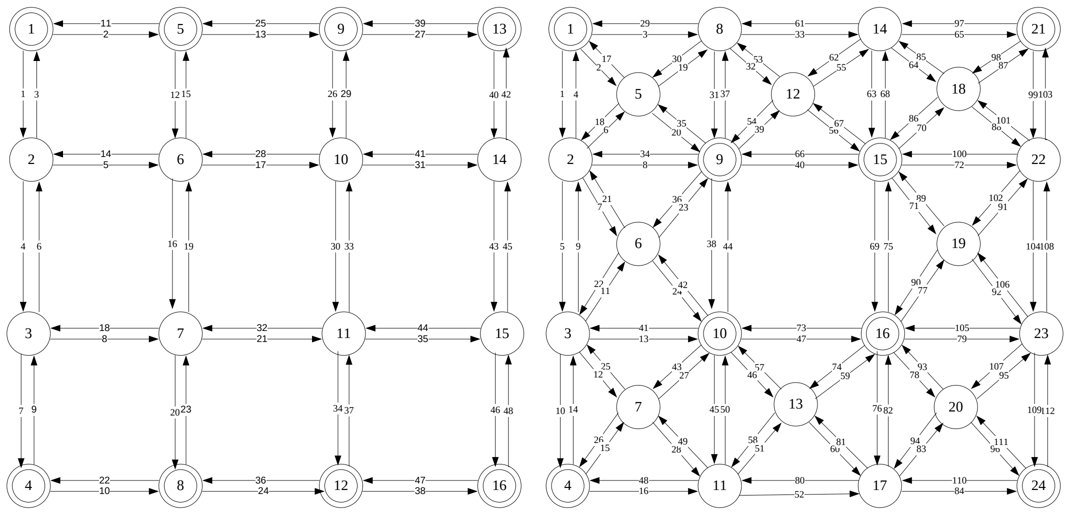

Figure 7 Topologies of experiment networks

For all the three scenarios, OD demands are loaded by a traffic equilibrium approach, to obtain near equilibrium actual route and link flows in step 2. (We reemphasize that the model and method developed here does not depend on equilibrium flows, but simply on a known route choice set.)

3.6.2 Numerical Results

Table 2 gives the RMAE results for the four location models (RVR-perfect, LFC, RCC, RFC) and three priors with different reliability ranges for grid network 1. For example, the columns marked with $u \in [0, 0.1]$ indicate that errors in priors range from 0%-10%. The table gives the averaged RMAE scores for every five instances for five numbers of sensor locations. For example, the first row in Table 2 represents the average RMAE scores for N=1 to N=5.

The best scores are in italicized and underlined in Table 2. In general, our proposed RVR-perfect outperforms other models in most cases. Note that, on the instances corresponding to values of N from 36 to 48 all models perform the same since most of the links are now with sensors and all models have the same locations and estimation errors. Since in real cases the decision-makers are interested in locating sensors only on few links due to budget considerations, our study focus is to monitor fewer than half of the links and compare the models for these ranges. The last row of Table 2 sums all of the average RMSE scores and emphasizes the advantage of RVR-perfect compared to other models for grid network 1. Observe, however, that LFC model sometimes has the best score, and often is the second best when RVR-perfect is the best. Nevertheless as it will be discussed later, RVR-perfect statistically significantly outperforms LFC.

Table 2 RMAE Results using RVR-perfect for Grid Network 1

	$\tilde{u}^k \in [0, 0.05]$				$\tilde{u}^k \in [0, 0.1]$				$\tilde{u}^k \in [0, 0.15]$			
N	RVR	LFC	RCC	RFC	RVR	LFC	RCC	RFC	RVR	LFC	RCC	RFC
1-5	0.90	***0.89***	0.94	0.94	2.30	2.31	***2.28***	2.32	***3.39***	3.44	3.50	3.38
6-10	0.85	***0.83***	0.84	0.84	2.08	2.14	***2.06***	2.08	3.14	***3.11***	3.25	3.19
11-15	***0.74***	0.75	0.80	0.80	***1.82***	2.00	2.00	2.03	3.01	3.17	***2.94***	2.98
16-20	0.63	0.67	***0.62***	***0.62***	***1.68***	1.77	1.89	1.83	***2.72***	2.78	2.73	2.74
21-25	***0.59***	0.60	0.60	0.60	1.56	***1.44***	1.73	1.73	***2.36***	2.49	2.60	2.60
26-30	***0.57***	0.59	0.60	0.60	1.30	***1.26***	1.70	1.70	***2.20***	2.26	2.59	2.59
31-35	0.53	***0.52***	0.57	0.57	***1.26***	***1.26***	1.52	1.52	***2.14***	***2.14***	2.41	2.41
36-48	***0.52***	***0.52***	***0.52***	***0.52***	***1.26***	***1.26***	***1.26***	***1.26***	***2.14***	***2.14***	***2.14***	***2.14***
Total	***5.32***	5.37	5.48	5.48	***13.25***	13.43	14.44	14.47	***21.09***	21.54	22.14	22.02

We have similar comparisons among the RMAE scores for the four location models applied to the grid network 2 where there are many more routes for each OD pair. The average RMAEs are shown in Table 3. The RVR-perfect model performs the best in most cases and on the overall performance as indicated in the last row.

Table 3 RMAE Results using RVR-perfect for Grid Network 2

	$\tilde{u}^k \in [0, 0.05]$				$\tilde{u}^k \in [0, 0.1]$				$\tilde{u}^k \in [0, 0.15]$			
N	RVR	LFC	RCC	RFC	RVR	LFC	RCC	RFC	RVR	LFC	RCC	RFC
1-20	***1.31***	1.34	1.33	1.34	2.35	2.36	***2.33***	2.34	3.59	3.59	3.58	***3.55***
21-40	***1.24***	1.28	1.30	1.30	***2.13***	2.18	2.21	2.21	***3.35***	3.38	***3.35***	***3.35***
41-60	***1.19***	1.21	1.29	1.29	***2.07***	2.07	2.19	2.19	***3.26***	***3.26***	3.33	3.34
61-80	***1.16***	***1.16***	1.21	1.21	***2.05***	***2.05***	2.11	2.11	***3.19***	***3.19***	3.22	3.22
81-112	***1.16***	***1.16***	***1.16***	***1.16***	***2.05***	***2.05***	***2.05***	***2.05***	***3.18***	***3.18***	***3.18***	***3.18***
Total	***6.07***	6.16	6.29	6.30	***10.65***	10.72	10.89	10.91	***16.56***	16.59	16.65	16.62

Table 4 tabulates the problem instances when at most half the links can be detectorized, that is, average scores for all instances from *N*=1 to *N*=½ $|A|$. Table 4 shows that the average value of RMSE is always lower when using RVR-perfect.

Table 4 RMAE Results using RVR-perfect for Problem Instances for *N*=1 to *N*=1/2|*A*|

		RVR	LFC	RCC	RFC
Grid Net 1	prior1	***0.747***	0.754	0.765	0.765
	prior2	***1.907***	1.960	2.006	2.012
	prior3	***2.948***	3.029	3.019	2.993
Grid Net 2	prior1	***1.253***	1.284	1.310	1.312
	prior2	***2.190***	2.215	2.246	2.252
	prior3	***3.414***	3.420	3.427	3.418
Houston	prior1	***1.203***	1.204	1.205	1.204
	prior2	***2.448***	2.451	2.458	2.456
	prior3	***3.650***	3.655	3.658	3.659

Because of the random effects in the framework of the RVR-perfect problem, average scores is only one way to compare the performance. Further comparison was conducted using statistical paired-*t* tests to compare RVR-perfect scores with those of each of the other models. Here one is interested in testing if the mean of scores of RVR-perfect is significantly lower than the mean of scores of each of the other models. The hypothesis for the paired-*t* test is:

$$H_0 : \mu_{RVR\_p} \geq \mu_{others}$$

$$H_1 : \mu_{RVR\_p} < \mu_{others}$$

If the *P*-value for a paired-*t* test is lower than 0.05, it means we can reject H_0 at 95% confidence level and have a statistical indication that RVR-perfect is better than the compared model. For each network scenario, we conducted three similar paired-*t* tests to compare RMAE of RVR-perfect with the scores of LFC, RCC and RFC separately; the *p*-

values are shown in Table 5. The table shows that RVR-perfect significantly outperforms all the other models in all the three scenarios because the *p*-values of the paired-*t* test are essentially zero.

Table 5 *P*-values for Paired-*t* Tests Comparison of RMSE of RVR-perfect with Other Location Models

Alternative hypothesis: $\mu_{\text{RVR\_P}} <$	μ_{LFC}	μ_{RCC}	μ_{RFC}
Grid Net 1	0.000	0.000	0.000
Grid Net 2	0.000	0.000	0.000
Houston	0.000	0.000	0.000

3.7 Chapter Conclusions

In this chapter, a new linear integer programming model in the location stage of *SLFE* problem was developed and evaluated in terms of the quality of the point estimates of route flows. In this model, prior knowledge on routes and flows is explicitly modeled. As in many other models that have appeared in the literature, the route choice set is assumed to be known to generalize the equilibrium flow patterns. In this model we assume that we have both priors for the route flows and their reliabilities. In other words, for each route flow prior we have a confidence interval for the actual flow on the route.

The model was shown to be similar to the LFC (link flow coverage) that has appeared in the literature but the RVR-perfect model explicitly considers the level of uncertainties in the priors in the computation of the "weights" attached to prior link flows. Computational experiments for three sets of network scenarios were conducted. In the experiments described in the chapter, the LFC performed well but in most cases RVR-perfect performed better than LFC. The better performance of RVR-perfect compared to LFC is statistically significant with a *p*-value of 0.05. The evaluation showed that RVR-perfect also performed significantly better than the other two models RCC and RFC.

Chapter 4

LOCATING SENSORS TO ESTIMATE ROUTE FLOWS USING DATA FROM NOISY SENSORS AND ASSUMING NORMAL DISTRIBUTIONS FOR PRIOR UNCERTAINTIES

4.1 Introduction and Problem Description

4.1.1 Introduction to Bayes' Theorem and Bayesian Linear Estimation

Let $B_1, B_2, \ldots$ be a set of mutually exclusive and exhaustive events and let A be an event. Then, Bayes' Theorem, or referred as Bayes' Rule, is stated as, for each $i = 1, 2, \ldots,$

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{\infty} P(A \mid B_i)P(B_i)} \tag{4.1}$$

Eq. (4.1) can be interpreted in the following way. Starting with an initial prior probability $P(B_i)$ of the interested event B_i, when thinking of the B_i s as set of hypotheses, the occurrence of event B_i is equal to say that hypothesis i is true. The posterior probability $P(B_i \mid A)$ is the proper description of how likely hypothesis i is true when event A is known to have occurred. Observing event A changes the prior probability $P(B_i)$ to posterior probability $P(B_i \mid A)$. Notice that the posterior probabilities for all events sum to one. This is because all B_i are mutually exclusive, which result that one and only one hypothesis is true. The denominator in Eq. (4.1) is a weighted average of the probabilities $P(A \mid B_i)$ with the weights being the $P(B_i)$. The occurrence of A increases the probability of B_i if $P(A \mid B_i)$ is greater than the average of all the $P(A \mid B_i)$ s.

Bayes' Theorem can be restated in terms of random variables instead of events as:

$$f(\mathbf{x}|\mathbf{y}) = \frac{f(\mathbf{x})f(\mathbf{y}|\mathbf{x})}{\int f(\mathbf{x})f(\mathbf{y}|\mathbf{x})d\mathbf{x}} \qquad (4.2)$$

where **y** is the data and **x** is the parameter to inference about. The prior density $f(\mathbf{x})$ represents the prior information about parameter **x**. Bayes' Theorem constructs the posterior density $f(\mathbf{x}|\mathbf{y})$ as the proportional to the product of the prior density $f(\mathbf{x})$ and the likelihood $f(\mathbf{y}|\mathbf{x})$.

The Bayesian method comprises the following principal steps:

(1) Obtain the likelihood function $f(\mathbf{y}|\mathbf{x})$. This step simply describes the process giving rise to the data **x** in terms of the unknown parameters **x**.

(2) Obtain the prior density $f(\mathbf{x})$. The prior distribution expresses what is known about parameter **x** prior to observe the data **y**.

(3) Apply Bayes' Theorem to derive the posterior density $f(\mathbf{x}|\mathbf{y})$. The posterior will express what is known about parameter **x** after observing the data **y**.

(4) Derive appropriate inference statements from the posterior distribution. This step is designed to bring out the information expressed in the posterior distribution. Examples of inferences include: point estimates, interval estimates or probabilities of hypotheses. A good inference is the one which effectively conveys information about **x** from the posterior distribution.

Consider a linear model in the form

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon} \qquad (4.3)$$

where **y** is an $n \times 1$ vector of observations, **H** is an $n \times p$ matrix of known coefficients, **x** is a $p \times 1$ vector of parameters and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors. The elements of $\boldsymbol{\varepsilon}$ are assumed to have zero mean and known variance-covariance $\boldsymbol{\Sigma}$. The linear model as Eq. (4.3) can be used to derive the likelihood function $f(\mathbf{y} \mid \mathbf{x}, \Sigma)$ as the conditional distribution of $\mathbf{y}$ given parameters $(\mathbf{x}, \Sigma)$ .When using the likelihood function $f(\mathbf{y} \mid \mathbf{x}, \Sigma)$ from a linear model Eq. (4.3), the four-step Bayesian method is called Bayesian linear estimation.

In this chapter, the route flows will be estimated by Bayesian linear model and the sensor location problem will be addressed within this Bayesian platform.

4.1.2 Problem Description

The flow of interest in this chapter is the deterministic OD demand. With the assumption of a known route choice set, this chapter first introduces a model for good point estimates of route flow volumes in a traffic network. The counting devices used in this chapter are "noisy sensors", which introduces random measurement errors into observation data. The route flows are estimated by a linear Bayesian method and the estimates represent the means of the route flow volumes. In this chapter, the prior knowledge for route flow *i* is assumed to have a probability distribution with mean μ_0^i and variance $\eta_{ii}^{(0)}$. The mean μ_0^i represents the magnitude of the prior information and the variance $\eta_{ii}^{(0)}$ is the measurement about how reliable the mean value μ_0^i is. The uncertainty of the prior information describes the model's knowledge towards the perception errors in decision space. Conjugate prior distribution (*MVN*) is chosen in this chapter in order to produce a mathematical tractable result of posterior distribution. This chapter then develops a decision model for the location of limited number of noisy sensors — the observations from which will be used in the Bayesian estimation model.

The objective of second stage is to maximize the uncertainty reduction in the posterior knowledge of route flows. A new integer programming model and a sequential algorithm are developed to obtain the location solution.

4.1.3 Notations

We will first provide the notation only so that readers familiar with this problem can quickly get to the modeling constructs. Readers unfamiliar with this topic can come back to this notation section to follow the modeling developments. Since there is significant notation, we classify notation as it relates to (a) network topology, (b) definition of route flows, (c) links flows, (d) observations or measurements, and (e) decision variables.

Network Topology Parameters:

R: Route-choice set for a network

A: Set of all the links in the network

A': Set of links where sensors are located

A_f: Set of links feasible for sensors to be located

$|R|$: Number of routes in a network

$|A|$: Number of links in a network

N: Number of sensors to locate

$\rho_a^i \in (0,1)$: Link-route parameter. $\rho_a^i = 1$ if route *i 3*uses link *a*, otherwise, $\rho_a^i = 0$

$$\mathbf{H} = \begin{bmatrix} \rho_1^1 & \cdots & \rho_1^{|R|} \\ \vdots & \ddots & \vdots \\ \rho_{|A|}^1 & \cdots & \rho_{|A|}^{|R|} \end{bmatrix}$$

Route Flows:

x^i : Route flow parameter (real mean) of i^{th} route

$\mathbf{x} = (x^1, x^2, ..., x^{|R|})'$

μ_0^i : Mean of prior distribution of x^i

$\boldsymbol{\mu}_0 = (\mu_1^{(0)}, \mu_2^{(0)}, ..., \mu_{|R|}^{(0)})'$

$\eta_{ij}^{(0)}$: Covariance of prior distribution between x^i and x^j (variance of prior distribution if $i = j$)

$$\mathbf{V_0} = \begin{bmatrix} \eta_{11}^{(0)} & \cdots & \eta_{1|R|}^{(0)} \\ \vdots & \ddots & \vdots \\ \eta_{|R|1}^{(0)} & \cdots & \eta_{|R||R|}^{(0)} \end{bmatrix}$$

u^i : Prior's reliability of route i

$u^{\max}$: Bound of prior's reliability

ξ_i : Real error in route flow prior of route i

μ_1^i : Mean of posterior distribution of x^i

$\boldsymbol{\mu}_1 = (\mu_1^{(1)}, \mu_2^{(1)}, ..., \mu_{|R|}^{(1)})'$

$\eta_{ij}^{(1)}$: Covariance of posterior distribution between x^i and x^j (variance of posterior distribution if $i = j$)

$$\mathbf{V}_1 = \begin{bmatrix} \eta_{11}^{(1)} & \cdots & \eta_{1|R|}^{(1)} \\ \vdots & \ddots & \vdots \\ \eta_{|R|1}^{(1)} & \cdots & \eta_{|R||R|}^{(1)} \end{bmatrix}$$

Link Flows:

v_a : Real link flow on link a

$\mathbf{v} = (v_1, v_2, ..., v_{|A|})'$

$\hat{v}_a$: Potential link flow observation if a sensor is locatedon link *a*

$\hat{\mathbf{v}} = (\hat{v}_1, \hat{v}_2, ..., \hat{v}_{|A|})'$

ε_a : Measurement error on link *a*

τ_a^2 : Variance of link flow measurement on link *a*

$$\Sigma = \begin{bmatrix} \tau_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tau_{|A|}^2 \end{bmatrix}$$

t^a : Sensor's reliability when located on link *a*

$t^{\max}$: Bound of sensors' reliability

$\tilde{v}_a$: Mean of Link flow prior means on link *a*

Observations by sensors:

y_n : Link flow observed by sensor *n*

$\mathbf{Y} = (y_1, y_2, ..., y_N)'$

g_n : An instance of link flow observed by sensor *n*

$\mathbf{g} = (g_1, g_2, ..., g_N)'$

φ_n^2 : Variance of observation by sensor *n*

Decision Variables for Location Models:

$y_a \in (0,1)$: $y_a = 1$ if locating sensor on link *a*, otherwise $y_a = 0$ (used for existing location models)

$z_{na} \in (0,1)$: $z_{na} = 1$ if locating n^{th} sensor on link *a*, otherwise $z_{na} = 0$.

$$\mathbf{z} = \begin{bmatrix} z_{11} & \cdots & z_{1|A|} \\ \vdots & \ddots & \vdots \\ z_{N1} & \cdots & z_{N|A|} \end{bmatrix}$$

$$\mathbf{z}_n = [z_{n1}, ..., z_{n|A|}]$$

Among these parameters, ρ_a^i are obtained from definitions of the given route choice sets. **x** is a column vector for the means or real route flow with elements x^i for i=1,…,|R|. **x** has a prior distribution and we will discuss later. $\hat{v}_a$ represents the flow observed on link a in the link set A, and it is the elements of $\hat{\mathbf{v}}$, which is a |A|-dimension column vector. The relationship between $\hat{v}_a$ and x^i is

$$\hat{v}_a = \sum_{k \in R} \rho_a^k x^k + \varepsilon_a, \forall a \in A \tag{4.4}$$

or

$$\hat{\mathbf{v}} = \mathbf{Hx} + \boldsymbol{\varepsilon} \tag{4.5}$$

where ε_a is the random measurement error from sensor if one is located on link a. Eq. (4.4) indicates that the observation of link flow on link a is equal to the sum of all route flows that use link a, plus a random measurement error coming from the device. This relationship always exists for entire link set A, however, we can measure only those in set A' where sensors are located. If link a is not covered by a sensor, Eq. (4.4) will not be active in the analysis (for estimating route flows).

The decision variables in this problem are binary and defined as z_{na} ($n = 1, ..., N$, $a = 1, ..., |A|$), where n is the subscript for sensors and a is the subscript for links. N is the number of sensors'(a given limit) and |A| is the total number of links. $z_{na} = 1$ means the n^{th} sensor is located on link a. Note that each sensor can only be located on only one link;

and for each specific link *a*, we only allow at most one sensor. These restrictions can be expressed as

$$\sum_{a=1}^{|A|} z_{na} = 1, n = 1, ..., N \tag{4.6}$$

$$\sum_{n=1}^{N} z_{na} \le 1, \forall a \in A \tag{4.7}$$

We can define a matrix **z** for the decision variables, using z_{na} as elements. The dimension of **z** is *N* by |*A*|, and the number of binary variables for each problem is *N**|*A*|.

Since $\hat{\mathbf{v}}$ is a column vector with dimension |*A*|, the rows in Eq. (4.5) that can be accessed by the estimation model include only the rows (i.e., links) where a sensor is located, in particular when $z_{na} = 1$ By the definition of **z**, the matrix multiplication $\mathbf{z}\hat{\mathbf{v}}$ selects the links with sensors.

Take an example of a four-link network with link flow volume $\hat{\mathbf{v}} = (1, 20, 30, 4)'$, $\mathbf{z} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ indicates that the first sensor is located on link 2 (in row 1, "1" appears in column 2 of **z**) and the second sensor is located on link 4 (in row 2, "1" appears in column 4 of **z**). Hence $\mathbf{z}\hat{\mathbf{v}} = (20, 4)'$ simply selects the volumes for link 2 and 4.

Therefore, we can modify Eq. (4.5) by Eq. (4.8)

$$\mathbf{Y} = \mathbf{z}\hat{\mathbf{v}} = \mathbf{z}(\mathbf{Hx} + \boldsymbol{\varepsilon}) = \mathbf{zHx} + \mathbf{z}\boldsymbol{\varepsilon} \tag{4.8}$$

which eliminates the links without sensors and captures actual observations collected by candidate sensors **z**.

We assume the random measurement error follows a normal distribution with mean zero and variance τ_a^2 and observations on different links are independent, that is $\boldsymbol{\varepsilon} \sim MVN(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a diagonal "*dispersion*" matrix for link flow observations.

Note that the parameter $\mathbf{\Sigma}$ can be estimated from the same type of sensor at similar locations or related studies in other areas. Because of Eq. (4.5), $\hat{\mathbf{v}} \,|\, \mathbf{x} \sim MVN(\mathbf{Hx}, \mathbf{\Sigma})$ is the conditional probability of link flows $\hat{\mathbf{v}}$ given route flow **x**. This conditional probability can be used for the likelihood function of the potential link flow observations $\hat{\mathbf{v}}$. Again, because of the nature of *SLFE* problem, which is a sequential decision-making procedure for first locating sensors and then estimating flow of interest, the likelihood that is used for computing the posterior distribution, is only available for the links with sensors at the estimation stage. Hence, given sensors' location, the accessible measurement errors considering the sensors' location is $\mathbf{z\varepsilon} \sim MVN(0, \mathbf{z\Sigma z'})$ (Mardiaet al., 1979), and from Eq. (4.8) link flow observations **Y** can be expressed as $\mathbf{Y} \,|\, \mathbf{x} \sim MVN(\mathbf{zHx}, \mathbf{z\Sigma z'})$.

Recall that route flow **x** has prior information. In this chapter, we assume the prior distributions are multivariate normal (*MVN*) which provides accurate approximation for the true distribution of traffic volumes (Maher, 1983). Define prior distribution $\mathbf{x} \sim MVN(\boldsymbol{\mu}_0, \mathbf{V}_0)$, where $\boldsymbol{\mu}_0$ and $\mathbf{V}_0$ are mean and variance of prior distribution. *MVN* is the conjugate prior distribution for the likelihood chosen in this chapter, which gives the mathematically tractable procedures to obtain posterior distribution for the estimates.

4.2 Bayesian Linear Model of Deterministic Route Flows

4.2.1 Route Flow Estimation with Observations from Multiple Links

As the lower level of *SLFE* problem, the procedure of estimating flow variables from observations, which is analogous to obtaining a posteriori knowledge based on observation evidence and prior knowledge, will be conducted once the sensors are located. If the flows of interest are route flow volumes, the corresponding problem is to reproduce the estimated link flows that are close to the observed link flows from the counting sensors

Originated from the Bayesian statistical approach of Maher (1983), the Bayesian method in this chapter is developed as follows:

With given locations $\mathbf{z}$, if $\mathbf{x} \sim MVN(\boldsymbol{\mu}_0, \mathbf{V}_0)$ and $\mathbf{Y} = \mathbf{zHx} + \mathbf{z}\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim MVN(0, \boldsymbol{\Sigma})$, so that $\mathbf{Y} \,|\, \mathbf{x} \sim MVN(\mathbf{zHx}, \mathbf{z\Sigma z'})$, the posterior distribution of $\mathbf{x}$ given observations $\mathbf{Y = g}$ is also *MVN*.

The probability densities of $\mathbf{x}$ and $\mathbf{Y} \,|\, \mathbf{x}$ are:

$$p(\mathbf{x}) = \frac{(2\pi)^{-n/2}}{|\det \mathbf{V}_0|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_0)' \mathbf{V}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) \right] \qquad (4.9)$$

$$f(\mathbf{Y} \,|\, \mathbf{x}) = \frac{(2\pi)^{-p/2}}{|\det \mathbf{z} \sum \mathbf{z}'|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{Y} - \mathbf{zHx})' (\mathbf{z} \sum \mathbf{z}')^{-1} (\mathbf{Y} - \mathbf{zHx}) \right] \qquad (4.10)$$

By Bayesian theorem, the posterior as:

$$f(\mathbf{x} \,|\, \mathbf{Y = g}) \propto p(\mathbf{x}) f(\mathbf{Y = g} \,|\, \mathbf{x}) \sim MVN(\boldsymbol{\mu}_1, \mathbf{V}_1) \qquad (4.11)$$

Therefore, $f(\mathbf{x} \,|\, \mathbf{Y = g})$

$$\propto \exp\left[-\frac{1}{2} (\mathbf{g} - \mathbf{zHx})' (\mathbf{z\Sigma z'})^{-1} (\mathbf{g} - \mathbf{zHx}) \right] \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_0)' \mathbf{V}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) \right]$$

$$= \exp\Big[-\frac{1}{2} \Big(\mathbf{g}'(\mathbf{z\Sigma z'})^{-1}\mathbf{g} - \mathbf{x'H'z'}(\mathbf{z\Sigma z'})^{-1}\mathbf{g} - \mathbf{g}'(\mathbf{z\Sigma z'})^{-1}\mathbf{zHx} + \mathbf{x'H'z'}(\mathbf{z\Sigma z'})^{-1}\mathbf{zHx}$$

$$+ \mathbf{x'V}_0^{-1}\mathbf{x} - \boldsymbol{\mu}_0'\mathbf{V}_0^{-1}\mathbf{x} - \mathbf{x'V}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\mu}_0'\mathbf{V}_0^{-1}\boldsymbol{\mu}_0 \Big) \Big]$$

$$= \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_1)' \mathbf{V}_1^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) \right]$$

$$= \exp\left[-\frac{1}{2} \left(\mathbf{x'V}_1^{-1}\mathbf{x} - \boldsymbol{\mu}_1'\mathbf{V}_1^{-1}\mathbf{x} - \mathbf{x'V}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_1'\mathbf{V}_1^{-1}\boldsymbol{\mu}_1 \right) \right]$$

Equating the corresponding terms gives that

$$\boldsymbol{\mu}_1 = \mathbf{V}_1(\mathbf{V}_0^{-1}\boldsymbol{\mu}_0 + \mathbf{H'z'(z\Sigma z')^{-1}g}) \quad (4.12)$$

$$\mathbf{V}_1 = (\mathbf{V}_0^{-1} + \mathbf{H'z'(z\Sigma z')^{-1}zH})^{-1} \quad (4.13)$$

Further matrix operations on Eq. (4.12) and Eq. (4.13) give that

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \mathbf{V}_0\mathbf{H'z'(z\Sigma z' + zHV_0H'z')^{-1}(g - zH}\boldsymbol{\mu}_0) \quad (4.14)$$

$$\mathbf{V}_1 = \mathbf{V}_0 - \mathbf{V}_0\mathbf{H'z'(z\Sigma z' + zHV_0H'z')^{-1}zHV_0} \quad (4.15)$$

Therefore, for the prior distribution of route flow $\mathbf{x} \sim MVN(\boldsymbol{\mu}_0, \mathbf{V}_0)$ and the posterior distribution $\mathbf{x} \,|\, \mathbf{Y} = \mathbf{g} \sim MVN(\boldsymbol{\mu}_1, \mathbf{V}_1)$, Eq. (4.14) and Eq. (4.15) update for the mean vector and dispersion matrix for any given sensors' location **z**.

The change in posterior variance is

$$\mathbf{V}_1 - \mathbf{V}_0 = -\mathbf{V}_0\mathbf{H'z'(z\Sigma z' + zHV_0H'z')^{-1}zHV_0} \quad (4.16)$$

From Eq. (4.16), the posterior dispersion matrix varies according to the different sensors' location matrix **z**. Because $\mathbf{z\Sigma z' + zHV_0H'z'}$ is positive definite, its inverse is positive definite as well. Therefore, Eq. (4.16) indicates that the posterior variances $\mathbf{V}_1$ always involves a variance reduction comparing to $\mathbf{V}_0$. The trace of Eq. (4.16) is the amount of variances reduction in posterior using above Bayesian' approach to update route flows information.

4.2.2 Route Flow Estimation with Observations from Single Links

When using Eq. (4.14) and Eq. (4.15) as the updating equations for mean and dispersion matrix of route flows, the matrix inverse is a time-consuming operation in objective function. Maher indicated that because of the sequential nature of Bayes'

Theorem, one can iteratively conduct the Bayesian procedure using independent individual observations, such that the posterior from one observation becomes the prior for the next. The sequence of selecting links for measurements does not have effect on the posterior distribution in the final step.

Suppose that a single observation of the *n*th sensor is taken from location $\mathbf{z_n}$, which is a |*A*|-dimension row vector — the n^{th} row from **z** in particular. If $z_{na} = 1$, the n^{th} sensor is located on link *a* and then the dispersion matrix **Σ** contains a single element $\mathbf{z_n \Sigma z_n'}$, which can be defined as the uncertainty brought in by sensor *n*. The sensors' uncertainty is simplified as $\varphi_n^2 = \sum_{a=1}^{|A|} z_{na}\tau_a^2$, because for each *n*, there is one and only one z_{na} to be non-zero. Matrix **H** becomes a row vector with only a^{th} row from original **H** left. Such reduced matrix **H** can be described by

$$\mathbf{z_n H} = [\sum_{a=1}^{|A|} z_{na}\rho_a^1, ..., \sum_{a=1}^{|A|} z_{na}\rho_a^{|R|}] = [h_n^1, ..., h_n^{|R|}] \tag{4.17}$$

Because the inverse of a scalar $\mathbf{z_n \Sigma z_n' + z_n H V_0 H' z_n'}$ is the reciprocal, the updating equations become:

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \frac{\mathbf{V_0 H' z_n'}}{\mathbf{z_n \Sigma z_n' + z_n H V_0 H' z_n'}}(\mathbf{g_n} - \mathbf{z_n H} \boldsymbol{\mu}_0) \tag{4.18}$$

$$\mathbf{V}_1 = \mathbf{V_0} - \frac{\mathbf{V_0 H' z_n' z_n H V_0}}{\mathbf{z_n \Sigma z_n' + z_n H V_0 H' z_n'}} \tag{4.19}$$

The elements $\mu_i^{(1)}$ and $\eta_{ij}^{(1)}$ ($i = 1,...,|R|,\ \ j = 1,...,|R|$) of $\boldsymbol{\mu}_1$ and $\mathbf{V}_1$ are given by

$$\mu_i^{(1)} = \mu_i^{(0)} + \frac{\sum_{j=1}^{|R|} h_n^j \eta_{ij}^{(0)}}{\varphi_n^2 + \sum_{i=1}^{|R|} h_n^i \left(\sum_{j=1}^{|R|} h_n^j \eta_{ij}^{(0)}\right)}(g_n - \sum_{r=1}^{|R|} h_n^r \mu_r^{(0)}) \tag{4.20}$$

$$\eta_{ij}^{(1)} = \eta_{ij}^{(0)} - \frac{\left(\sum_{j=1}^{|R|} h_n^j \eta_{ij}^{(0)}\right)\left(\sum_{i=1}^{|R|} h_n^i \eta_{ij}^{(0)}\right)}{\varphi_n^2 + \sum_{i=1}^{|R|} h_n^i \left(\sum_{j=1}^{|R|} h_n^j \eta_{ij}^{(0)}\right)} \qquad (4.21)$$

Denoting $\sum_{j=1}^{|R|} h_n^j \eta_{ij}^{(0)}$ by S_n^i ($i = 1,...,|R|, \quad n = 1,...,N$) and $\sum_{i=1}^{|R|} h_n^i S_n^i$ by T_n , $\sum_{r=1}^{|R|} h_n^r \mu_r^{(0)}$ is the mean of link flow prior (calculated as the summation of prior means of routes passing that link) denoted as $\tilde{v}_n$. Then Eq. (4.20) and Eq. (4.21) are simplified as

$$\mu_i^{(1)} = \mu_i^{(0)} + \frac{S_n^i}{\varphi_n^2 + T_n}(g_n - \tilde{v}_n) \qquad (4.22)$$

$$\eta_{ij}^{(1)} = \eta_{ij}^{(0)} - \frac{S_n^i S_n^j}{\varphi_n^2 + T_n} \qquad (4.23)$$

S_n^i describes the variation on flow of route *i* influenced by the *n*th. T_n is the variation on all routes influenced by sensor *n*'s observation. φ_n^2 is the variation of measurement sensor *n* itself. If sensor's location is decided (*n* is fixed), $\varphi_n^2 + T_n$ will be fixed. The route *i* with large variation S_n^i will be given large weight for sharing the information from observation g_n . The rationale behind is that if the prior route information is not reliable (i.e., has a large variance), it requires more information from observations in order to produce a reliable posterior and vice versa. If the observation is not reliable (φ_n^2 is large), the corresponding ratio will become lower and the observation will be considered with less weight in computation of the distributions of the corresponding routes.

4.3 Locating Sensors to Maximize Variance Reduction

Because the trace of Eq. (4.16) measures the total variance reduction of posterior from prior, the strategy for sensor location is to select a subset of links which have largest potential to reduce the variance of the posterior distributions. The objective function of the sensor location model is to maximize the trace of Eq. (4.16) by selecting different decision variable **z**. This new sensor location model to maximize variance reduction using Bayesian estimator can be formulated as

[RVR-noisy] $$\max tr(\mathbf{V_0H'z'(z\Sigma z' + zHV_0H'z')^{-1}zHV_0}) \quad (4.24)$$

$$s.t. \quad \sum_{a=1}^{|A|} z_{na} = 1, n = 1,...,N \quad (4.25)$$

$$\sum_{n=1}^{N} z_{na} \le 1, \forall a \in A \quad (4.26)$$

$$z_{na} \in (0,1), n = 1,...,N, a \in A \quad (4.27)$$

We can simplify the formulation for independent observations. Suppose N sensors are to be located in the network. Instead of solving [RVR-noisy] directly, we can use an iterative method that locates one link with capability to reduce most posterior variance at a time. After selecting one link, the route flow mean and variance are updated using Eq. (4.22) and Eq. (4.23). The posterior becomes the prior and process repeats for another unselected link. We define A_f as the set of links which have not been selected yet and $\bar{A}_f$ is the set of selected links (and cannot be selected again). The process is repeated until all N links are determined. For each sensor n, its location is decided by the model, formulated as:

[RVRS-noisy] $$\max \sum_{i=1}^{|R|} \frac{(S_n^i)^2}{\varphi_n^2 + T_n} \quad (4.28)$$

$$s.t.\quad h_n^i = \sum_{a=1}^{|A|} z_{na}\rho_a^i, \quad \forall i = 1,...,|R| \tag{4.29}$$

$$\varphi_n^2 = \sum_{a=1}^{|A|} z_{na}\tau_a^2 \tag{4.30}$$

$$S_n^i = \sum_{j=1}^{|R|} h_n^j \eta_{ij}^{(0)}, \quad \forall i = 1,...,|R| \tag{4.31}$$

$$T_n = \sum_{j=1}^{|R|} h_n^j S_n^i \tag{4.32}$$

$$\sum_{a=1}^{|A|} z_{na} = 1 \tag{4.33}$$

$$z_{na} = 0, \forall a \in \bar{A}_f \tag{4.34}$$

$$z_{na} \in (0,1), n = 1,...,N, a \in A \tag{4.35}$$

The objective function Eq. (4.28) in formulation [RVRS-noisy] maximizes total variance reduction $\sum_{i=1}^{|R|}(\eta_{ii}^{(0)} - \eta_{ii}^{(1)})$. Constraints Eq. (4.29) to Eq. (4.32) are introducing notations to simplify objective function Eq. (4.28). Constraint Eq. (4.33) restricts sensor *n* will be located on only one link. Constraint Eq. (4.34) forces the decision of sensor's location be made only from the feasible links' set A_f. Constraint Eq. (4.35) is the binary constraint for decision variables.

4.4 Algorithm

Locating only one sensor in the network for problem MVRS can be solved by a *greedy algorithm*: for each candidate link, calculate the objective value $\sum_{i=1}^{|R|} \frac{(S_n^i)^2}{\varphi_n^2 + T_n}$ and select the one with largest objective value. The solution of MVRS can be found by $\mathcal{O}(|A|)$ time.

Therefore, the entire algorithm for solving [RVR-noisy] is:

BAYESIAN BASED LOCATION ALGORITHM [RVR-NOISY]

Input: $\boldsymbol{\mu}_0, \mathbf{V}_0, \mathbf{H}, N$

Output: $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_N)'$, $\boldsymbol{\mu}_1, \mathbf{V}_1$

(1) Set $n := 1$.

Set $A_f = A$.

Set $\bar{A}_f = \varnothing$.

(2) For each link $a \in A_f$, calculate $\sum_{i=1}^{|R|} \frac{(S_n^i)^2}{\varphi_n^2 + T_n}$.

(3) Select the link $a \in A_f$ with largest value calculated in Step 2 and $\mathbf{z_n}$ is a row vector with only non-zero element in *a*th column. Remove link *a* from A_f and add it into $\bar{A}_f$.

(4) Using observation on link *a* to update posterior using Eq. (4.14) and Eq. (4.15) for $\boldsymbol{\mu}_\mathbf{1}, \mathbf{V}_\mathbf{1}$.

(5) Increment *n* by 1.

(6) If $n = N + 1$ then stop.

Else set $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_\mathbf{1}$, $\mathbf{V}_0 = \mathbf{V}_\mathbf{1}$ and go to Step 2.

The running time for this algorithm is polynomial because the entire algorithm will run for *N* loops; each loop consists of a preprocessing (step 2) and a searching (step 3) process that can be conducted in polynomial time.

The optimality of above algorithm can be proved by the sequential nature of Bayes' Theorem. Assume our sensor location **z** containing two independent observations $(\mathbf{z}_1, \mathbf{z}_2)'$, the likelihood is:

$$l(\mathbf{z} \mid \mathbf{x}) = l(\mathbf{z}_1, \mathbf{z}_2 \mid \mathbf{x}) = l(\mathbf{z}_1 \mid \mathbf{x}) \cdot l(\mathbf{z}_2 \mid \mathbf{x}) \tag{4.36}$$

The posterior route flows when using observation **z** is:

$$p(\mathbf{x} \mid \mathbf{z}) \propto p(\mathbf{x}) l(\mathbf{z} \mid \mathbf{x}) = p(\mathbf{x}) l(\mathbf{z}_1 \mid \mathbf{x}) l(\mathbf{z}_2 \mid \mathbf{x}) \tag{4.37}$$

where $p(\mathbf{x})$ is the prior distribution of route flows.

Now consider a two-stage process of sequentially updating priors using observations $\mathbf{z}_1$ and $\mathbf{z}_2$. The first posterior probability for route flow **x** after applying observation $\mathbf{z}_1$ can be derived as:

$$p(\mathbf{x} \mid \mathbf{z}_1) \propto p(\mathbf{x}) l(\mathbf{z}_1 \mid \mathbf{x}) \tag{4.38}$$

Updating posterior probability $p(\mathbf{x} \mid \mathbf{z}_1)$ using Eq. (4.38) and then using it as the prior probability in the next stage, the second posterior probability is:

$$p(\mathbf{x} \mid \mathbf{z}_1, \mathbf{z}_2) \propto p(\mathbf{x} \mid \mathbf{z}_1) l(\mathbf{z}_2 \mid \mathbf{x}) \tag{4.39}$$

Substituting Eq. (4.38) in Eq. (4.39) gives the posterior probability when sequentially using observations $\mathbf{z}_1$ and $\mathbf{z}_2$ as:

$$p(\mathbf{x} \mid \mathbf{z}_1, \mathbf{z}_2) \propto p(\mathbf{x}) l(\mathbf{z}_1 \mid \mathbf{x}) l(\mathbf{z}_2 \mid \mathbf{x}) \tag{4.40}$$

Because the right hand sides of Eq. (4.37) and Eq. (4.40) are the same ($p(\mathbf{x}) l(\mathbf{z}_1 \mid \mathbf{x}) l(\mathbf{z}_2 \mid \mathbf{x})$), we have that:

$$p(\mathbf{x} \mid \mathbf{z}) = p(\mathbf{x} \mid \mathbf{z}_1, \mathbf{z}_2) \tag{4.41}$$

Eq. (4.41) confirms that updating for route flow distribution from $p(\mathbf{x})$ to $p(\mathbf{x}\,|\,\mathbf{z})$ directly using both observations at the same time is equivalent to a two-stage process that sequentially updates the distribution using individual sensor measurements.

Therefore, the posterior with all N observations (solution for model RVR-noisy) can be obtained by a series of application of Bayesian updates using individual sensor measurements.

4.5 Experimental Result

4.5.1 Experiment Setup

The proposed RVR-noisy model and algorithm are applied to two experiment networks and the performances are compared with three existing models: Link Flow Coverage (LFC), Route Flow Coverage (RFC) and RVR-perfect model from chapter 3. The reason to select LFC and RFC in comparison is that both models are widely reported and rely on network structures and the magnitude of prior information; RVR-perfect model relies on the variance or degree of believe of prior information (no measurement errors considered). All the experiments were conducted in Matlab.

Figure 8 shows the scheme of the experimental procedure. First (Steps 1-2), problem scenarios are generated by defining the network (supply), loading the assumed traffic flows (demand) and obtaining the "actual" mean of network traffic x^i and v_a. Assumed route flow prior distribution (mean and variance), sensors' reliabilities and potential link flow observations are generated by linking the reliabilities to the actual values (Step 3). For each route, the prior's reliability is first randomly generated by $u^i \in (0, u^{max})$; then the prior's variance is calculated by $\eta_{ii}^{(0)} = x^i u^i$; the prior's mean is generated from a normal random generator $\mu_i^{(0)} \sim N(x^i, \eta_{ii}^{(0)})$; all covariance are assumed to be zero in the initial priors. Similar logic is used to generate sensors' reliabilities and potential observations. For each link, the sensor's reliability t_a is randomly generated within range $[0, t^{\max}]$ and variance of link flow measurement is calculated by $\tau_a^2 = v_a t_a$; the potential observation on such link is generated from a normal random generator $\hat{v}_a \sim (v_a, \tau_a^2)$. In Step 4, RVR-noisy and other location models are solved using Matlab to obtain location decisions **z**. Then, with these decisions, posterior mean and variance of route flows are obtained using the Bayesian estimation model in section 4.2 (Step 5). To evaluate each location model (Step 6), the posterior means are compared with actual route flows in terms of the bias. Total variances (trace of $\mathbf{V}_1$) are compared as well. The evaluation criterion is the sum-squared error (SSE) of the posterior distribution, which is defined by the summation of variance and squared bias:

$$SSE = BIAS^2 + Variance = \sum_{i \in R} (\mu_i^{(1)} - x^i)^2 + \sum_{i \in R} \eta_{ii}^{(1)} \qquad (4.42)$$

where x^i is the true mean value of the route flow *i*. Because *SSE* measures the combination of bias and variance, if the magnitude of variance is significantly lower than the squared bias, *SSE* will be dominated by the effect of bias. Because our proposed RVR-noisy already minimizes the posterior variances, the criterion *SSE* is more sensitive for showing the effect of bias.

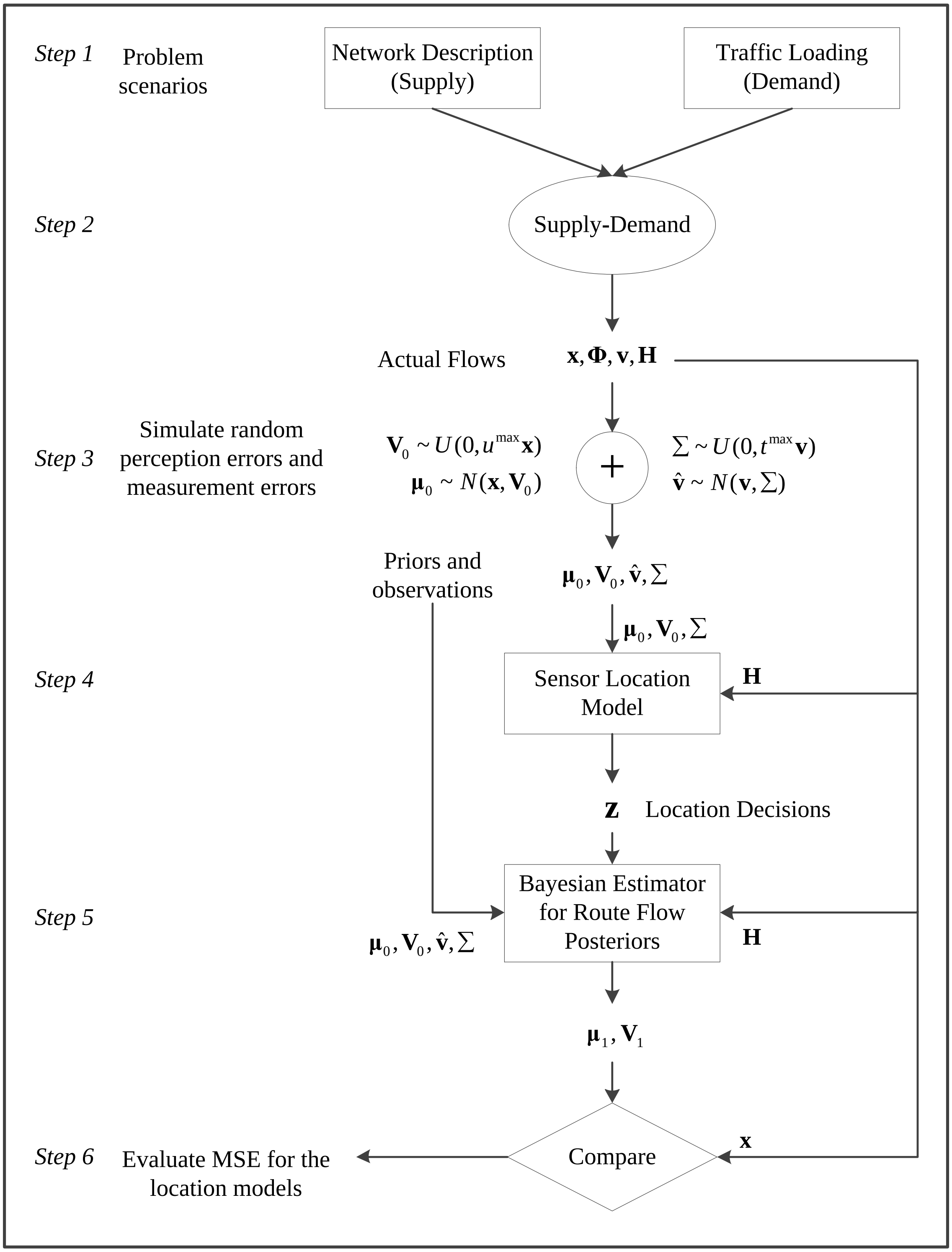

Figure 8 Procedure for the computational experiments for RVR-noisy

Finally, note that for each network scenario, we can have different problem instances by locating different number of sensors between 1 and (say) *N*, the number of possible sensor sites. Furthermore, we can have priors with different qualities, in particular prior variance and bias; sensors can have various reliabilities. Hence, each problem instance in our computational experiments is defined by:

(1) Network supply

(2) Traffic demand

(3) Reliability of priors on traffic flows to be estimated

(4) Reliability of sensors

(5) Number of sensors to be located.

Solving the location model for each problem instance, for the four models being compared, gives us a score for each model, as defined by the SSE.

Two sets of network scenarios were used in the computational experiments.

(1) A grid network with 16 OD pairs, 43 routes and 48 links.

(2) A grid network with 16 OD pairs, 204 routes and 112 links.

4.5.2 Numerical Results

Four network scenarios are tested with two initial prior distributions. We simulated two sets of link flow observations for each grid network. When generating the experiments, u^{max} is set to be 0.15 and t^{max} is 0.1. This parameter setting reflects that we assume counting devices are more accurate then our prior knowledge in practice.

Table 6 shows the *SSE* results for the four comparative location models. The table gives the averaged *SSE* for every five instances locating different numbers of sensor locations. For example, the first row in Table 6 represents the average SSE scores for

N=1 to *N*=5. The best scores are in bold in Table 6. In general, our proposed RVR-noisy (denoted by RVR-N in the following tables) outperforms other models in most cases, especially when few sensors are in use. The row marked as “Total” in Table 6 sums all of the *SSE* scores and emphasizes on the advantage of RVR-noisy compared to other models for grid network 1. Observe, however, that RVR-perfect (denoted by RVR-P in the following tables) model sometimes has the best score, and often is the second best when RVR-noisy is the best.

Table 6 SSE Results using RVR-noisy for Grid Network1

		obs_1				obs_2			
		RVR-N	RVR-P	LFC	RFC	RVR-N	RVR-P	LFC	RFC
prior_1	1-5	***481.04***	486.86	821.60	832.85	***288.16***	288.61	825.64	823.68
	6-10	***455.34***	462.25	808.11	566.57	270.54	***269.36***	800.22	402.88
	11-15	***441.49***	441.76	474.10	502.76	259.35	***248.85***	253.25	294.90
	16-20	***437.53***	437.47	472.21	491.71	226.68	***223.72***	248.50	286.06
	21-25	***439.34***	440.04	469.49	478.66	***220.23***	220.38	249.93	263.22
	26-30	***440.92***	441.22	452.77	455.85	***218.40***	222.44	237.03	262.63
	31-34	***441.26***	441.88	445.33	447.31	218.02	***217.22***	221.08	222.92
	Total	***3136.91***	3151.49	3943.63	3775.71	1701.37	***1690.57***	2835.65	2556.28
prior_2	1-5	***114.36***	156.33	140.54	122.59	***118.80***	128.93	139.14	127.45
	6-10	***136.22***	150.23	162.73	147.97	***103.33***	116.14	130.70	123.33
	11-15	***125.14***	150.63	139.41	151.94	106.75	***101.83***	127.85	126.88
	16-20	***129.06***	145.52	133.88	152.74	100.66	***94.94***	108.52	119.38
	21-25	140.50	140.12	***137.63***	155.63	***83.66***	85.28	103.27	109.53
	26-30	137.87	***135.17***	136.93	146.41	***83.94***	85.73	91.05	107.55
	31-34	134.66	***134.12***	135.57	143.75	87.50	86.82	***86.75***	97.41
	Total	***917.80***	1012.13	986.69	1021.03	***684.63***	699.68	787.27	811.53

We have similar comparisons in Table 7 among the *SSE* scores for the four location models applied to the grid network 2. The RVR-noisy model performs the best in most cases and on the overall performance in the rows marked “Total”.

Table 7 SSE Results using RVR-noisy for Grid Network 2

	SSE	obs_1				obs_2			
		RVR-N	RVR-P	MFC	RFC	RVR-N	RVR-P	MFC	RFC
prior_1	1-10	***31.47***	33.11	46.28	42.94	34.80	***34.25***	47.20	40.78
	11-20	***21.89***	22.28	38.84	44.33	***22.68***	24.35	42.88	40.49
	21-30	19.82	***19.82***	29.95	43.42	20.69	***18.77***	37.42	42.53
	31-40	***17.80***	18.33	21.62	40.71	***17.42***	18.38	23.41	45.02
	41-50	***17.74***	18.08	20.04	29.26	***18.86***	20.04	21.58	35.59
	51-60	***17.45***	17.83	18.49	24.71	***19.60***	19.66	20.50	30.81
	61-69	17.28	***17.19***	17.22	19.87	19.37	19.32	***19.30***	24.06
	Total	***143.46***	146.63	192.43	245.23	***153.41***	154.77	212.29	259.28
prior_2	1-10	14.19	***13.33***	20.60	15.02	13.21	***13.06***	24.92	15.41
	11-20	11.84	***11.66***	18.00	14.11	***8.78***	10.33	23.03	13.13
	21-30	10.52	***10.29***	11.89	13.81	7.74	***7.70***	17.39	17.77
	31-40	9.66	***9.58***	12.02	12.42	***7.68***	7.97	16.68	18.77
	41-50	***9.58***	9.91	9.84	11.26	***7.56***	7.84	8.23	17.11
	51-60	9.09	***8.77***	8.98	10.55	7.63	7.91	***7.57***	14.22
	61-69	8.63	***8.61***	8.60	9.71	7.73	7.71	***7.67***	9.03
	Total	73.50	***72.15***	89.91	86.88	***60.32***	62.52	105.48	105.44

Figure 9 and Figure 10 plot the average posterior variances when locating different number of sensors using RVR-noisy. Figure 9 and Figure 10 indicate that, for an instance, the more sensors in use, the lesser the posterior variances. However, the marginal reduction in posterior variances decreases when *N* is increased. In addition, the majority (more than 70%) posterior variances can be reduced when using data from up to 25% links in a network.

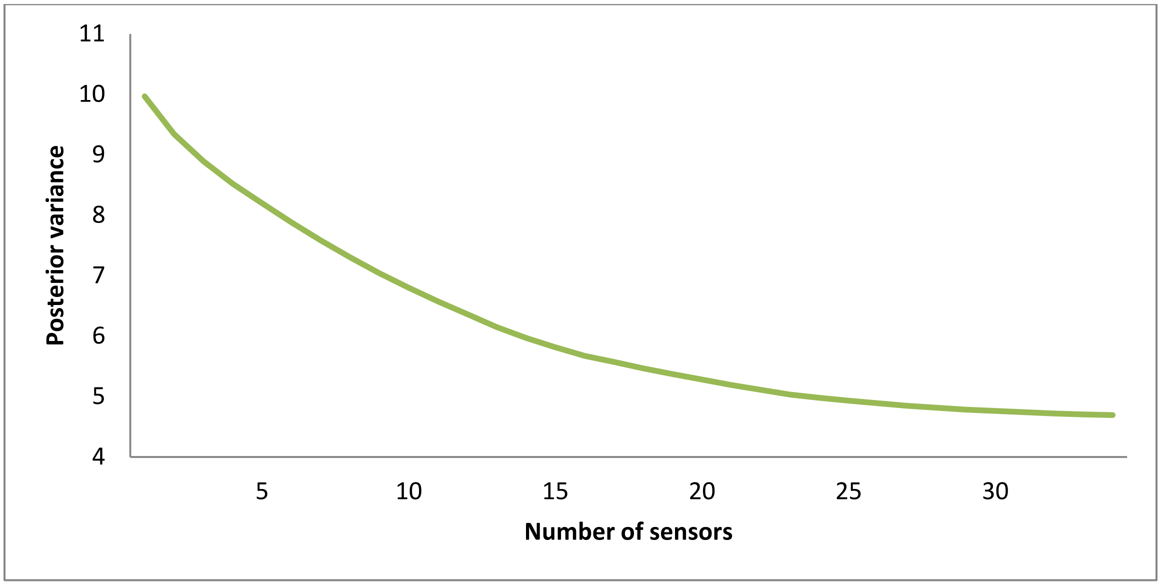

Figure 9 Route flow posterior variance using RVR-noisy in network 1

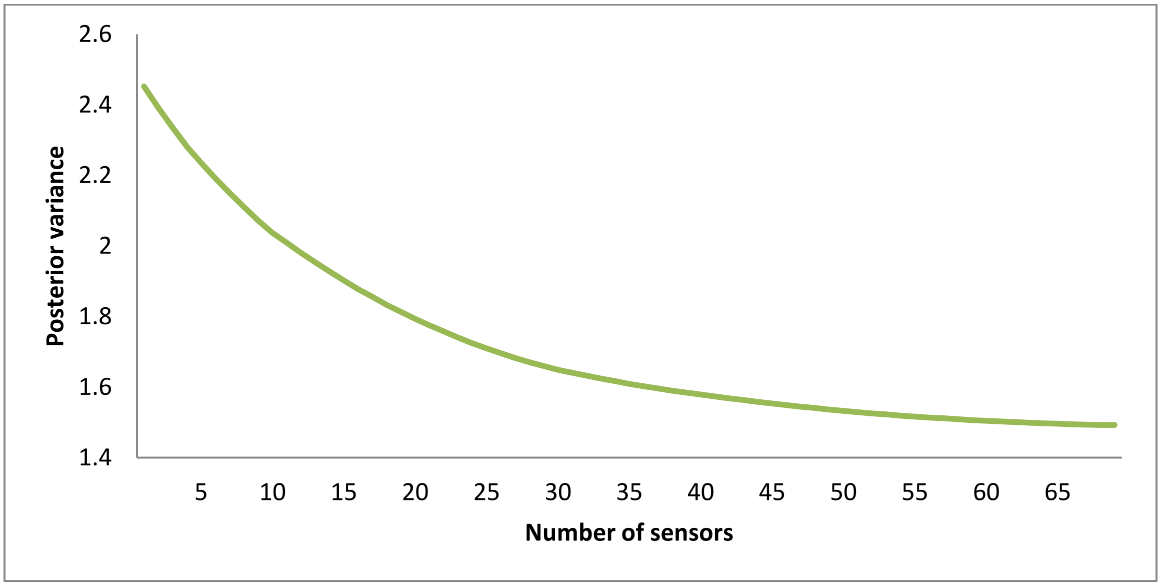

Figure 10 Route flow posterior variance using RVR-noisy in network 2

In practice, the decision-makers are more interested in locating sensors only on few links due to budget considerations. Table 8 and 9 tabulate the problem instances when at most half the links can be detectorized, that is, average scores for all instances from *N*=1 to *N*=1/2|*A*| and from *N*=1 to *N*=1/4|*A*| respectively. In each cell of the tables, the number on the top represents the value of *SSE*, while the rest two values are the variance and bias from the instance. Table 8 shows that the average *SSE* for RVR-noisy and RVR-perfect are always lower than the average *SSE* for LFC and RFC. For RVR-noisy and RVR-perfect, the average SSEs are close because when more noise sensors are used, more bias could be introduced in the estimates. Table 9 focuses on model performance when locating fewer sensors. RVR-noisy outperforms among all the instances in terms of SSE, variance and bias.

Table 8 SSE Results using RVR-noisy for Problem Instances for N=1 to N=1/2|A|

		RVR-N		RVR-P		MFC		RFC	
		SSE		SSE		SSE		SSE	
	(1st 50%)	Var	Bias	Var	Bias	Var	Bias	Var	Bias
Network 1	prior_1,obs_1	***456.73***		460.47		674.34		618.45	
		7.00	21.21	7.03	21.29	9.10	25.79	8.72	24.64
	prior_1,obs_2	268.31		***264.66***		581.99		482.60	
		6.15	16.19	6.19	16.08	8.62	23.95	8.74	21.71
	prior_2,obs_1	***125.29***		152.03		145.76		142.06	
		7.27	10.86	7.48	12.02	8.57	11.71	8.65	11.46
	prior_2,obs_2	***109.06***		113.49		130.58		126.16	
		7.04	10.10	7.15	10.31	8.22	11.06	8.92	10.75
Network 2	prior_1,obs_1	***23.52***		24.16		36.05		43.31	
		2.01	4.64	2.03	4.70	2.27	5.81	2.44	6.37
	prior_1,obs_2	24.99		***24.63***		39.95		41.78	
		2.04	4.79	2.06	4.75	2.29	6.14	2.45	6.24
	prior_2,obs_1	11.83		***11.46***		16.16		14.09	
		1.90	3.15	1.91	3.09	2.17	3.74	2.22	3.39
	prior_2,obs_2	***9.58***		10.02		21.13		15.92	
		1.90	2.77	1.91	2.85	2.18	4.35	2.23	3.65

Table 9 SSE Results using RVR-noisy for Problem Instances for N=1 to N=1/4|A|

		RVR-N		RVR-P		MFC		RFC	
		SSE		SSE		SSE		SSE	
	(1st 25%)	Var	Bias	Var	Bias	Var	Bias	Var	Bias
Network 1	prior_1,obs_1	***469.48***		475.78		816.39		721.42	
		7.97	21.48	7.99	21.63	10.30	28.39	9.58	26.65
	prior_1,obs_2	***280.05***		280.14		816.26		648.46	
		7.14	16.52	7.17	16.52	10.01	28.39	9.60	25.24
	prior_2,obs_1	***125.46***		153.29		150.35		133.21	
		8.30	10.82	8.51	12.03	9.34	11.87	9.36	11.07
	prior_2,obs_2	***112.52***		123.60		136.18		125.04	
		8.17	10.22	8.29	10.74	9.08	11.27	9.59	10.70
Network 2	prior_1,obs_1	***27.61***		28.80		43.85		43.51	
		2.23	25.38	2.24	26.56	2.47	41.38	2.56	40.71
	prior_1,obs_2	***29.92***		30.54		45.82		40.55	
		2.26	27.66	2.27	28.27	2.47	43.35	2.57	37.75
	prior_2,obs_1	***11.51***		12.14		24.24		14.45	
		2.10	9.41	2.11	10.02	2.35	21.88	2.37	11.88
	prior_2,obs_2	13.39		***12.80***		19.67		14.74	
		2.10	11.28	2.11	10.69	2.36	17.32	2.36	12.17

Because of the random effects in the experiment framework, average score is only one way to compare the performance. Further comparisons are conducted using statistical paired-*t* tests to compare *SSE* scores with those of other models. Here one is interested in testing if the mean of estimation scores using RVR-noisy is significantly lower than that of each of the other models. The hypothesis for the paired-*t* test is:

$$H_0 : \mu_{RVR\_p} \geq \mu_{others}$$

$$H_1 : \mu_{RVR\_p} < \mu_{others}$$

If the *P*-value for a paired-*t* test is lower than 0.05, it means we can reject H_0 at 95% confidence level and have a statistical indication that RVR-noisy is better than the compared model. In addition, if the 95% confidence interval does not include zero, it is a statistical proof that both means are different. We conduct three similar paired-*t* tests to compare *SSE* scores of RVR-noisy versus the *SSE* of RVR-perfect, LFC and RFC separately; the *P*-values and 95% confidence intervals by Minitab are shown in Table 10. The table shows that RVR-noisy significantly outperforms all the other models because the *P*-values are essentially zero and none of the confidence intervals include the value of zero.

Table 10 *P*-values for Paired-*t* Tests Comparison of SSE of RVR-noisy

with Other Location Models

Alternative hypothesis: $\mu_{RVR-N} <$	*p*-value	95% Confidence Interval
μ_{RVR-N}	0.000	(-2.18307, -0.83229)
μ_{LFC}	0.000	Upper bound is -21.9645
μ_{RFC}	0.000	Upper bound is -21.1212

4.6 Chapter Conclusions

In this chapter, Bayesian statistical procedure is used to combine the prior information and observation data to obtain the route flow estimates. In the Bayesian linear model, conjugate prior distributions are assumed for the route flows. A linear integer program is formulated with the objective function of maximizing the expected variance reduction in Bayesian procedure. The solution of the new model provides the optimal subset of links on which to locate sensors. A sub-model for optimally selecting a single link is developed in order to simplify the solution procedure. The proposed algorithm iteratively solves this sub-model and provides the optimal set of links at the end of solution process. The optimality of proposed solution algorithm is proven by sequentially applying the Bayes Theorem.

The experimental results also show that the proposed model RVR performs significantly better than the other two models LFC and RFC. The RVR-perfect performs well but in most cases RVR-noisy performs the best. The outperformance of RVR-noisy and RVR-perfect are statistically compared and the statistical evidence indicates that RVR-noisy outperforms RVR-perfect, which confirms the advantage of considering the reliabilities of measurement devices in the sensor location problems.

Chapter 5

LOCATING SENSORS TO ESTIMATE STOCHASTIC ROUTE FLOWS WITH KNOWN VARIANCE AND ASSUMING NORMAL DISTRIBUTIONS FOR PRIOR UNCERTAINTIES

5.1 Introduction and Problem Description

5.1.1 Introduction to Hierarchical Linear Models

In the four-step Bayesian method described in section 4.1.1, the statistical model and prior model together can form an ordered structure, in which the distribution of the data **Y** is conditioning on parameter **x** as $f(\mathbf{y}|\mathbf{x})$, and the distribution of **x** is written conditionally on hyperparameters $\boldsymbol{\gamma}$ as $f(\mathbf{x}|\boldsymbol{\gamma})$. The distribution of $\boldsymbol{\gamma}$ is $f(\boldsymbol{\gamma})$. Theoretically one could write the distribution of $\boldsymbol{\gamma}$ conditionally on some more parameter and extend to include further stages with new hyperparameters. Such models are called hierarchical models because the distribution of parameter in each level of the hierarchy depends on the parameters in the next level (Clark and Gelfand, 2006).

The parameters (**Y, x,** $\boldsymbol{\gamma}$) described above form a three-stage hierarchical model and they can be viewed as three entities with stochastic features. First, the data **y** is presumed to be drawn from some populations regarding underlying process **x**. Second, the process **x** involves uncertainty that will be estimated by parameter $\boldsymbol{\gamma}$. Third, the parameter $\boldsymbol{\gamma}$ is uncertain and expected to vary depending upon how and where the data is obtained. Because each entity is stochastic, the joint distribution of (**y, x,** $\boldsymbol{\gamma}$) is

$$f(\mathbf{Y},\mathbf{x},\boldsymbol{\gamma}) \propto f\left(\mathbf{Y}|\mathbf{x},\boldsymbol{\gamma}\right) \times f(\mathbf{x}|\boldsymbol{\gamma}) \times f(\boldsymbol{\gamma}), \quad (5.1)$$

which is

$$f(data, process, parameters)$$
$$\propto f(data|process, parameters) \times f(process|parameters) \times f(parameters).$$

It is usually assumed that the distribution of parameters at any stage of the hierarchy depends only on parameters at the next lower level, and independent of parameters at all levels below that. This assumption is based on a judgment that if we know process $\mathbf{x}$ then known parameter $\boldsymbol{\gamma}$ would not add any additional information about data $\mathbf{Y}$, because the prior knowledge about $\boldsymbol{\gamma}$ has been introduced as a way of

$$f(\mathbf{x}) = \int f(\mathbf{x} \mid \boldsymbol{\gamma}) f(\boldsymbol{\gamma}) d\boldsymbol{\gamma} \quad . \tag{5.2}$$

By this assumption, the likelihood $f\left(\mathbf{Y} \mid \mathbf{x}, \boldsymbol{\gamma}\right)$ in Eq. (5.1) is equivalent as $f\left(\mathbf{Y} \mid \mathbf{x}\right)$ and the joint distribution in Eq. (5.1) becomes to

$$f(\mathbf{Y}, \mathbf{x}, \boldsymbol{\gamma}) \propto f\left(\mathbf{Y} \mid \mathbf{x}\right) \times f(\mathbf{x} \mid \boldsymbol{\gamma}) \times f(\boldsymbol{\gamma}) . \tag{5.3}$$

A hierarchical model specifies the full joint distribution of all quantities in the way of equation Eq. (5.3). The joint distribution on the left side of Eq. (5.3) is provided in terms of three distributions on the right side. Usually the distributions on the right side may be easier to consider individually rather than the entire joint distribution.

Now consider a linear model in the form

$$\mathbf{Y} = \mathbf{Hx} + \boldsymbol{\varepsilon} , \tag{5.4}$$

where $\mathbf{Y}$ is an $n \times 1$ vector of data, $\mathbf{H}$ is an $n \times p$ matrix of known coefficients, $\mathbf{x}$ is a $p \times 1$ vector of process and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors. The elements of $\boldsymbol{\varepsilon}$ are assumed to have zero mean and known variance $\boldsymbol{\Sigma}$. In addition, the uncertainty in $\mathbf{x}$ is expressed as a random deviation $\boldsymbol{\delta}$ from parameter $\boldsymbol{\gamma}$

$$\mathbf{x} = \boldsymbol{\gamma} + \boldsymbol{\delta} , \tag{5.5}$$

where $\boldsymbol{\gamma}$ is an $n \times 1$ vector of parameter and $\boldsymbol{\delta}$ is an $n \times 1$ vector of random errors. The elements of $\boldsymbol{\delta}$ are assumed to have zero mean and known variance $\boldsymbol{\Phi}$. And finally parameter $\boldsymbol{\gamma}$ is described as prior distribution $f(\boldsymbol{\gamma})$.

Assume the random errors $\boldsymbol{\varepsilon}$, $\boldsymbol{\delta}$ and the prior for $\boldsymbol{\gamma}$ are all normally distributed, the three-stage hierarchy is

$$\mathbf{Y} \mid \mathbf{x} \sim MVN(\mathbf{Hx}, \Sigma),$$

$$\mathbf{x} \mid \boldsymbol{\gamma} \sim MVN(\boldsymbol{\gamma}, \Phi),$$

$$\boldsymbol{\gamma} \sim MVN(\boldsymbol{\mu}_0, \mathbf{V}_0).$$

The joint posterior distribution of $\mathbf{x}$ and $\boldsymbol{\gamma}$ can be derived using Bayesian method and inferences will be made from the derived posterior distribution.

5.1.2 Problem Description

The traffic demand is assumed to follow the stochastic demand pattern, which includes the regular traffic demand pattern (the long-term demand information under normal conditions) and the random fluctuations (the inherent stochastic nature and any effect of unobserved factors). This chapter first introduces a hierarchical linear model for stochastic route flows $\mathbf{x}$ in a traffic network. The stochastic nature of route flow $\mathbf{x}$ is described by a probability distribution with two parameters mean $\boldsymbol{\gamma}$ and variance-covariance $\boldsymbol{\Phi}$. The mean $\boldsymbol{\gamma}$ is to describe the regular demand information, and the variance-covariance $\boldsymbol{\Phi}$ describes the day-to-day variations inherent in route flows. This chapter assumes the variance-covariance $\boldsymbol{\Phi}$ is known. The prior knowledge for the route flow mean $\boldsymbol{\gamma}$ is described as a multivariate normal distribution (*MVN*) with the mean $\boldsymbol{\mu}_0$ and the variance-covariance $\mathbf{V}_0$. The reason for the choice of prior distribution for $\boldsymbol{\gamma}$ is that *MVN* is the conjugate prior distribution for the parameter of mean. The noisy sensors in use introduce the measurement errors from counting devices and the measurement errors are randomly distributed with a mean of zero and a known variance $\boldsymbol{\Sigma}$. $\boldsymbol{\Sigma}$ is determined by the reliability of the measuring device. Therefore, the mean of route flows $\boldsymbol{\gamma}$ is to be estimated based on link flow observations from a subset of links $\mathbf{Y}$ by using Bayesian estimation approach.

The second task of this chapter is to derive a decision model for the location of limited number of noisy sensors — the observations from which will be used in the hierarchical Bayesian estimation model. The objective for locating sensors is to maximize the expected uncertainty reductions by hierarchical Bayesian estimator when using the observed traffic data from the candidate set of links.

5.1.3 Notation

The notation is first defined in this chapter. Similar to previous chapters, the notation is classified as it relates to (a) network topology, (b) definition of route flows, (c) links flows, (d) observations or measurements, and (e) decision variables.

Network Topology Parameters:

R: Route-choice set for a network

A: Set of all the links in the network

A': Set of links where sensors are located

A_f: Set of links feasible for sensors to be located

$|R|$: Number of routes in a network

$|A|$: Number of links in a network

N: Number of sensors to locate

$\rho_a^i \in (0,1)$: Link-route parameter. $\rho_a^i = 1$if route $i$uses link a, otherwise,$\rho_a^i = 0$

$$\mathbf{H} = \begin{bmatrix} \rho_1^1 & \cdots & \rho_1^{|R|} \\ \vdots & \ddots & \vdots \\ \rho_{|A|}^1 & \cdots & \rho_{|A|}^{|R|} \end{bmatrix}$$

Route Flows:

x^i: Real route flow of i^{th} route

$\mathbf{x} = (x^1, x^2, \ldots, x^{|R|})'$

γ^i: The parameter to describe the mean of route flow x^i

$\boldsymbol{\gamma} = [\gamma^1, \gamma^2, \ldots, \gamma^{|R|}]$

δ^i: The parameter to describe the variation of route flow x^i (random deviation of x^i fromγ^i)

$\boldsymbol{\delta} = (\delta^1, \delta^2, \ldots, \delta^{|R|})'$

ϕ_{ij}: Covariance of flows between routes i and j(variance of route flow i if $i = j$)

$$\boldsymbol{\Phi} = \begin{bmatrix} \phi_{11} & \cdots & \phi_{1|R|} \\ \vdots & \ddots & \vdots \\ \phi_{|R|1} & \cdots & \phi_{|R||R|} \end{bmatrix}$$

μ_0^i: Mean of prior distribution for parameterγ^i

$\boldsymbol{\mu_0} = (\mu_1^{(0)}, \mu_2^{(0)}, \ldots, \mu_{|R|}^{(0)})'$

$v_{ij}^{(0)}$: Covariance of prior distribution between parameterγ^i and γ^j (variance of prior distribution if $i = j$)

$$\mathbf{V_0} = \begin{bmatrix} v_{11}^{(0)} & \cdots & v_{1|R|}^{(0)} \\ \vdots & \ddots & \vdots \\ v_{|R|1}^{(0)} & \cdots & v_{|R||R|}^{(0)} \end{bmatrix}$$

μ_1^i: Mean of posterior distribution for parameterγ^i

$\boldsymbol{\mu}_1 = (\mu_1^{(1)}, \mu_2^{(1)}, \dots, \mu_{|R|}^{(1)})'$

$v_{ij}^{(1)}$: Covariance of posterior distribution between parameterγ^i and γ^j(variance of posterior distribution if *i* =*j*)

$$\mathbf{V}_1 = \begin{bmatrix} v_{11}^{(1)} & \cdots & v_{1|R|}^{(1)} \\ \vdots & \ddots & \vdots \\ v_{|R|1}^{(1)} & \cdots & v_{|R||R|}^{(1)} \end{bmatrix}$$

Link Flows:

v_a: Real link flow on link *a*

$\mathbf{v} = (v_1, v_2, \dots v_{|A|})'$

$\hat{v}_a$: Potential link flow observation if a sensor is locatedon link *a*

$\hat{\mathbf{v}} = (\hat{v}_1, \hat{v}_2, \dots \hat{v}_{|A|})'$

τ_a^2: Variance of link flow measurement on link *a*

$$\boldsymbol{\Sigma} = \begin{bmatrix} \tau_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tau_{|A|}^2 \end{bmatrix}$$

ε_a: Measurement error on link *a*

$\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{|A|}]$

Observations by sensors:

y_n: Link flow observed by sensor *n*

$\mathbf{Y} = (y_1, y_2, \dots, y_N)'$

g_n: An instance of link flow observed by sensor *n*

$\mathbf{g} = (g_1, g_2, \dots, g_N)'$

φ_n^2: Variance of observation by sensor *n*

Decision Variables for Location Models:

$y_a \in (0,1)$: $y_a = 1$ if locating sensor on link a, otherwise $y_a = 0$ (used for existing location models)

$z_{na} \in (0,1)$: $z_{na} = 1$if locating n^{th} sensor on link a, otherwise $z_{na} = 0$.

$$\mathbf{z} = \begin{bmatrix} z_{11} & \cdots & z_{1|A|} \\ \vdots & \ddots & \vdots \\ z_{N1} & \cdots & z_{N|A|} \end{bmatrix}$$

$$\mathbf{z}_n = \left[z_{n1}, \dots, z_{n|A|} \right]$$

5.2 Hierarchical Linear Bayesian Model of Stochastic Route Flows

5.2.1 Route Flow Mean Estimation with Observations from Multiple Links

The stochastic route flow demand will be modeled by a three-stage hierarchical Bayesian model. The first hierarchy describes the relationship between link flow observations $\hat{\mathbf{v}}$ and route flows $\mathbf{x}$ as

$$\hat{\mathbf{v}} = \mathbf{Hx} + \boldsymbol{\varepsilon}. \tag{5.6}$$

Eq. (5.6) indicates a linear relationship between $\hat{\mathbf{v}}$ and $\mathbf{x}$. A random measurement error ε_a is associated with each link observation from the sensor. We assume ε_a is from a normal distribution $N(0, \tau_a^2)$ where τ_a^2 is known as the reliability of the device. When observations on different links are independent, we have $\boldsymbol{\varepsilon} \sim MVN(0, \Sigma)$, where Σ is a diagonal "dispersion" matrix and the diagonal elements are τ_a^2 for link a.

The linear relationship Eq. (5.6) only exists for the link set *A'*, where sensors are located. Hence, the matrix multiplication $\mathbf{z}\hat{\mathbf{v}}$ identifies the links with sensors, and Eq. (5.6) can be modified to Eq. (5.7) as

$$\mathbf{Y} = \mathbf{z}\hat{\mathbf{v}} = \mathbf{z}(\mathbf{Hx} + \boldsymbol{\varepsilon}) = \mathbf{zHx} + \mathbf{z}\boldsymbol{\varepsilon}. \tag{5.7}$$

In the first hierarchy, the distribution of link flow observation $\mathbf{Y}$ given $\mathbf{z}$, $\mathbf{H}$, $\mathbf{x}$ and Σ, which defines the likelihood, is $MVN(\mathbf{zHx}, \mathbf{z}\Sigma\mathbf{z}')$.

The second hierarchy is to model the stochastic feature in route demand $\mathbf{x}$. By defining two parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ to describe the mean and variation in route flow demand, $\mathbf{x}$ is expressed as

$$\mathbf{x} = \boldsymbol{\gamma} + \boldsymbol{\delta} . \qquad (5.8)$$

The random deviation term $\boldsymbol{\delta}$ is assumed to have *MVN* distribution with zero mean and known variance Φ, the element Φ_{ij} of which, represents the covariance between route i and j. In the second hierarchy, the hierarchical prior distribution for $\mathbf{x}$ given $\boldsymbol{\gamma}$ and Φ is $\mathbf{x} \sim MVN(\boldsymbol{\gamma}, \Phi)$. The parameter representing route flow means $\boldsymbol{\gamma}$ is the parameter to be estimated.

In the final hierarchy, the prior distribution of $\boldsymbol{\gamma}$ is $MVN(\boldsymbol{\mu}_0, \mathbf{V}_0)$, which $\boldsymbol{\mu}_0$ and $\mathbf{V}_0$ are known and to be updated from observation data $\mathbf{Y}$ by Bayesian method.

To sum up, the three-stage hierarchy to model the stochastic route flow is

$$\mathbf{Y} \mid \mathbf{x} \sim MVN(\mathbf{zHx}, \mathbf{z}\Sigma\mathbf{z}'),$$

$$\mathbf{x} \mid \boldsymbol{\gamma} \sim MVN(\boldsymbol{\gamma}, \Phi),$$

$$\boldsymbol{\gamma} \sim MVN(\boldsymbol{\mu}_0, \mathbf{V}_0).$$

Maher (1983) indicated that multivariate normal seems to be the accurate approximation for random variables concerned with counts, so that it is an appropriate choice for the distribution of $\boldsymbol{\gamma}$ and the random errors $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ in order to produce mathematical tractable results.

Following from these assumptions, the posterior distribution of $\boldsymbol{\gamma}$ is also *MVN*. This will be shown in the following theorem:

Theorem. If $\mathbf{Y = zHx + z\varepsilon}$ *where* $\boldsymbol{\varepsilon} \sim MVN(0, \Sigma)$, *then* $\mathbf{Y} \mid \mathbf{x}, \mathbf{z} \sim MVN(\mathbf{zHx}, \mathbf{z}\Sigma\mathbf{z}')$, *and the conditional distribution* $\mathbf{x} \mid \boldsymbol{\gamma} \sim MVN(\boldsymbol{\gamma}, \Phi)$ *is the hierarchical prior distribution for* $\mathbf{x}$ *given* $\boldsymbol{\gamma}$ *and the prior distribution for* $\boldsymbol{\gamma}$ *is* $MVN(\boldsymbol{\mu}_0, \mathbf{V}_0)$. *It follows the Bayes' Theorem then the posterior distribution of* $\boldsymbol{\gamma}$ *given observation* $\mathbf{Y}$ *is also MVN.*

When substituting Eq. (5.8) into Eq. (5.7), it gives

$$\mathbf{Y = zHx + z\varepsilon = zH\gamma + (zH\delta + z\varepsilon)}. \qquad (5.9)$$

Replacing $\mathbf{zH}$ by $\mathbf{A}$ and $\mathbf{(zH\delta + z\varepsilon)}$ by $\mathbf{B}$, Eq. (5.9) corresponds to the standard form of linear model

$$\mathbf{Y = A\gamma + B}. \qquad (5.10)$$

Since $\mathbf{zH\delta} \sim MVN(\mathbf{0}, \mathbf{zH\Phi H'z'})$ and $\mathbf{z\varepsilon} \sim MVN(\mathbf{0}, \mathbf{z\Sigma z'})$ (see Mardiaet al. 1979 for details), the error term in Eq. (5.10) $\mathbf{B = zH\delta + z\varepsilon}$ is distributed as

$$\mathbf{zH\delta + z\varepsilon} \sim MVN(\mathbf{0}, \mathbf{zH\Phi H'z' + z\Sigma z'}).$$

Then the distribution of the data $\mathbf{Y}$ given parameter $\boldsymbol{\gamma}$ is $MVN(\mathbf{zH\gamma}, \mathbf{zH\Phi H'z' + z\Sigma z'})$, which gives the likelihood function as

$$f(\mathbf{Y} \mid \boldsymbol{\gamma}) \propto \exp\{-(\mathbf{Y - zH\gamma})'(\mathbf{zH\Phi H'z' + z\Sigma z'})^{-1}(\mathbf{Y - zH\gamma})/2\}. \qquad (5.11)$$

Replacing $\mathbf{zH\Phi H'z' + z\Sigma z'}$ by $\mathbf{W}$, Eq. (5.11) is simplified as

$$f(\mathbf{Y} \mid \boldsymbol{\gamma}) \propto \exp\{-(\mathbf{Y - A\gamma})'\mathbf{W}^{-1}(\mathbf{Y - A\gamma})/2\}. \qquad (5.12)$$

Because the prior distribution of $\boldsymbol{\gamma}$ be $MVN(\boldsymbol{\mu}_\mathbf{0}, \mathbf{V}_\mathbf{0})$, it gives

$$f(\boldsymbol{\gamma}) \propto \exp\{-(\boldsymbol{\gamma} - \boldsymbol{\mu}_\mathbf{0})'\mathbf{V}_\mathbf{0}^{-1}(\boldsymbol{\gamma} - \boldsymbol{\mu}_\mathbf{0})/2\}. \qquad (5.13)$$

Because *MVN* is the conjugate prior distribution, by Bayes' Theorem $f(\boldsymbol{\gamma} \mid \mathbf{Y}) \propto f(\mathbf{Y} \mid \boldsymbol{\gamma}) f(\boldsymbol{\gamma})$, the posterior distribution is $\boldsymbol{\gamma} \mid \mathbf{Y} \sim MVN(\boldsymbol{\mu}_1, \mathbf{V}_1)$.

From the derivation in chapter 4, the updating equations for $\boldsymbol{\mu}_1$ and $\mathbf{V}_1$ are

$$\boldsymbol{\mu}_1 = \mathbf{V_1}(\mathbf{V_0^{-1}}\boldsymbol{\mu}_\mathbf{0} + \mathbf{A'W^{-1}Y}), \quad (5.14)$$

$$\mathbf{V_1} = (\mathbf{V_0^{-1}} + \mathbf{A'W^{-1}A})^{-1}, \quad (5.15)$$

or

$$\boldsymbol{\mu}_\mathbf{1} = \boldsymbol{\mu}_\mathbf{0} + \mathbf{V_0A'(W + AV_0A')^{-1}(Y - A}\boldsymbol{\mu}_\mathbf{0}), \quad (5.16)$$

$$\mathbf{V}_1 = \mathbf{V_0} - \mathbf{V_0A'(W + AV_0A')^{-1}AV_0}. \quad (5.17)$$

Substituting $\mathbf{A = zH}$ and $\mathbf{W = zH\Phi H'z' + z\Sigma z'}$ in Eq. (5.14) and Eq. (5.15) gives alternative updating equations as

$$\boldsymbol{\mu}_\mathbf{1} = \mathbf{V_1}(\mathbf{V_0^{-1}}\boldsymbol{\mu}_\mathbf{0} + \mathbf{H'z'(zH\Phi H'z' + z\Sigma z')^{-1}Y}), \quad (5.18)$$

$$\mathbf{V_1} = (\mathbf{V_0^{-1}} + \mathbf{H'z'(zH\Phi H'z' + z\Sigma z')^{-1}zH})^{-1}. \quad (5.19)$$

Substituting $\mathbf{A = zH}$ and $\mathbf{W = zH\Phi H'z' + z\Sigma z'}$ in Eq. (5.16) and Eq. (5.17) gives

$$\boldsymbol{\mu}_\mathbf{1} = \boldsymbol{\mu}_\mathbf{0} + \mathbf{V_0H'z'(z\Sigma z' + zH\Phi H'z' + zHV_0H'z')^{-1}(Y - zH}\boldsymbol{\mu}_\mathbf{0}), \quad (5.20)$$

$$\mathbf{V_1} = \mathbf{V_0} - \mathbf{V_0H'z'(z\Sigma z' + zH\Phi H'z' + zHV_0H'z')^{-1}zHV_0}. \quad (5.21)$$

Similar to the updating equation in Eq. (4.13) for deterministic route flow estimation, Eq. (5.19) shows the similar format that the precision of posterior (inverse of variance-covariance) is the summation of prior precision and precisions related with the linear measurement models. The variances in the first and second hierarchies are combined with equal importance in the way of $\mathbf{z\Sigma z' + zH\Phi H'z'}$ to produce the total variances related with link flow measurements from candidate sensors' location $\mathbf{z}$ and the route flow variances which have been observed by sensors **z**. Hence, the posterior variances when using link data from candidate set **z** is determined by the uncertainty of sensors $\mathbf{z\Sigma z'}$, the inherent variance in route flows $\mathbf{zH\Phi H'z'}$, and the prior variances $\mathbf{zHV_0H'z'}$.

Updating equation Eq. (5.18) expresses the posterior mean of parameter γ as a matrix-weighted average of the prior mean $\boldsymbol{\mu}_0$ and the observation data **Y**. If the prior information is more reliable (variance of prior information is lower than the variance in the linear measurement model), then the prior mean μ_0 should contribute more when producing the posterior means. Otherwise, the observations **Y** should have more weight when calculating posterior means.

In addition, rearranging Eq. (5.21) will give the change of variances in route flow mean estimates due to the Bayesian estimator, such that

$$\mathbf{V_1 - V_0 = -V_0H'z'(z\Sigma z' + zH\Phi H'z' + zHV_0H'z')^{-1} zHV_0}. \qquad (5.22)$$

In Eq. (5.22), because $\mathbf{z\Sigma z' + zH\Phi H'z' + zHV_0H'z'}$ is positive definite, $\mathbf{(z\Sigma z' + zH\Phi H'z' + zHV_0H'z')^{-1}}$ is also positive definite. Therefore, the variances in posterior are always reduced comparing to prior variances.

5.2.2 Route Flow Mean Estimation with Observations from Single Link

Define the single observation by the *n*th sensor as $\mathbf{Y_n}$, and $\mathbf{z_n}$ indicates the location of such sensor *n*. The posterior updating equations from single observation $\mathbf{Y_n}$ are revised from Eq. (5.20) and Eq. (5.21) as:

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \frac{\mathbf{V_0H'z_n'}}{\mathbf{z_n\Sigma z_n' + z_nH(V_0 + \Phi)H'z_n'}}(\mathbf{Y_n - z_nH}\boldsymbol{\mu}_0), \qquad (5.23)$$

$$\mathbf{V_1} = \mathbf{V_0} - \frac{\mathbf{V_0 H' z_n' z_n H V_0}}{\mathbf{z_n \Sigma z_n'} + \mathbf{z_n H (V_0} + \mathbf{\Phi) H' z_n'}} \ . \tag{5.24}$$

The elements $\mu_i^{(1)}$ and $\eta_{ij}^{(1)}$ (i=1,…,|R|; j=1,…,|R|) of $\boldsymbol{\mu_1}$ and $\mathbf{V_1}$ are given by

$$\mu_i^{(1)} = \mu_i^{(0)} + \frac{\sum_{j=1}^{|R|} h_n^j \eta_{ij}^{(0)}}{\varphi_n^2 + \sum_{i=1}^{|R|} h_n^i \left(\sum_{j=1}^{|R|} h_n^j (\eta_{ij}^{(0)} + \phi_{ij}) \right)} (Y_n - \sum_{r=1}^{|R|} h_n^r \mu_r^{(0)}) \tag{5.25}$$

$$\eta_{ij}^{(1)} = \eta_{ij}^{(0)} - \frac{\left(\sum_{j=1}^{|R|} h_n^j \eta_{ij}^{(0)} \right) \left(\sum_{i=1}^{|R|} h_n^i \eta_{ij}^{(0)} \right)}{\varphi_n^2 + \sum_{i=1}^{|R|} h_n^i \left(\sum_{j=1}^{|R|} h_n^j (\eta_{ij}^{(0)} + \phi_{ij}) \right)} \tag{5.26}$$

The updating equations Eq. (5.25) and Eq. (5.26) could be iteratively applied using only one observation data in each step.

5.3 Sensor Location Model for Stochastic Route Flow Mean Estimation

When the lower level of *SLFE* problem is a Bayesian estimator, the upper level sensors' location becomes a Bayesian experimental design problem that the statistical inference about the route flows can be improved by appropriately selecting which subset of links to be located with sensors. Decision theory provides a mathematical foundation for the optimal designs, particularly in this dissertation for decisions on sensor locations for data collection. In order to conduct a more informative experiment, prior information, observational studies or subjective beliefs from observations can be valuable. The Bayesian approach to experimental design provides a way to incorporate all information into the upper level of *SLFE* problem.

According to Lindley (1972), the two-part decision theoretic approach involves specification of a utility function which reflects the purpose and costs of the experiment. The Bayesian solution is to find the best location of sensors to maximize the expected utility. In the estimation stage, given the observed data, the best estimates are found to maximize the posterior expected utility, where the expectation is taken with respect to the posterior distribution of route flows and properly reflects uncertainty in estimates *after* observations are collected. Assuming $\mathbf{z}$ is the subset of links with sensors located on, $\mathbf{Y}$ is the link count data observed from a sample space $\mathcal{Y}$; d is the decision selected from possible decision rules D to address the terminal goal of the experiment (e.g. obtaining best estimates of route flows); the unknown parameters are $\boldsymbol{\theta}$ and the parameter space is Θ. A general utility function is of the form $U(d, \boldsymbol{\theta}, \mathbf{z}, \mathbf{Y})$ which represents the costs and consequences of using links $\mathbf{z}$ and corresponding observations $\mathbf{Y}$ and route flow parameter $\boldsymbol{\theta}$. The optimization problem in estimation stage is formulated as:

$$\max_d \int_\Theta U(d, \boldsymbol{\theta}, \mathbf{z}, \mathbf{Y})\, p(\boldsymbol{\theta} \mid \mathbf{z}, \mathbf{Y})\, d\boldsymbol{\theta} = U(\mathbf{z}, \mathbf{Y}) . \qquad (5.27)$$

In the sensor location stage, the optimization problem involves finding the best design that maximizes the *pre-posterior* expected utility, which is obtained by integrating posterior expected utility function Eq. (5.27) over possible outcomes in the sample space. The general formulation to find the optimal design for optimal locations $\mathbf{z}$*:

$$U(\mathbf{z}^*) = \max \int_y \max_d \int_\Theta U(d, \boldsymbol{\theta}, \mathbf{z}, \mathbf{Y})\, p(\boldsymbol{\theta} \mid \mathbf{z}, \mathbf{Y})\, p(\mathbf{Y} \mid \mathbf{z})\, d\boldsymbol{\theta}\, d\mathbf{Y} . \qquad (5.28)$$

Usually, when the goal of conducting an experiment is to obtain the point estimate of parameters, a quadratic loss function is an appropriate utility function which leads to the pre-posterior expected utility as

$$U(\mathbf{z}) = -\int_y \int_\Theta (\boldsymbol{\theta} - \boldsymbol{\theta}) \mathbf{A} (\boldsymbol{\theta} - \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \mathbf{Y}, \mathbf{z})\, p(\mathbf{Y} \mid \mathbf{z})\, d\boldsymbol{\theta}\, d\mathbf{Y} \ , \qquad (5.29)$$

where $\mathbf{A}$ is a symmetric nonnegative definite matrix to assign weights to parameters according to different levels of interest. In this case, a design can be chosen to maximize the expected utility in Eq. (5.29). According to Chaloner and Verdinelli (1995), the Bayesian procedure yields an expected utility

$$U(\mathbf{z}) = -tr\{\mathbf{A}\mathbf{V}_1\} \;, \tag{5.30}$$

where $\mathbf{V}_1$ is the posterior variance-covariance matrix in Eq. (5.19) or Eq. (5.21). A design that maximize $U(\mathbf{z})$ in Eq. (5.30) is called Bayesian A-optimality criterion. If $\mathbf{A}$ is the identity matrix (it means that the modeler treats each route with importance), the sensor location problem to provide Bayesian *A*-optimal design is formulated as:

[A-opt_1]

$$\min tr(\mathbf{V}_0^{-1} + \mathbf{H'z'(zH\Phi H'z' + z\Sigma z')^{-1}zH})^{-1} \tag{5.31}$$

$$s.t. \quad \sum_{a=1}^{|A|} z_{na} = 1, n = 1, \dots, N \tag{5.32}$$

$$\sum_{n=1}^{N} z_{na} \leq 1, \forall a \in A \tag{5.33}$$

$$z_{na} \in (0,1), n = 1, \dots, N, a \in A \tag{5.34}$$

The objective function Eq. (5.31) minimizes the total posterior variances, which satisfies the Bayesian A-optimal criterion that the trace of posterior variance matrix is minimized. Constraint Eq. (5.32) ensures that every sensor *n* has to be located on one and only one link. Constraint Eq. (5.33) forces that each link allows at most one sensor to be located on. Constraint Eq. (5.34) is the binary constraint for decision variable z_{na}. Constraints Eq. (5.32) to Eq. (5.34) indicate the design problem is an exact design which the number of observations at each potential point is an integer. Finding optimal exact designs is often a difficult problem (Clyde, 2001). In order to simplify the computational effect to evaluate the objective value, e.g. reducing the number of the matrix inverse operations, the alternative formulation for Bayesian A-optimal design is

[A-opt_2] $$\max tr(\mathbf{V_0 H'z'(z\Sigma z' + zH\Phi H'z' + zHV_0H'z')^{-1} zHV_0}) \quad (5.35)$$

$s.t.$ (5.32) – (5.34)

The interpretation of objective function Eq. (5.35) is to maximize the total variance reduction in posterior distribution of **x** by choosing the sensor locations **z** that provide observation data.

When considering selecting single link from the candidate set, formulation [A-opt_2] can be revised to [A-opt_3] as:

[A-opt_3] $$\max \sum_{i=1}^{|R|} \frac{(S_n^i)^2}{\varphi_n^2 + T_n + Q_n} \quad (5.36)$$

$$s.t. \quad h_n^i = \sum_{a=1}^{|A|} z_{na} \rho_a^i, \quad \forall i = 1, ..., |R| \quad (5.37)$$

$$\varphi_n^2 = \sum_{a=1}^{|A|} z_{na} \tau_a^2 \quad (5.38)$$

$$S_n^i = \sum_{j=1}^{|R|} h_n^j \eta_{ij}^{(0)}, \quad \forall i = 1, ..., |R| \quad (5.39)$$

$$T_n = \sum_{j=1}^{|R|} h_n^j S_n^i \quad (5.40)$$

$$Q_n = \sum_{i=1}^{|R|} h_n^i \left(\sum_{j=1}^{|R|} h_n^j \phi_{ij} \right) \quad (5.41)$$

$$\sum_{a=1}^{|A|} z_{na} = 1 \quad (5.42)$$

$$z_{na} = 0, \forall a \in \bar{A}_f \quad (5.43)$$

$$z_{na} \in (0,1), n = 1, ..., N, a \in A \quad (5.44)$$

The objective function Eq. (5.36) in formulation [A-opt_3] maximizes total route flow variance reduction by selecting only one link from candidate set A_f. Constraints Eq. (5.37) to Eq. (5.41) are introducing notations to simplify objective function Eq. (5.36).

Constraint Eq. (5.42) restricts sensor n will be located on only one link. Constraint Eq. (5.43) forces the candidate location of sensors only from the feasible links' set A_f ($\overline{A_f}$ is the set of links which have been located with sensors and are not available any more). Constraint Eq. (5.44) is the binary constraint for decision variables.

5.4 Algorithm

5.4.1 A Sequential Method for Independent Observations

Instead of solving formulation [A-opt_2] which has an objective function of a complex form, one can iteratively select the location of one sensor by solving [A-opt_3] each time from link set A_f. At the end of the iteration, updating set A_f and the prior variances $\mathbf{V_0}$ to posterior variance estimates which becomes prior $\mathbf{V_0}$ required for the next iteration. The process is repeated until N links are determined for locating the sensors.

To locate only one sensor in the network, formulation [A-opt_3] can be solved by a *greedy algorithm*: for each candidate link, calculate the objective value $\sum_{i=1}^{|R|} \frac{(S_n^i)^2}{\varphi_n^2 + T_n + Q_n}$ and select the one with largest objective value. The solution of [A_opt_3] can be found by $\mathcal{O}(|A| \cdot |R|^2)$ time.

Therefore, the sequential method for solving [A-opt_2] is as follows:

BAYESIAN BASED SEQENTIAL LOCATION ALGORITHM [SEQ]

Input: $\mathbf{V}_0$, $\mathbf{H}$, N

Output: $\mathbf{z}=(\mathbf{z}_1,\mathbf{z}_2,...,\mathbf{z}_N)'$, $\mathbf{V}_1$

(1) Set $n:=1$.

Set $A_f = A$.

Set $A' = \varnothing$.

(2) For each link $a \in A_f$, calculate $\sum_{i=1}^{|R|}\frac{(S_n^i)^2}{\varphi_n^2+T_n+Q_n}$.

(3) Select the link $a \in A_f$ with largest value calculated in Step 2 and $\mathbf{z_n}$ is a row vector with only non-zero element in *a*th column. Remove link *a* from A_f and add link *a* into A'.

(4) Using the observation on link *a* to update posterior variance $\mathbf{V}_1$ by Eq. (5.24).

(5) Increment *n* by 1.

(6) If $n = N+1$ then stop.

Else set $\mathbf{V}_0 = \mathbf{V}_1$ and go to Step 2.

The algorithm runs for N loops and within each loop the running time is polynomial $\mathcal{O}(|A|\cdot|R|^2)+\mathcal{O}(1)+\mathcal{O}(|R|^2)=\mathcal{O}(|A|\cdot|R|^2)$ (preprocessing in step 2 runs for $\mathcal{O}(|A|\cdot|R|^2)$, searching for largest value of variance reduction in step 3 runs for

($\mathcal{O}(|A|)$), and updating for posterior mean and variance-covariance runs for $\mathcal{O}(|R|^2)$). The complexity for this optimality algorithm is $\mathcal{O}(N \cdot |A| \cdot |R|^2)$.

This method is same as the algorithm developed and implemented in Chapter 4, which has been proven to provide an optimal solution for deterministic route flow estimation with the independent observations (see section 4.4 for details).

The three stage hieratical linear model can be revised to a linear model as below:

$$\mathbf{Y} \,|\, \boldsymbol{\gamma} \sim MVN(\mathbf{zH}\boldsymbol{\gamma}, \mathbf{z}(\mathbf{HWH'} + \boldsymbol{\Sigma})\mathbf{z'})$$

$$\boldsymbol{\gamma} \sim MVN(\boldsymbol{\mu}_0, \mathbf{V}_0)$$

Although the independent assumption for link flow measurement provides a diagonal matrix $\boldsymbol{\Sigma}$, the entire variance term in likelihood distribution Eq. (5.45) is diagonal only if when the route flows $\mathbf{x}$ are independent and the variance-covariance matrix $\boldsymbol{\Phi}$ is diagonal. Therefore, the above sequential procedure gives optimal solution with the assumption that route flows $\mathbf{x}$ are independent and link flow measurement errors are independent as well. When the route flow $\mathbf{x}$ are dependent, such sequential procedure can be used to provide high-quality near-optimal solution.

5.4.2 A Greedy Heuristic for Independent Observations

Because the method in section 5.4.1 is not efficient for large networks, another greedy heuristic can be used to find a solution within reasonable computational time. This greedy heuristic selects all the N links at one time based on the value of variance reduction by single link. The heuristic is as follows:

BAYESIAN BASED GREEDY LOCATION HEURISTIC [GRE]

Input: $\mathbf{V}_0$, $\mathbf{H}$, N

Output: $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_N)'$

(1) For each link $a \in A$, calculate $\sum_{i=1}^{|R|} \frac{(S_n^i)^2}{\varphi_n^2 + T_n + Q_n}$.

(2) Sort the links according to the values calculated in step (1).

(3) Select the first N links with largest values calculated in step (1).

The heuristic will operate the three steps without any looping. The run time in step (1) is $\mathcal{O}(|A| \cdot |R|^2)$; the run time in step (2) depends on the sorting algorithms in use (e.g. the quicksort algorithm can be run in $\mathcal{O}(|A|^2)$); and the last step is conducted in $\mathcal{O}(1)$. The total run time for the heuristic is $\mathcal{O}(|A| \cdot |R|^2)$.

5.4.3 A Sequential Method for Dependent Observations

When the link flow observations are dependent and the correlations between some pairs of link measurements are substantial, the previous proposed methods (SEQ and GRE) are not accurate because they do not accurately calculate the covariance between measurements. For this case, the sequential method is adjusted to consider the covariance between pairs of measurements. Initially the solution link set is empty. In each iteration, one link is added to the solution link set which has been established by previous iterations. The criterion to select a link is that the variance reduction from prior variance $\mathbf{V}_0$ by this additional link (together with the existing link set) is minimized. Unlike the sequential procedure in section 5.4.1, this method does not update the variance-covariance matrix iteratively. However, the approach for link selection is greedy because the solution in each step is a local optimum.

BAYESIAN BASED SEQUENCIAL LOCATION HEURISTIC FOR DEPENDENT OBSERVATIONS [DSEQ]

Input: $\mathbf{V}_0$, $\mathbf{H}$, N

Output: $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_N)'$

(1) Set $n := 1$.

Set $A_f = A$.

Set $A' = \varnothing$.

(2) For each link $a \in A_f$, using the links $\mathbf{z}$ from set $\{a\} \cup A'$ to calculate the potential variance reduction value $\mathbf{V_0 H'z'(z \Sigma z' + z H \Phi H'z' + z H V_0 H'z')^{-1} z H V_0}$.

(3) Select the link a with largest value calculated in Step 2. Remove link a from A_f and add link a into A'.

(4) Increment n by 1.

(5) If $n = N + 1$ then stop.

Else, go to Step 2.

The approach has the same complexity as the sequential method in section 5.4.1, which is $\mathcal{O}(N \cdot |A| \cdot |R|^2)$.

5.5 Experimental Result

5.5.1 Experiment Setup

We apply the proposed sequential procedure in section 5.4.1 (SEQ) and the greedy heuristic in section 5.4.2 (GRE) to three problem scenarios and the performances are compared with two existing models: Link Flow Coverage (LFC) and Route Flow Coverage (RFC). The reason to select LFC and RFC for comparison is that both models are often studied for deterministic route flow estimation and rely on different sources of information. For each scenario, the baseline is provided by a sensor location model "Random" (Rand), which selects a random set of links to observe. The "Rand" model provides the effect of variance reduction by incrementing the number of sensors. All the experiments are conducted in Matlab.

Figure 11 shows the scheme of the experimental procedure. First (Steps 1-2), problem scenarios are generated by defining the network (supply), loading the assumed traffic flows (demand) and obtaining the "actual" mean of network traffic on routes $\boldsymbol{\gamma}$. The "actual" variance-covariance matrix $\boldsymbol{\Phi}$ can be obtained from other models which are independent with this experiment, for example, a Bayesian model learning the variance-covariance on routes from link observations. By simulating the random perception errors and measurement errors, route flow prior distribution (mean and variance), sensors' reliabilities and potential link flow observations are generated by linking the errors to the actual values (Step 3). The prior's reliability is first randomly generated by $\mathbf{u} \sim U(\mathbf{0}, \mathbf{u}^{\mathbf{max}})$; then the diagonal elements of prior's variance-covariance matrix $\mathbf{V_0}$ is calculated by $\eta_{ii}^{(0)} = \gamma^i u^i$; all covariance are assumed to be zero in the initial priors; the prior's mean is generated from a normal random generator $\mathbf{u_0} \sim N(\boldsymbol{\gamma}, \mathbf{V_0})$. The random deviation from $\boldsymbol{\gamma}$ is generated by $\boldsymbol{\delta} \sim N(\mathbf{0}, \boldsymbol{\Phi})$ and such deviation is added to the "actual" route flow mean values. Hence the "actual" value of link flows is calculated by using the link-route parameter **H** by $\mathbf{v} = \mathbf{H}(\boldsymbol{\gamma} + \boldsymbol{\delta})$. The sensors' reliability **t** is randomly generated from $\mathbf{t} \sim U(\mathbf{0}, \mathbf{t}^{\mathbf{max}})$; and the diagonal elements in variance-covariance matrix of link flow measurement $\boldsymbol{\Sigma}$ can be calculated by $\tau_a^2 = v^a t^a$; a random correlation coefficient for each pair of observations is simulated to produce the covariance $\boldsymbol{\Sigma}$; in the potential observation on such link was generated from a normal random generator $\mathbf{Y} \sim N(\mathbf{v}, \boldsymbol{\Sigma})$. In Step 4, [A-opt_2] and other location models are solved to obtain location decisions**z**. Both sequential algorithm and the greedy method based heuristic proposed in section 5.4 are implemented to solve [A-opt_2]. Then, with these sensor location decisions $\mathbf{z}$, posterior distribution for route flow means is obtained using the Bayesian estimation model in section 5.2 (Step 5). To evaluate each location model (Step 6), the posterior means are compared with actual route flow means in terms of the bias and total variance (trace of $\mathbf{V_1}$). The evaluation criterion is the sum-squared error (*SSE*) of the posterior distribution, which is defined by the summation of variance and squared bias:

$$SSE = BIAS^2 + Variance = \sum_{i \in R} \left(\mu_i^{(1)} - \gamma^i \right)^2 + \sum_{i \in R} \eta_{ii}^{(1)} \qquad (5.45)$$

where γ^i is the true mean value of the route flow *i*. Because our proposed model already ensures the minimum posterior variances, the criterion SSE is sensitive to show the effect of bias provided by the comparative locational models.

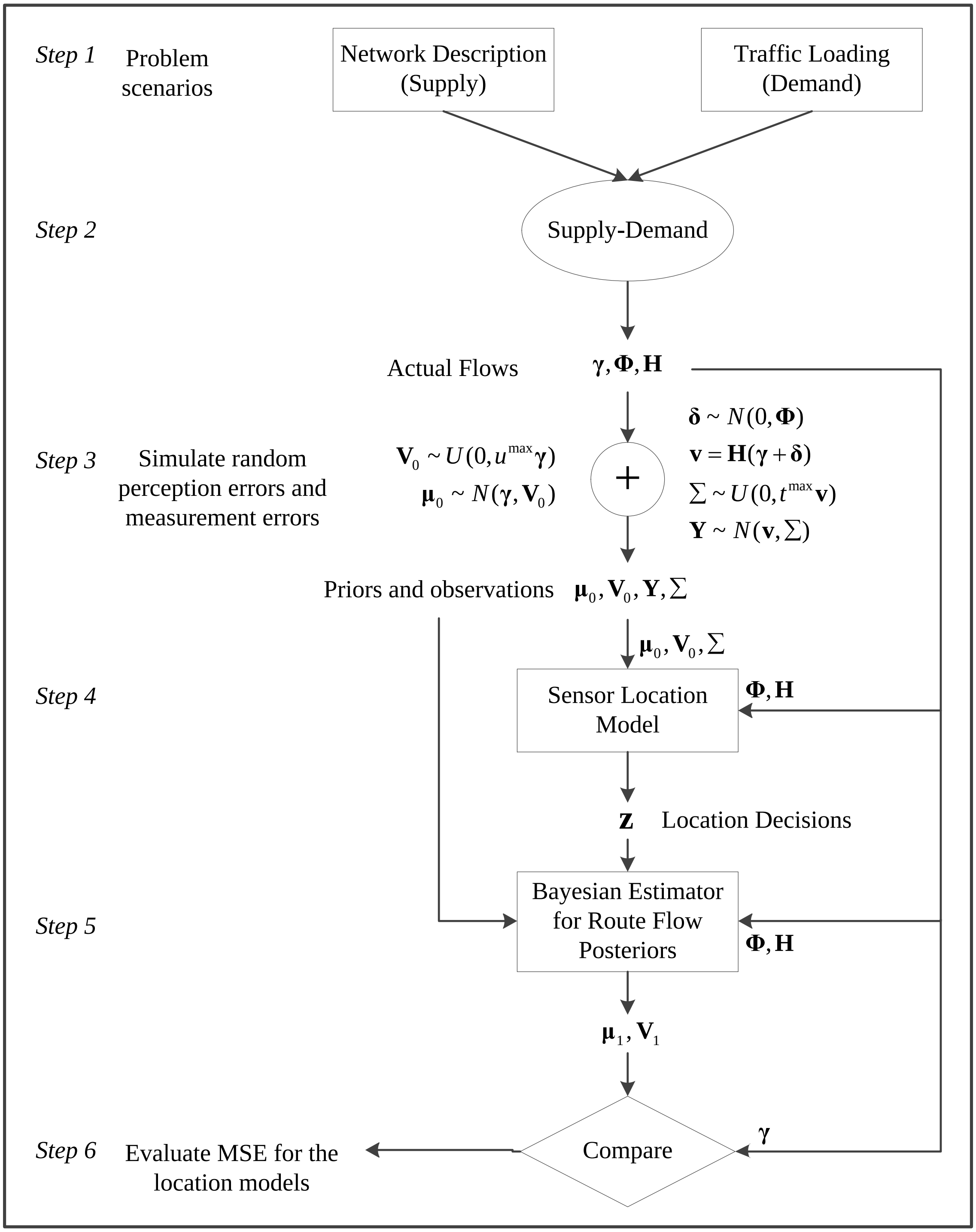

Figure 11 Procedure for the computational experiments for Chapter 5

Finally, note that for each network scenario, we can have different problem instances by locating different number of sensors between 1 and (say) *N*, the number of possible sensor sites. Hence, each problem instance in our computational experiments is defined by:

(1) Network supply

(2) Traffic demand

(3) Reliability (variance) of priors on traffic flows to be estimated

(4) Reliability (variance) of sensors

(5) Number of sensors to be located.

Solving each problem instance by each of the four models, gives us an *SSE* score for each model. Because of the stochastic nature of the experiment framework, for each network supply and traffic demand, we run the experiments several times, each of which uses different priors and observations generated randomly. The average SSEs is calculated in the end to evaluate the performance of four models.

Three sets of network scenarios are used in the computational experiments.

(1) A grid network with 16 OD pairs, 36 routes and 34 links (grid network 1, left side of Figure 7).

(2) A grid network with 16 OD pairs, 145 routes and 69 links (grid network 2, right side of Figure 7).

(3) Houston data used by Mirchandani et al. (2009), with 1768 routes and 468 links with flows.

When generating the experiments, u^{max} is set to be 0.15 and t^{max} is 0.1, which means the maximum measurement error is within 10% deviation from the actual link flow, while the maximum perception error in prior is within 15% off from the actual mean of route flow. This parameter setting reflects that we assume counting devices are generally more accurate than our prior knowledge in practice. We assume that each testing network has unique "real" route flow distribution with parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\Phi}$, and the unique route choice set with link route parameter $\mathbf{H}$. For the first two test networks, 1000 scenarios with different priors and observations are randomly generated within the specific range. Because of the large scale of the last network, only 30 scenarios are generated for experiments. The last network from practice is used to validate the conclusions studied from first two testing networks.

5.5.2 Numerical Results for Independent Measurements

This section shows the numerical results for cases when measurement errors are assumed to be independent. Five sensor location models are first compared using the posterior variances for the route flow means var($\boldsymbol{\gamma} | \mathbf{Y}$) in 1000 randomly generated scenarios. The results are shown in Table 11. Each cell in Table 11 represents the average posterior variances in 1000 scenarios when locating certain number of sensors in the network. For example, the first row (named as "0~20%") averages the posterior variances when locate sensors to cover from zero to twenty percent of links, while the row named as "21%~40%" averages the posterior variances when covering 21% to 40% of links in the network.

Table 11 Posterior Variances from Bayesian Estimation using Different Sensors Location

		SEQ	GRE	LFC	RFC	Rand
Network 1	0~20%	***8.89***	9.05	9.93	9.69	10.22
	21%~40%	***7.05***	7.40	8.27	8.10	8.81
	41%~60%	***6.25***	6.53	7.33	7.19	7.62
	61%~80%	***5.87***	6.01	6.54	6.39	6.66
	81%~100%	***5.71***	5.73	5.93	5.80	5.93
	Total	***33.77***	34.72	38.01	37.17	39.24
network2	0~20%	***2.34***	2.39	2.49	2.51	2.60
	21%~40%	***1.98***	2.05	2.18	2.27	2.33
	41%~60%	***1.82***	1.87	1.97	2.08	2.09
	61%~80%	***1.74***	1.76	1.82	1.90	1.90
	81%~100%	***1.71***	1.71	1.72	1.75	1.75
	Total	***9.59***	9.78	10.18	10.51	10.67

As the baseline, random selection of sensors' location (Rand) performs worst in variance reduction from the results in Table 11. The sequential procedure (SEQ) has the best performance to reduce the variances in posterior route flow means because the average posterior variances are minimal comparing to other models. The greedy heuristic (GRE) is the second best model in terms of variances reduction, and its performance is close to SEQ. The other two comparison models LFC and RFC perform worse than our proposed procedures but can show extra effects in variance reduction comparing to the baseline when using sensors' location from random choices.

Figure 12 and Figure 13 plots the average of relative posterior variances for each of the sensor location models. The x-axis is the number of sensors in use, while the y-axis is the average of relative posterior variances. When assuming the prior variances of route flow means (V_0) as one, and the posterior variances when all the links in the network are covered by sensors (V_N) as zero, any values of posterior variances when using link flow measurements from *i* sensors (V_i) can be scaled as:

$$Relative\ posterior = \frac{v_i - v_N}{v_0 - v_N} \quad (5.46)$$

From Figure 12 and Figure 13, the percentage of variance reduction by Bayesian estimation can be easily illustrated and compared since all the posterior variances have been scaled between zero and one. First, in both figures for networks 1 and 2, the baseline scenario, which is established by randomly selecting the sensors' location, indicates that the reliability of posterior information is increasing when incrementing the number of sensors, and the marginal effect of variance reduction is constant. Therefore, the performances of different location models are shown as how fast the curves decrease in Figure 12 and Figure 13. SEQ gives best performance in variance reductions in both test networks, in particular 80% of the variances are reduced when 40% of the links are measured. The variance reduction by GRE is close to but not as good as SEQ, e.g. 80% of the variance reductions by GRE require additional 10% of link coverage (50% of the links need to be measured). The curves for other comparison model LFC are between the baseline and our proposed procedures (SEQ and GRE). This is because LFC does not directly address the variance reduction objective and does not rely on the reliability of prior information and potential measurement errors. The performance of RFC model is close to but does not have a steady pattern in both test networks compared to LFC. An explanation for this is that due to different network topologies and traffic assignments, RFC model has varied performances. When all routes in a network have been covered by limited number of links and the covered route flows has been maximized, the additional link is selected randomly because the routes covered by such link have already covered. Hence, the effect of such additional links on variance reductions is similar as in baseline scenario. This is why the curves of “Rand” and “RFC” overlap in Figure 12 after more than half of the network is covered.

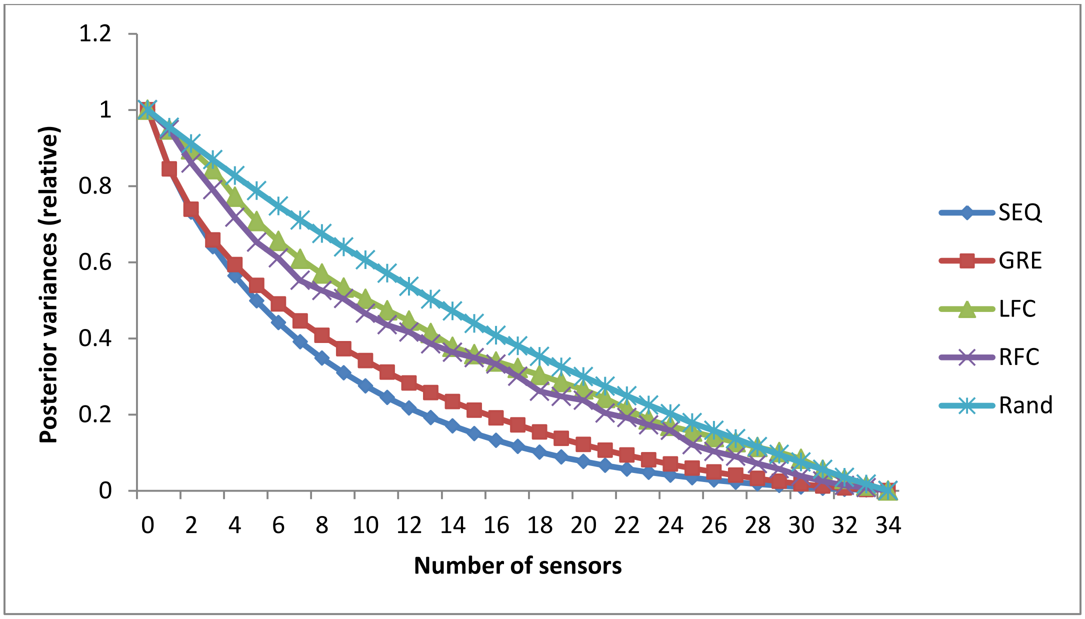

Figure 12 Relative Posterior Variances from Bayesian Estimation when Link Measurements are Independent (Network 1)

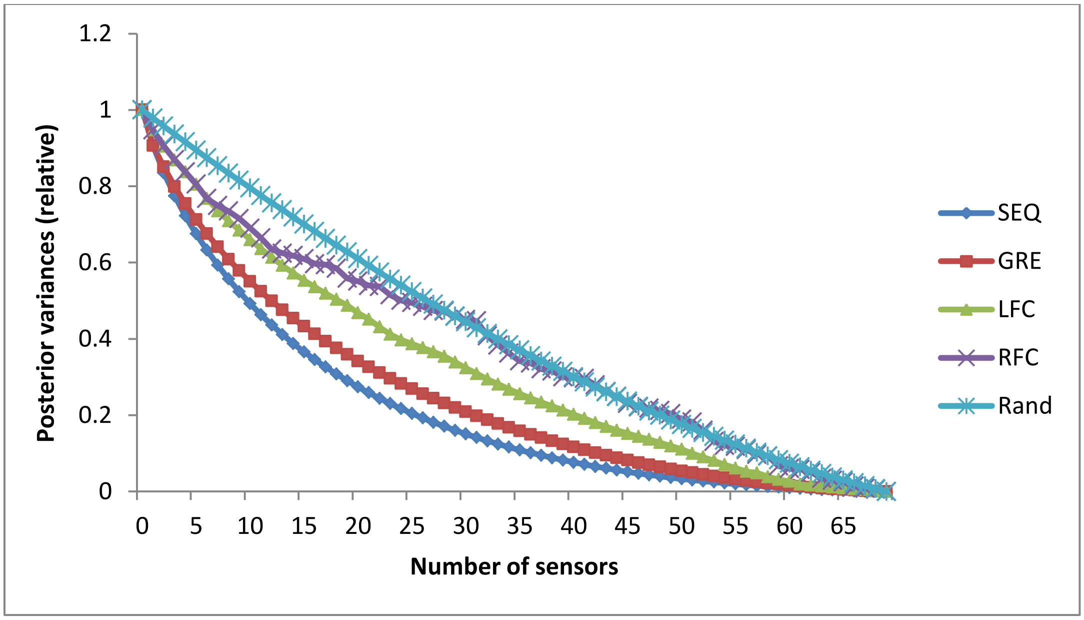

Figure 13 Relative Posterior Variances from Bayesian Estimation when Link Measurements are Independent (Network 2)

Secondly, posterior *SSE* is the evaluation criterion that takes both bias and variance together. The average posterior *SSE*s in 1000 random generated scenarios are shown in Table 12. Similar to Table 11, each row in Table 12 averages the posterior *SSE*s when locating sensors to cover a range of links in a network. From Table 12, SEQ has the best performance in estimating posterior route flow means because the average posterior *SSE*s are minimal comparing to other models. GRE is the second best model in terms of *SSE*s. Rand performs worst, and the other two comparison models LFC and RFC perform better than baseline but worse than our proposed procedures. The result of posterior *SSE* is similar to the result of posterior variances.

Table 12 Posterior SSE for Five Sensor Location Models when Link Measurements are Independent

		SQU	GRE	LFC	RFC	Rand
Network 1	0~20%	***186.82***	193.47	255.64	237.51	279.01
	21%~40%	***133.95***	145.96	198.85	181.12	233.45
	41%~60%	***122.13***	129.56	164.71	157.25	190.16
	61%~80%	***118.65***	121.93	142.92	135.37	158.04
	81%~100%	***118.21***	118.76	124.99	121.00	128.25
	Total	***679.76***	709.69	887.10	832.25	988.91
network2	0~20%	***20.31***	21.14	25.04	25.84	27.68
	21%~40%	***15.66***	16.60	20.68	22.77	23.50
	41%~60%	***14.54***	15.10	17.71	19.64	20.14
	61%~80%	***14.22***	14.40	15.58	16.98	17.29
	81%~100%	***14.08***	14.12	14.21	14.75	15.01
	Total	***78.81***	81.35	93.22	99.97	103.61

Figure 14 and Figure 15 plots the average of relative posterior SSEs for each of the sensor location models. The x-axis is the number of sensors in use, while the y-axis is the average of relative posterior *SSE*s. Same as posterior variances, posterior SSEs are scaled between zero and one.

The patterns in Figures 14 and Figure 15 are similar to observed patterns in Figure 12 and Figure 13. The baseline established by "Rand" model shows constant marginal effects on *SSE*s when locating one sensor at a time in network; other models show faster improvements in the estimation quality of the posterior distributions. Among all the comparison models, SEQ performs best and GRE is the second best model in terms of posterior *SSE*s. LFC and RFC models are significantly worse than our proposed procedures. Again, the non-steady pattern of RFC curves is because of the characteristic of the model.

When comparing the performance of different location models by posterior *SSE*s, our proposed procedures SEQ and GRE show more benefits. 80% of *SSE* reduction can be achieved by covering only 18% of the links in a network (from Figure 14 and 15), while 80% of *SSE* reduction requires 40% of the links to be measured (Figure 12 and Figure 13). Therefore, in practice, in order to obtain better route flow estimates (e.g. a 80% *SSE* reduction), analysts can invest in sensors to provide a 18% of links coverage. In additional, 50% of *SSE* reduction can be achieved by measuring less than10% of the links.

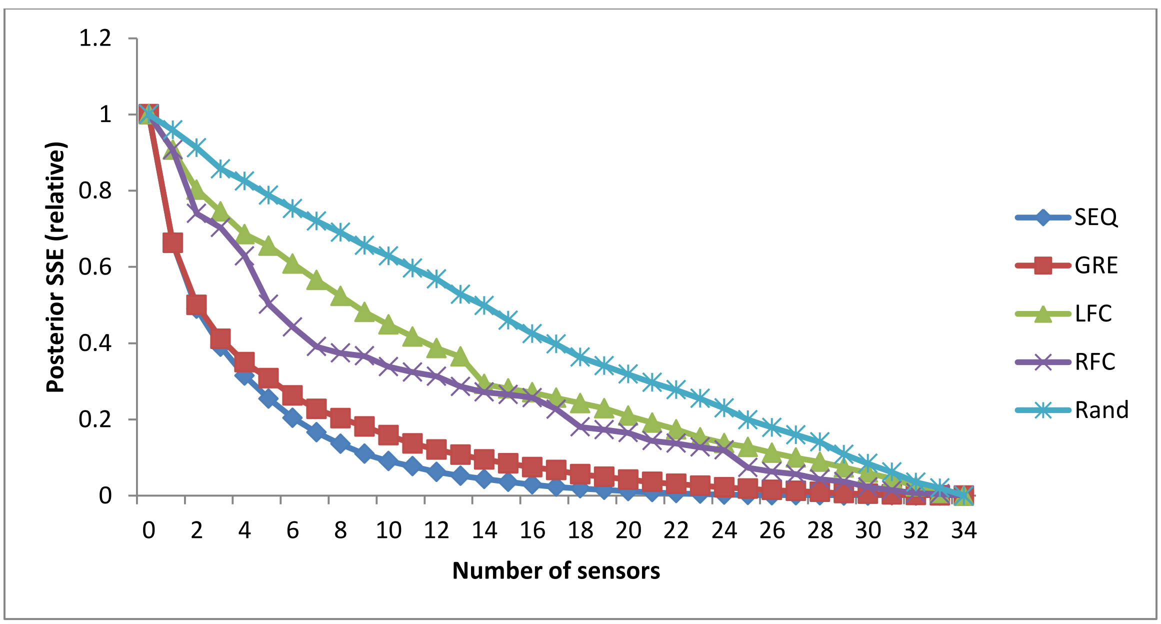

Figure 14 Relative SSE from Bayesian Estimation when Link Measurements are Independent (Network 1)

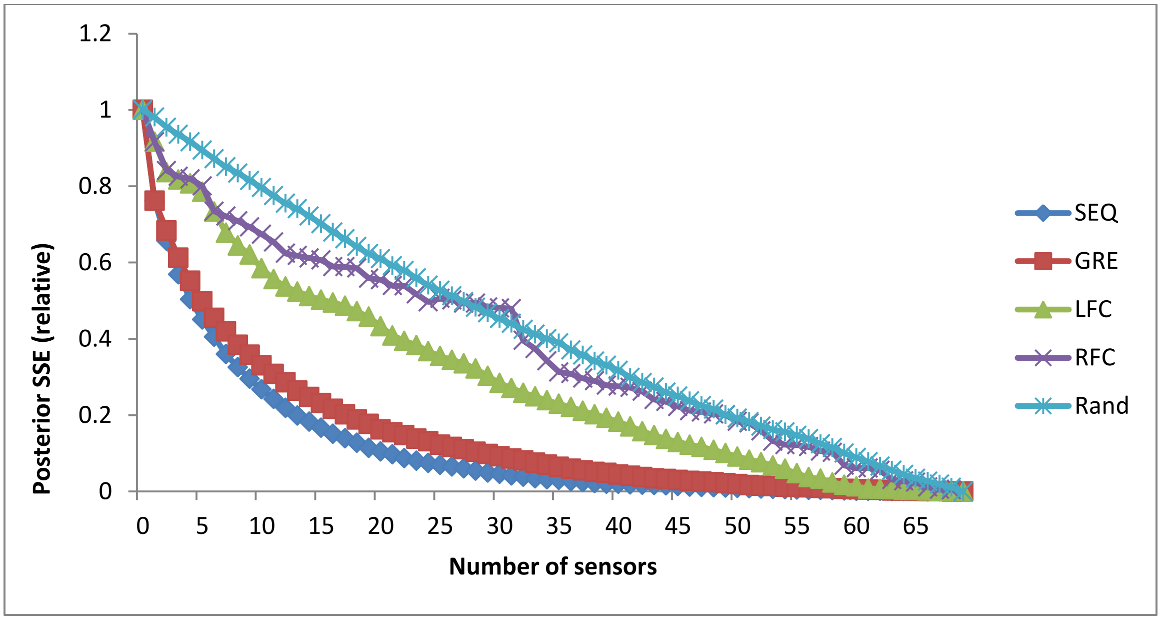

Figure 15 Relative SSE from Bayesian Estimation when Link Measurements are Independent (Network 2)

The last test network is constructed based on a large traffic network in Houston to validate the observations and conclusions obtained from test networks 1 and 2. From previous test results, the performance by GRE is very close to SEQ but computational times are significantly lower. Therefore, we chose to test four models, GRE plus LFC, RFC and Rand, all of which are computationally manageable. Figure 16 plots the relative posterior *SSE*s in 30 random generated scenarios. The patterns in Figure 16 are similar with previous observations. The posterior *SSE*s decrease fastest for the proposed method GRE, for which 90% of *SSE* is reduced by covering only 6% of the links (30 links).

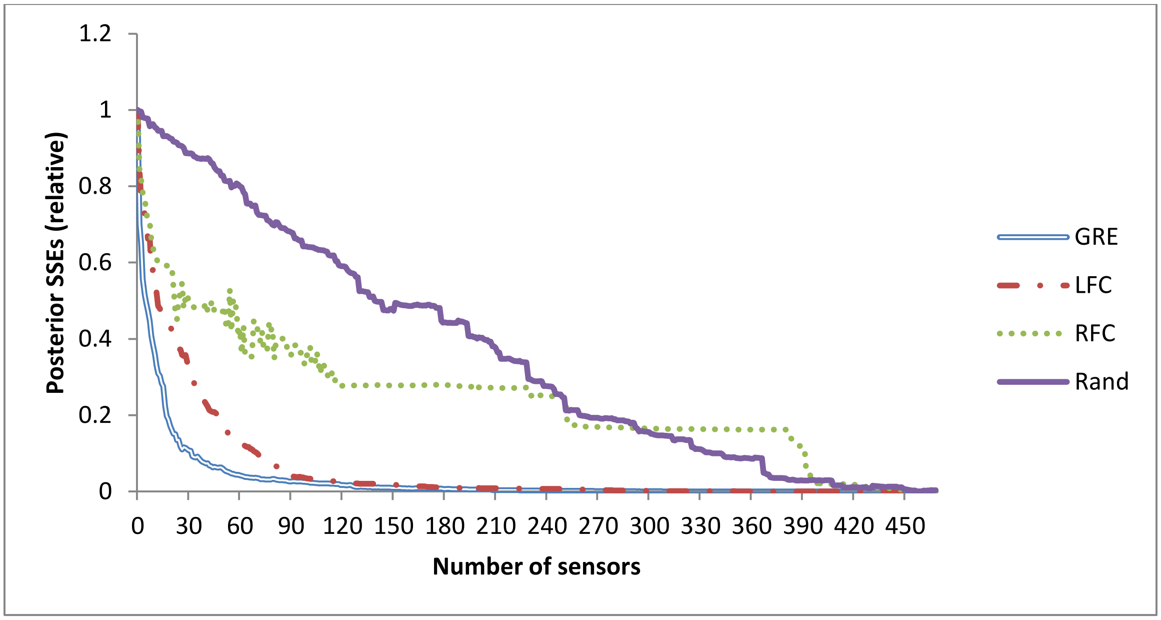

Figure 16 Relative Posterior SSE from Bayesian Estimation when Link Measurements are Independent (Houston Network)

Finally, we focus on the comparison between SEQ and GRE on scenarios where only a small number of sensors are available in Houston network (because when *N* is large, computational time for SEQ becomes very large). The posterior *SSE*s are plotted in Figure 17. It indicates that, in comparing with GRE, the SEQ method further reduces the *SSE* and improves the quality of route flow mean estimates. However, as discussed earlier, the processing time of SEQ is much higher.

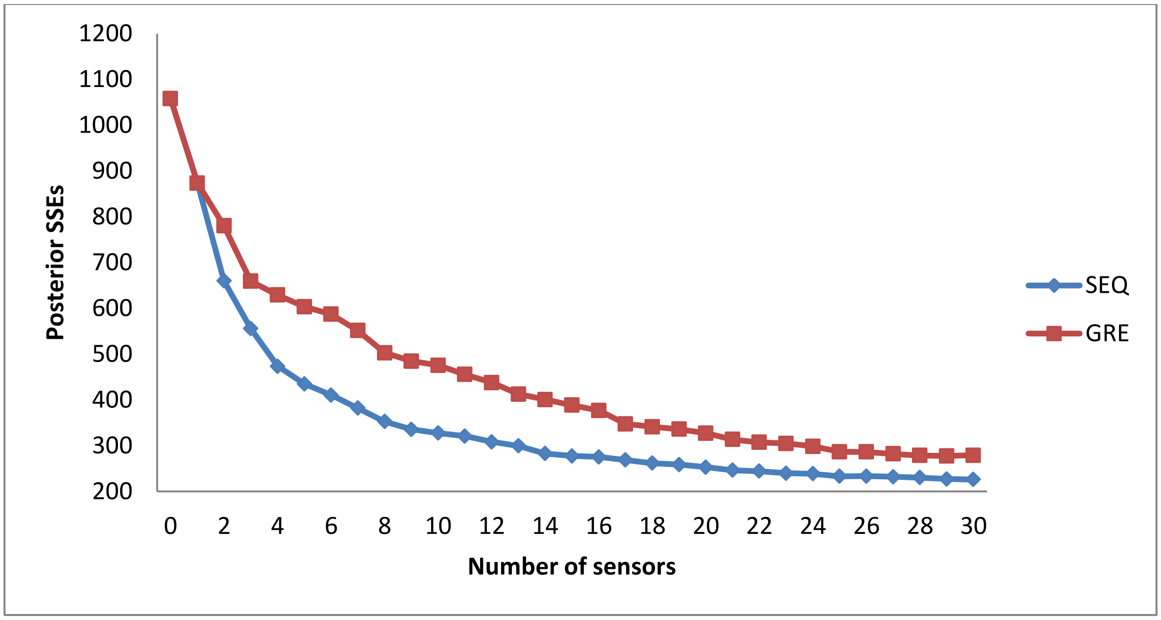

Figure 17 Posterior SSE from Bayesian Estimation when Link Measurements are Independent and Observe a Few Links (Houston Network)

5.5.3 Numerical Results for Dependent Measurements

The results for the experiment instances with dependent measurement are presented in this section. First, the posterior variances and bias for the route flow mean estimates for the four sensor location models averaged in 1000 random generated scenarios are tableted in Table 13. From Table 13, the sequential procedure (DSEQ) has the best performance because the average posterior variances and bias are minimal compared to other models. The random selection of sensors' location (Rand) performs worst.

Table 13 Posterior variances and bias from Bayesian Estimation when Link Measurements are Dependent

		Posterior variances				Posterior bias			
		DSEQ	LFC	RFC	Rand	DSEQ	LFC	RFC	Rand
Network1	0~20%	***8.92***	9.95	9.72	10.25	***163.91***	240.07	217.17	272.78
	21%~40%	***7.07***	8.30	8.13	8.81	***105.69***	164.86	144.96	213.51
	41%~60%	***6.26***	7.35	7.22	7.63	***89.93***	123.78	119.45	165.38
	61%~80%	***5.88***	6.57	6.40	6.68	***83.94***	102.71	96.23	124.11
	81%~100%	***5.72***	5.95	5.82	5.95	***81.89***	87.59	83.58	93.02
	Total	***33.86***	38.11	37.28	39.32	***525.37***	719.00	661.39	868.80
Network2	0~20%	***2.34***	2.49	2.50	2.60	***17.03***	20.42	21.17	24.61
	21%~40%	***1.97***	2.18	2.27	2.32	***11.44***	15.00	17.41	19.43
	41%~60%	***1.81***	1.97	2.08	2.08	***9.81***	12.03	14.37	15.03
	61%~80%	***1.74***	1.82	1.90	1.89	***9.21***	10.26	11.70	11.78
	81%~100%	***1.70***	1.72	1.74	1.75	***9.02***	9.09	9.54	9.66
	Total	***9.57***	10.18	10.49	10.64	***56.50***	66.80	74.20	80.51

Figure 18 and Figure 19 plot the average of relative posterior *SSE*s for each of the sensor location models in networks 1 and 2. The patterns in Figure 18 and Figure 19 are similar to the results for the independent measurement cases. The baseline scenario by model "Rand" indicates that the marginal effect on SSE reduction due to each additional sensor is nearly constant. DSEQ gives best performance in SSE reductions in both test networks; in particular note approximately 80% of the SSE reduction occurs with only 20% of the links providing measurements. The curves for models LFC and RFC are between the baseline and our proposed procedure DSEQ.

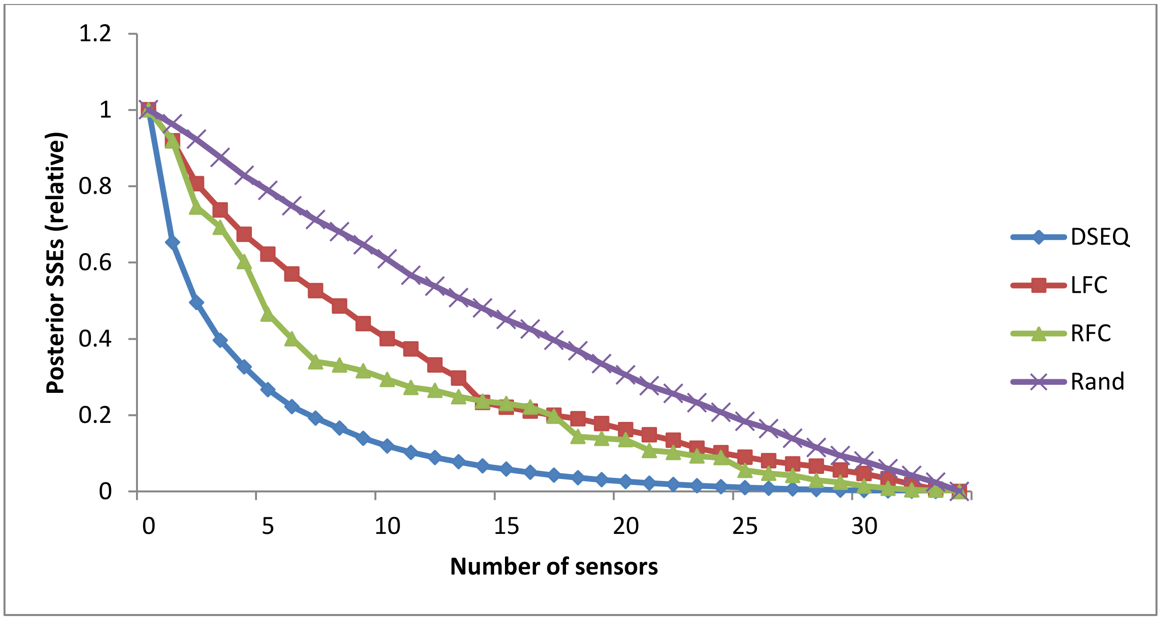

Figure 18 Relative Posterior SSE Relative SSE from Bayesian Estimation when Link Measurements are Dependent (Network 1)

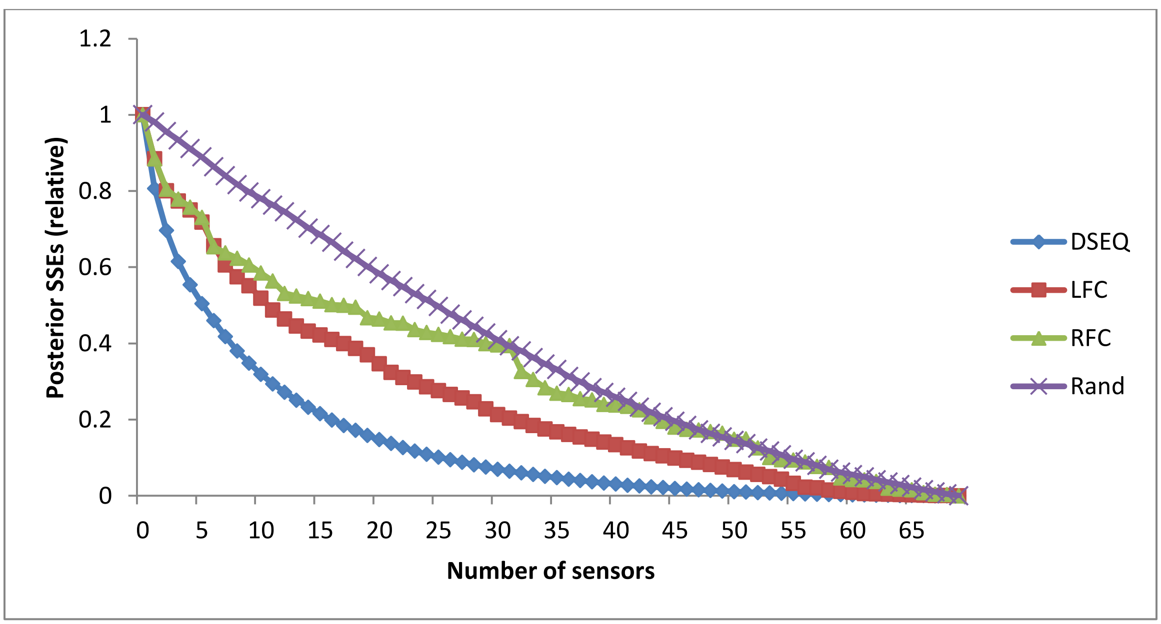

Figure 19 Relative Posterior SSE Relative SSE from Bayesian Estimation when Link Measurements are Dependent (Network 2)

The last experiment is to examine the sensor location models' performance in Houston network when observations are taken dependently. We focus the comparison on the cases when a small number of sensors are available. The pattern in Figure 20 validates the earlier conclusions for small networks and illustrates the significant outperformance of the proposed method DSEQ.

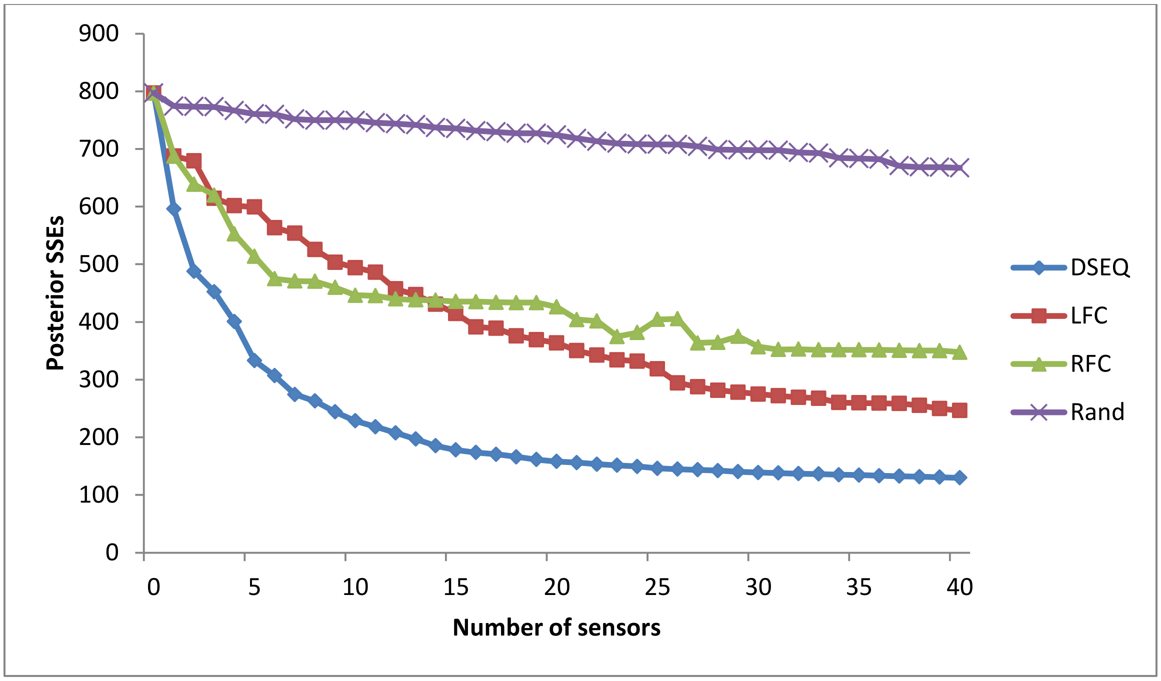

Figure 20 Posterior SSEs from Bayesian Estimation when Link Measurements are Dependent and Observe a Few Links (Houston Network)

5.6 Chapter Conclusions

In this chapter, a hierarchical linear Bayesian model was proposed for the stochastic route flow estimation. In the hierarchical model, the stochastic route flows $\mathbf{x}$ in the network was modeled as a multivariate normal distribution with two parameters: mean ($\boldsymbol{\gamma}$) and variance-covariance matrix ($\boldsymbol{\Phi}$). The prior knowledge for route flow mean was described by a multivariate normal distribution. The variance-covariance matrix for route flow was assumed to be known. The posterior distribution for the mean of route flows was updated from the link measurements on a subset of links using a Bayesian approach. We then proposed a sensor location model, the objective of which is to maximize the total variance reductions, with a constraint on the numbers of sensors. The objective function of sensor location model is formulated by minimizing the trace of variance reduction matrix, which forms a Bayesian A-optimal design model. Three solution methods were proposed and tested. First, the SEQ method, for the cases with independent link flow observations, sequentially selects one link at a time providing largest value of additional variance reduction, and updates the posterior variance for next iteration. The second method (GRE) calculates variance reduction by all single links, and selects first N links with largest values. GRE is proposed for the cases with independent link measurements in order to overcome the computational burden of SEQ. The last method (DSEQ) was designed for the cases when the measurements are dependent. Starting from an empty link set, DSEQ sequentially adds on link, at a time which maximizes the variance reduction with respect to posterior distribution due to all the selected links. All the proposed methods show significant advantage in terms of SSE of posterior distribution.

Chapter 6

LOCATING SENSORS TO ESTIMATE STOCHASTIC ROUTE FLOWS WITH UNKNOWN VARIANCE AND ASSUMING NORMAL DISTRIBUTION FOR PRIOR UNCERTAINTIES

6.1 Introduction

6.1.1 Problem Description

The traffic demand interest in this chapter is the stochastic traffic demand. First, a hierarchical linear model for stochastic route flows $\mathbf{x}$ in a network is developed. The counting sensors used in this chapter are "noisy sensors", the measurement errors from which are randomly distributed with a mean of zero and a known variance-covariance $\mathbf{\Sigma}$. Different from chapter 5, this chapter assumes that the mean $\boldsymbol{\gamma}$ and variance-covariance $\mathbf{\Phi}$ of route flows are both unknown and their prior knowledge are described by conjugate prior distributions. Hence, the third hierarchy of the Bayesian linear hierarchical model in chapter 5 is extended to include a prior distribution of the variance-covariance matrix $\mathbf{\Phi}$. The conjugate prior distribution for parameter of mean $\boldsymbol{\gamma}$ is multivariate normal distribution (*MVN*) with the mean $\boldsymbol{\mu}_\mathbf{0}$ and variance-covariance $\mathbf{V}_\mathbf{0}$. The mean $\boldsymbol{\mu}_\mathbf{0}$ represents the magnitude of the prior information and the variance $\mathbf{V}_\mathbf{0}$ is the measurement of the reliability for the prior information. The conjugate prior distribution for route flow variance-covariance $\mathbf{\Phi}$ is inverse-Wishart distribution (*IW*) with parameters $\mathbf{\Phi}_0$ and *d*, where $\mathbf{\Phi}_0$ is the scale parameter and *d* is the degree of freedom.

The first task of this chapter is to obtain the consistent Bayesian estimation of posterior route flow means $\boldsymbol{\gamma} \mid \mathbf{Y}$ for a given set of link flow observations $\mathbf{Y}$. The second task of this chapter is to determine the strategy to locate limited number of noisy sensors in the traffic network, which will be a good design in order to improve the estimation quality.

6.1.2 Notation

The notation is first defined in this chapter. Similar to previous chapters, the notation is classified as it relates to (a) network topology, (b) definition of route flows, (c) links flows, (d) observations or measurements, and (e) decision variables.

Network Topology Parameters:

R: Route-choice set for a network

A: Set of all the links in the network

A': Set of links where sensors are located

A_f: Set of links feasible for sensors to be located

$|R|$: Number of routes in a network

$|A|$: Number of links in a network

N: Number of sensors to locate

$\rho_a^i \in (0,1)$: Link-route parameter. $\rho_a^i = 1$ if route i uses link a, otherwise, $\rho_a^i = 0$

$$\mathbf{H} = \begin{bmatrix} \rho_1^1 & \cdots & \rho_1^{|R|} \\ \vdots & \ddots & \vdots \\ \rho_{|A|}^1 & \cdots & \rho_{|A|}^{|R|} \end{bmatrix}$$

Route Flows:

x^i: Real route flow of i^{th} route

$\mathbf{x} = (x^1, x^2, \dots, x^{|R|})'$

γ^i: The parameter to describe the mean of route flow x^i

$\boldsymbol{\gamma} = [\gamma^1, \gamma^2, \dots, \gamma^{|R|}]$

δ^i: The parameter to describe the variation of route flow x^i (random deviation of x^i from γ^i)

$\boldsymbol{\delta} = (\delta^1, \delta^2, \dots, \delta^{|R|})'$

ϕ_{ij}: Covariance of flows between routes *i* and *j*(variance of route flow *i* if *i* =*j*)

$$\boldsymbol{\Phi} = \begin{bmatrix} \phi_{11} & \cdots & \phi_{1|R|} \\ \vdots & \ddots & \vdots \\ \phi_{|R|1} & \cdots & \phi_{|R||R|} \end{bmatrix}$$

μ_0^i: Mean of prior distribution for parameter γ^i

$\boldsymbol{\mu_0} = (\mu_1^{(0)}, \mu_2^{(0)}, \dots, \mu_{|R|}^{(0)})'$

$v_{ij}^{(0)}$: Covariance of prior distribution between parameter γ^i and γ^j (variance of prior distribution if *i* =*j*)

$$\mathbf{V_0} = \begin{bmatrix} v_{11}^{(0)} & \cdots & v_{1|R|}^{(0)} \\ \vdots & \ddots & \vdots \\ v_{|R|1}^{(0)} & \cdots & v_{|R||R|}^{(0)} \end{bmatrix}$$

μ_1^i: Mean of posterior distribution for parameter γ^i

$\boldsymbol{\mu_1} = (\mu_1^{(1)}, \mu_2^{(1)}, \dots, \mu_{|R|}^{(1)})'$

$v_{ij}^{(1)}$: Covariance of posterior distribution between parameter γ^i and γ^j(variance of posterior distribution if *i* =*j*)

$$\mathbf{V_1} = \begin{bmatrix} v_{11}^{(1)} & \cdots & v_{1|R|}^{(1)} \\ \vdots & \ddots & \vdots \\ v_{|R|1}^{(1)} & \cdots & v_{|R||R|}^{(1)} \end{bmatrix}$$

$\boldsymbol{\Phi_0}$: First parameter of prior distribution of $\boldsymbol{\Phi}$

d: Second parameter of prior distribution of $\boldsymbol{\Phi}$

Link Flows:

v_a: Real link flow on link *a*

$\mathbf{v} = (v_1, v_2, \dots v_{|A|})'$

$\hat{v}_a$: Potential link flow observation if a sensor is locatedon link *a*

$\hat{\mathbf{v}} = (\hat{v}_1, \hat{v}_2, \dots \hat{v}_{|A|})'$

τ_a^2: Variance of link flow measurement on link *a*

$$\mathbf{\Sigma} = \begin{bmatrix} \tau_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tau_{|A|}^2 \end{bmatrix}$$

ε_a: Measurement error on link *a*

$\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{|A|}]$

Observations by sensors:

y_n: Link flow observed by sensor *n*

$\mathbf{Y} = (y_1, y_2, \dots, y_N)'$

g_n: An instance of link flow observed by sensor *n*

$\mathbf{g} = (g_1, g_2, \dots, g_N)'$

φ_n^2: Variance of observation by sensor *n*

Decision Variables for Location Models:

$y_a \in (0,1)$: $y_a = 1$ if locating sensor on link a, otherwise $y_a = 0$ (used for existing location models)

$z_{na} \in (0,1)$: $z_{na} = 1$if locating *n*th sensor on link *a*, otherwise $z_{na} = 0$.

$$\mathbf{z} = \begin{bmatrix} z_{11} & \cdots & z_{1|A|} \\ \vdots & \ddots & \vdots \\ z_{N1} & \cdots & z_{N|A|} \end{bmatrix}$$

$\mathbf{z}_n = [z_{n1}, \ldots, z_{n|A|}]$

6.2 Hierarchical Linear Model for Stochastic Route Flow Estimation

The stochastic route flow demand is modeled by a three-stage hierarchical Bayesian model. The first hierarchy of the model describes the relationship between link flow observations $\hat{\mathbf{v}}$ and route flows $\mathbf{x}$ is

$$\hat{\mathbf{v}} = \mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon}. \tag{6.1}$$

Eq. (6.1) indicates a linear relationship between $\hat{\mathbf{v}}$ and $\mathbf{x}$. Random measurement error $\boldsymbol{\varepsilon}$ is assumed generated from a multivariate normal distribution $MVN(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the variance-covariance matrix that is related to the reliability of the sensors and assumed to be known. Because this linear relationship only exists for the link set *A'*, where sensors are located, Eq. (6.1) will be modified by Eq. (6.2)

$$\mathbf{Y} = \mathbf{z}\hat{\mathbf{v}} = \mathbf{z}(\mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon}) = \mathbf{z}\mathbf{H}\mathbf{x} + \mathbf{z}\boldsymbol{\varepsilon}. \tag{6.2}$$

Hence, in the first hierarchy, the distribution of link flow observation $\mathbf{Y}$ given $\mathbf{x}$, which is the likelihood, is $MVN(\mathbf{zHx}, \mathbf{z}\boldsymbol{\Sigma}\mathbf{z}')$. Other parameters $\mathbf{z}$, $\mathbf{H}$ and $\boldsymbol{\Sigma}$ are constant.

The second hierarchy is to model the stochastic feature in route demand $\mathbf{x}$. By defining two parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ to describe the long term pattern and random fluctuations in route flow demand, $\mathbf{x}$ is expressed as

$$\mathbf{x} = \boldsymbol{\gamma} + \boldsymbol{\delta}. \tag{6.3}$$

Parameter $\boldsymbol{\delta}$ for random fluctuation is assumed to have *MVN* distribution with zero mean and variance-covariance $\boldsymbol{\Phi}$. In the second hierarchy, the conditional prior distribution given $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ for $\mathbf{x}$ is $MVN(\boldsymbol{\gamma}, \boldsymbol{\Phi})$.

In the final hierarchy, the prior knowledge for parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\Phi}$ are both described as probability distributions. Choose the conjugate prior distribution for $\boldsymbol{\gamma}$ of $MVN(\boldsymbol{\mu_0}, \mathbf{V_0})$, and conjugate prior distribution for $\boldsymbol{\Phi}$ of $IW(\boldsymbol{\Phi_0}, d)$, where $\boldsymbol{\mu_0}$, $\mathbf{V_0}$, $\boldsymbol{\Phi_0}$ and d are known. The posterior of route flow mean $\boldsymbol{\gamma}$ from using link flow observation $\mathbf{Y}$are interested to be inferred.

To sum up, the three-stage hierarchy to model the stochastic route flow is

$$\mathbf{Y} \,|\, \mathbf{x} \sim MVN(\mathbf{zHx}, \mathbf{z\Sigma z'}),$$

$$\mathbf{x} \,|\, \boldsymbol{\gamma}, \boldsymbol{\Phi} \sim MVN(\boldsymbol{\gamma}, \boldsymbol{\Phi}),$$

$$\boldsymbol{\gamma} \sim MVN(\boldsymbol{\mu_0}, \mathbf{V_0}), \boldsymbol{\Phi} \sim IW(\boldsymbol{\Phi_0}, d)$$

The posterior distribution for route flow mean $\boldsymbol{\gamma} \,|\, \mathbf{Y}$ will be calculated for a given sensor location $\mathbf{z}$.

6.3 Bayesian Computation for Posterior Route Flow Mean

The Bayesian method consists of combining the prior distribution and likelihood to derive the posterior distribution by Bayes' theorem. The posterior distribution could be written as the product of the likelihood and the prior distribution. In order to express the posterior information in a usable form, and to serve as formal inferences, it is important to calculate relevant summaries of the posterior distribution, such as the mean and variance. In the models of chapters 4 and 5, the prior distributions and likelihoods are of sufficiently convenient forms to obtain the necessary results by straightforward mathematics. However, in practice, the combination of likelihood and prior in the more complex models will generally produce a posterior distribution too complex for mathematical summarization, even if the two constituents separately are sufficiently simple. Our proposed hierarchical linear model in section 6.2 is such a complex one. For these complex models, we need general computational tools to calculate a variety of summaries from posterior distribution with mathematical complexity.

Popular computing tools in Bayesian practice are Markov Chain Monte Carlo (MCMC) methods. The advantage of MCMC method in Bayesian computation is their ability to enable inference from posterior distributions of high dimension by reducing the problem to one of recursively treating a sequence of lower-dimensional problems. Like traditional Monte Carlo methods, MCMC methods work by producing a sample of values $\{\boldsymbol{\theta}^{(g)}, g = 1,...,G\}$ from this distribution. A histogram or kernel density estimate based on such a sample is typically sufficient for reliable inference. The accuracy of the estimate can be increased by the Monte Carlo sample size *G*. However, unlike traditional Monte Carlo methods, MCMC algorithms produce correlated samples for the posterior, since they arise from recursive draws from a particular Markov chain, which is, starting from an arbitrary $\boldsymbol{\theta}^{(0)}$, and $\boldsymbol{\theta}^{(i+1)}$ is independent of $\boldsymbol{\theta}^{(i-1)}$, $\boldsymbol{\theta}^{(i-2)}$, … given the immediately preceding value, $\boldsymbol{\theta}^{(i)}$. Because of the Markov property of $\boldsymbol{\theta}^{(i)}$, the sequence has only a one-step memory. Based on results in Markov chain theory, subject to appropriate conditions, the distribution of $\boldsymbol{\theta}^{(i)}$ converges to the invariant (or stationary) distribution of that chain when *i* becomes large. Therefore, when the required conditions hold, all the $\boldsymbol{\theta}^{(i)}$ for sufficiently large *i* can be regarded as sampled from the invariant (stationary) distribution, which is from the true posterior distribution. The estimates for posterior distribution can be made from the samples $\boldsymbol{\theta}^{(i)}$ drawn from the stationary distribution when *i* become large.

One MCMC algorithm is known as Gibbs sampling. Assume our model features k parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$. Gibbs sampler involves successive sampling from the complete conditional densities

$$p(\theta_p \mid \theta_1, ..., \theta_{p-1}, \theta_{p+1}, ..., \theta_k, \mathbf{y})$$

which conditions on both the data $\mathbf{y}$ and the other parameter θ_i when $i \neq p$. Given an arbitrary set of starting values $\{\theta_2^{(0)}, ..., \theta_k^{(0)}\}$, the Gibbs sampling algorithm proceeds as follows:

For $(t \in 1:T)$, repeat:

Step 1: Draw $\theta_1^{(t)}$ from

$$p(\theta_1 \mid \theta_2^{(t-1)}, \theta_3^{(t-1)}, ..., \theta_k^{(t-1)}, \mathbf{y})$$

Step 2: Draw $\theta_2^{(t)}$from

$$p(\theta_2 \mid \theta_1^{(t)}, \theta_3^{(t-1)}, ..., \theta_k^{(t-1)}, \mathbf{y})$$

…

Step k: Draw $\theta_k^{(t)}$from

$$p(\theta_k \mid \theta_1^{(t)}, \theta_2^{(t)}, ..., \theta_{k-1}^{(t)}, \mathbf{y})$$

These distributions are always known up to proportionality constant because they take the form of the [likelihood] * [prior with everything fixed but θ_p] in step *p*. Such successive samples may involve simple sampling from standard densities (Normal, gamma, etc.), or sampling from non-standard densities. The latter case involves the rejection sampling approach. If the full conditionals are non-standard but a certain mathematical form (log-concave), the adaptive rejection sampling (Gilks and Wild, 1992) may be used within the Gibbs sampling for those parameters. Otherwise, alternative schemes based on the Metropolis-Hastings algorithm (Metropolos et al. 1953) may be used to sample from non-standard densities. The collection of complete conditional distributions uniquely determines the joint posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y})$, as well as all marginal posterior distributions $p(\theta_i \mid \mathbf{y}), i = 1, \ldots, k$.

Under mild regularity conditions (Roberts and Smith 1993), the *k*-tuple, $(\theta_1^{(t)}, \ldots, \theta_k^{(t)})$ obtained at iteration *t* converges in distribution to a draw from the true joint posterior distribution $p(\theta_1, \ldots, \theta_k \mid \mathbf{y})$. For t sufficiently large (greater than t_0), $\{\boldsymbol{\theta}^{(t)}, t = t_0 + 1, \ldots, T\}$ is a correlated sample from the true posterior. The time from $t = 0$ to $t = t_0$ is known as the burn-in period. Hence, the statistics for the posterior distribution can be estimated, such as a histogram of the $\{\theta_i^{(t)}, t = t_0 + 1, \ldots, T\}$, and a sample mean to estimate the posterior mean as:

$$\hat{E}(\theta_i \mid \mathbf{y}) = \frac{1}{T - t_0} \sum\nolimits_{t=t_0+1}^{T} \theta_i^{(t)} \qquad (6.4)$$

In practice, statisticians run *m* parallel Gibbs sampling chains in order to assess sampler convergence. The samples drawn from the burn-in period should be discarded and the remaining samples are used to obtain the posterior mean estimates as:

$$\hat{E}(\theta_i \mid \mathbf{y}) = \frac{1}{m(T - t_0)} \sum_{j=1}^{m} \sum_{t=t_0+1}^{T} \theta_{i,j}^{(t)} \qquad (6.5)$$

where *j* indicates the chain number.

Suppose we have a single chain of *N* post-burn-in samples of a parameter θ, a variance estimate from independent samples is given by

$$\text{var}_{iid} = \frac{s_\theta^2}{N} = \frac{1}{N(N-1)}\sum_{t=1}^{N}(\theta^{(t)} - \theta_N)^2 \qquad (6.6)$$

where $\theta_N = E(\theta \mid \mathbf{y}) = \frac{1}{N}\sum_{t=1}^{N}\theta^{(t)}$ is the posterior mean estimator, and

$s_{\boldsymbol{\theta}}^2 = \frac{1}{N-1}\sum_{t=1}^{N}(\theta^{(t)} - \theta_N)^2$ is the sample variance. This estimate is an underestimate because the MCMC samples have positive autocorrelation. Define the autocorrelation time for θ as

$$\kappa(\theta) = 1 + 2\sum_{k=1}^{\infty}\rho_k(\theta) \qquad (6.7)$$

where $\rho_k(\theta)$ is the autocorrelation at lag k for the parameter θ. Hence, the effective sample size (ESS) is defined as

$$ESS = N / \kappa(\theta) \qquad (6.8)$$

The variance estimate is then

$$\text{var}_{ESS} = \frac{s_\theta^2}{ESS} = \frac{\kappa(\theta)}{N(N-1)}\sum_{t=1}^{N}(\theta^{(t)} - \theta_N)^2 \qquad (6.9)$$

For the correlated samples, we have $\kappa(\theta) > 1$ and $ESS < N$, so that $\text{var}_{ESS} > \text{var}_{iid}$. In practice, $\kappa(\theta)$ is estimated using sample autocorrelations estimated from the MCMC chain.

Because the MCMC method is based on the assumption that after the burn-in-period, the Markov chain goes to a stationary mode and all the samples are drawn from true posterior. The challenge for implementing MCMC is convergence diagnosis, which is to decide when it is safe to stop the simulation procedure and summarize the output. The most common approach is to run a few parallel sampling chains with different starting points that are initially over-dispersed with respect to the true posterior. After running the *m* chains for 2*N* iterations each, we try to see whether the variation within the chains equals the total variation across all chains during the latter half iterations. The convergence is monitored by the statistic by Gelman and Rubin (1992):

$$\sqrt{\hat{R}} = \sqrt{\left(\frac{N-1}{N} + \frac{m+1}{mN}\frac{B}{W}\right)\frac{df}{df-2}} \tag{6.10}$$

where *B/N* is the variance across the *m* parallel chains, *W* is the average of the *m* within-chain variances, and *df* is the degrees of freedom of an approximating *t* density to the posterior distribution.

The hierarchical linear model for Bayesian route flow estimation model proposed in section 6.2 will be solved by MCMC algorithms which have been coded in the general purpose MCMC software WinBUGS (Spiegelhalter et al., 2003). WinBUGS is the Windows successor to Bayesian inference Using Gibbs Sampling. In WinBUGS, the parameters are updated using conditional distributions, or using Metropolis-Hastings algorithm if the prior distribution and the likelihood are not a conjugate pair. WinBUGS has an interactive environment that allows the user to specify models hierarchically and display plots for convergence diagnostics, model checks and comparisons.

In the proposed hierarchical linear models for stochastic route flows, the full conditional distributions are all in closed form. For example, with a normal prior at the first stage for route flow **x**, the conditional posteriors for the parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\Phi}$ arising from normal and inverse Wishart priors in the third-stage are also normal and inverse Wishart respectively. The remaining conditional distribution, for **x** given $\boldsymbol{\gamma}$ and $\boldsymbol{\Phi}$, is the standard posterior arising from a single-stage linear model with normal prior (the proposed model in Chapter 4). Therefore, Gibbs sampler can be implemented conveniently because the samplers can be drawn from each of the full conditional distributions with closed form.

Based on the proposed hierarchical linear model for stochastic route flows, with a given sensors location **z**, the mean of posterior route flow mean $E(\boldsymbol{\gamma} \mid \mathbf{zY})$ can be estimated by Eq. (6.5) using the samples drawn from MCMC algorithms coded in WinBUGS. The variance of posterior route flow mean $\text{var}_{ESS}(\boldsymbol{\gamma} \mid \mathbf{zY})$ is estimated according to Eq. (6.9) using the MCMC samples. This methodology for route flow mean estimate is different from the previous chapter. With the known variance $\boldsymbol{\Phi}$ in chapter 5, the variance of posterior route flow mean can be calculated by:

$$\text{var}(\boldsymbol{\gamma} \mid \mathbf{z}, \mathbf{Y}) = \mathbf{V_1} = (\mathbf{V_0^{-1}} + \mathbf{H'z'(zH\Phi H'z' + z\Sigma z')^{-1} zH})^{-1} \quad (6.11)$$

which is independent of the value of observation **Y**. However, the posterior variance estimates by MCMC $\text{var}_{ESS}(\boldsymbol{\gamma} \mid \mathbf{zY})$ depends on the observation **Y**'s value. In other words, the posterior variance in chapter 5 can be accurately predicted *before* actually making the observation; but in the model of chapter 6 we do not have ability to precisely predict posterior variance *before* obtaining the observations.

6.4 Sensor Location Modeling for Stochastic Route Flow Estimation

Although the convergence of MCMC algorithm can be diagnosed with the assistant of Gelman and Rubin's method, the assumption behind is that the process is stationary. Non-stationary process will not converge even when the burn-in-period N is set to be large. Therefore, the modeler has to ensure that the process is stationary in order to applying MCMC algorithm. *For the Sensor Location Flow-Estimation problem, the*

Markov chain process is stationary only when all routes are covered (observed) by sensors, which means the route coverage is a hard constraint in sensor location model in order to apply MCMC to estimate posterior route flows. Route coverage (or OD coverage) constraint specifies that the subset of links where sensors are located should ensure that each route (or OD pair) is observed by at least one located sensor. OD coverage is proposed by Yang et al. (1991) as a rule to determine sensors' location for OD estimation problem. It has been shown that if there are any OD pairs not observed by any sensors, then the corresponding "relative deviation between the estimated trips and the true ones" can tend to infinite. The OD coverage constraint ensures a finite error in the OD estimation.

The objective of optimal sensor location model is to minimize the expected posterior variances (or maximize the expected variance reduction), in particular minimize the trace of posterior variance-covariance matrix. This strategy will provide a sensor location model with Bayesian *A*-optimality. The conditional posterior route flow mean $\boldsymbol{\gamma}$ given route flow variance-covariance $\boldsymbol{\Phi}$ is shown as:

$$\mathrm{var}(\boldsymbol{\gamma} \mid \boldsymbol{\Phi}, \mathbf{z}, \mathbf{Y}) = \mathbf{V_1} = (\mathbf{V_0^{-1}} + \mathbf{H'z'}(\mathbf{zH\Phi H'z'} + \mathbf{z\Sigma z'})^{-1}\mathbf{zH})^{-1} \qquad (6.12)$$

Because the variance-covariance matrix of posterior mean $\mathrm{var}(\boldsymbol{\gamma} \mid \mathbf{z}, \mathbf{Y})$ does not have a close form expression when integrating the variance-covariance $\boldsymbol{\Phi}$ out of Eq. (6.12), $\mathrm{var}(\boldsymbol{\gamma} \mid \mathbf{z}, \mathbf{Y})$ has to be estimated from MCMC samples. In the sensor location stage before obtaining observation $\mathbf{Y}$, we can approximate the variance reduction by substituting the scale parameter $\boldsymbol{\Phi}_0$ of conjugate prior distribution of $\boldsymbol{\Phi}$ in the conditional posterior variance in Eq. (6.12). The sensor location model is formulated as:

$$\max tr(\mathbf{V_0 H'z'(z\Sigma z' + zH\Phi_0 H'z' + zHV_0 H'z')^{-1} zHV_0}) \quad (6.13)$$

$$\text{s.t. } \sum_{a=1}^{|A|} z_{na} = 1, n = 1,..., N \quad (6.14)$$

$$\sum_{n=1}^{N} z_{na} \leq 1, \forall a \in A \quad (6.15)$$

$$\sum_{a=1}^{|A|} \left(\rho_a^i \left(\sum_{n=1}^{N} z_{na} \right) \right) \geq 1 \quad (6.16)$$

$$z_{na} \in (0,1), n = 1,..., N, a \in A \quad (6.17)$$

The objective function Eq. (6.13) maximizes the approximated variance reduction in the posterior distribution. Constraint Eq. (6.14) ensures that every sensor *n* has to be located on one and only one link. Constraint Eq. (6.15) forces that each link allows at most one sensor to be located on. Constraint Eq. (6.16) is the route coverage constraint that forces every route is observed by at least one sensor. Constraint Eq. (6.17) is the binary constraint for decision variablez_{na}.

The linear approximation of objective function Eq. (6.13) can be formulated as follows: we first calculate the variance reduction $tr(\mathbf{V_0 H'z'(z\Sigma z' + zH\Phi_0 H'z' + zHV_0 H'z')^{-1} zHV_0})$ for each single sensor *a*, which is notated by vr_a; then approximate the variance reduction from multiple links by the summation of variance reduction by single link vr_a from all corresponding links. Therefore, the approximated linear model is formulated as:

[LRVR] $$\max \sum_{a=1}^{|A|} vr_a \quad (6.18)$$

s.t. (6-14) – (6-17)

Then the sensor location model can be solved by the solution engines of linear integer optimization models.

6.5 Experimental Result

6.5.1 Experiment Setup

We applied the proposed sensor location model LRVR and modified models of existing methods Link Flow Coverage (LFC) and Route Flow Coverage (RFC), which include additional set of route coverage constraints in the formulations. Because route coverage is a mandatory in all the sensor location models, the objective function of RFC becomes redundant because all routes have been observed. Hence, RFC model is same as locating sensors at random links with the route coverage constraint. The route flow mean is updated by MCMC procedures using the link observations from three different sensor location models. All the three sensor location models are solved by the linear mixed integer program solver in Matlab. Then the route flow means are estimated by MCMC algorithms which are coded in WinBUGS.

Figure 21 shows the scheme of the experimental procedure. First (Steps 1-2), problem scenarios are generated by defining the network (supply), loading the assumed traffic flows (demand) and obtaining the "actual" mean of network traffic on routes $\boldsymbol{\gamma}$. The "actual" variance-covariance matrix $\boldsymbol{\Phi}$ can be obtained from other models which are independent with this experiment, for example, a Bayesian model learning the variance-covariance on routes from link observations. By simulating the random perception errors on route flows and measurement errors, route flow prior distributions for the mean (multivariate normal distribution $\boldsymbol{\gamma} \sim MVN(\boldsymbol{\mu}_0, \mathbf{V}_0)$) and the variance-covariance (inverse-

Wishart distribution $\mathbf{\Phi} \sim IW(\mathbf{\Phi_0}, d)$), sensors' reliabilities $\mathbf{\Sigma}$ and potential link flow observations $\mathbf{Y}$ are generated by linking the errors to the actual values (Step 3). The reliability of the prior distribution of parameter $\boldsymbol{\gamma}$ for was first randomly generated by $\mathbf{u} \sim U(0, \mathbf{u}^{\max})$; then the diagonal elements of prior's variance-covariance matrix $\mathbf{V_0}$ is calculated by $\eta_{ii}^{(0)} = \gamma^i u^i$; all covariance are assumed to be zero in the initial priors; the mean of the prior distribution of parameter $\boldsymbol{\gamma}$ is generated from a normal random generator $\boldsymbol{\mu_0} \sim MVN(\boldsymbol{\gamma}, \mathbf{V_0})$. The scale parameter $\mathbf{\Phi_0}$ for the prior distribution of $\mathbf{\Phi}$ is simulated by adding a small deviation to each cell of $\mathbf{\Phi}$ (note $\mathbf{\Phi_0}$ is symmetric matrix). The degree of freedom d for the prior distribution of $\mathbf{\Phi}$ is set to equal to the number of routes. The random deviation from $\boldsymbol{\gamma}$ is generated by $\boldsymbol{\delta} \sim N(\mathbf{0}, \mathbf{\Phi})$ and such deviation is added to the "actual" route flow mean values. Hence the "actual" value of link flows can be calculated using the link-route parameter by $\mathbf{v} = \mathbf{H}(\boldsymbol{\gamma} + \boldsymbol{\delta})$. The sensors' reliability $\mathbf{t}$ is randomly generated from $\mathbf{t} \sim U(\mathbf{0}, \mathbf{t}^{\max})$; and the diagonal elements in variance-covariance matrix of link flow measurement $\mathbf{\Sigma}$ can be calculated by $\tau_a^2 = v^a t^a$; the potential observation on such link is generated from a normal random generator $\mathbf{Y} \sim N(\mathbf{v}, \mathbf{\Sigma})$. In Step 4, the location model is solved to obtain location decision **z**. Particularly for the proposed model, we first calculate the variance reduction by single sensor on link *a*; then the location variable **z** is solved from the mixed integer linear model LRVR by the solution engine in Matlab. The location solutions for LFC and RFC are obtained by the solver in Matlab as well. Then, with these sensor location decision **z**, posterior distribution for route flow means $\boldsymbol{\gamma} | \mathbf{z}, \mathbf{Y}$ are obtained using the MCMC algorithm in WinBUGS (Step 5). In MCMC implementation, in order to ensure the model convergence, we setup the burn-in-period as 10000, and the extra 5000 samples is drawn to calculate the posterior quantities. To evaluate each location model (Step 6), the sum-squared error (SSE) of the posterior distribution are compared as:

$$SSE = BIAS^2 + Variance = \sum_{i \in R} \left(\mu_i^{(1)} - \gamma^i \right)^2 + \sum_{i \in R} \eta_{ii}^{(1)} \qquad (6.19)$$

where γ^i is the true mean value of the route flow *i*.

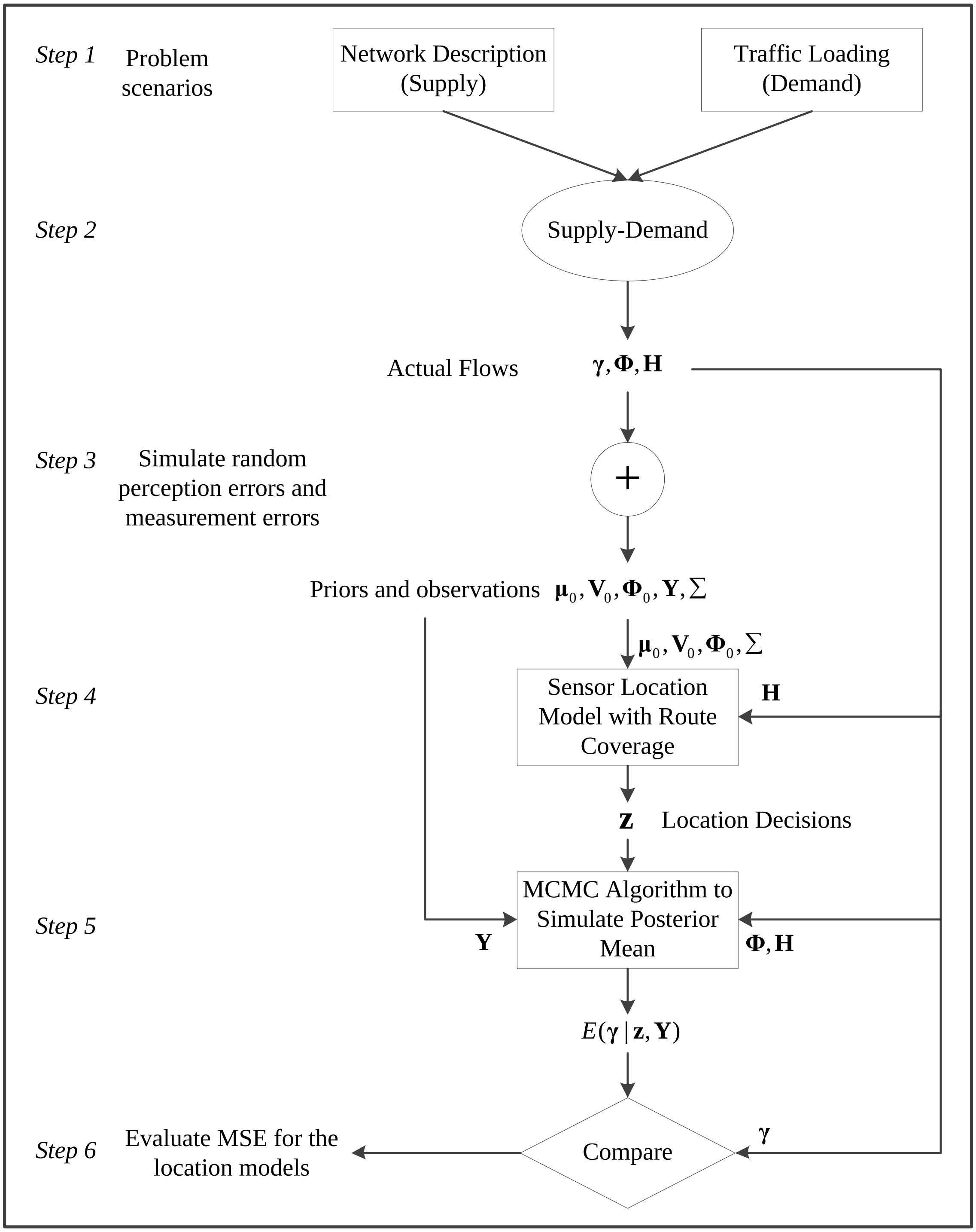

Figure 21 Procedure for the computational experiments for Chapter 6

Finally, note that for each network scenario, we can have different problem instances by locating different number of sensors N ($N \geq t$). t is the minimum number of sensors to cover all routes in a network. The value of t is solved by a linear integer model formulated as:

$$\min t \quad (6.20)$$

$$s.t. \sum_{a \in A} \rho_a^i y_a \geq 1, \forall i \in R \quad (6.21)$$

$$\sum_{a \in A} y_a \leq t \quad (6.22)$$

$$y_a \in (0,1), \forall a \in A \quad (6.23)$$

Solving each problem instance by each of the location models gives us a score as defined by the *SSE*. Because of the stochastic nature of the experiment framework, for each network supply and traffic demand, we can run the experiments several times--each with priors and observations randomly generated. The average *SSE*s is calculated in the end to evaluate the performance of three models. A grid network with 16 OD pairs, 36 routes and 34 links (grid network 1) is used to conduct all the experiments.

When generating the experiments, u^{max} is set to be 0.15 and t^{max} is 0.1, which means the maximum measurement error is within 10% deviation from the actual link flow, while the maximum perception error in prior is within 15% off from the actual mean of route flow. For each test network, 20 scenarios with different priors and observations are randomly generated within the specific range, and the priors and observations are used to first generate the locational solution **z**, and then the posterior mean of route flow parameter $\boldsymbol{\gamma}$.

6.5.2 Numerical Results

For the experiment network, 20 random generated scenarios (priors and observations) are used to conduct the experiment to compare the performance of three sensor location models. The SSE results are plotted in Figure 22. The minimum number of sensors required to observe all routes in experiment network is eight. From Figure 22, our proposed model LRVR outperforms the other two comparison models LFC and RFC when observing less than half of the networks. RFC is the worst performer in posterior SSE because it does not optimize the sensors location (any choice of sensors' location is equivalent for RFC with route coverage constraint). The performance of LFC and RFC are comparable when the number of sensors in use is large (when more than half of the links are observed). This is because the linear approximation of variance reduction for multiple links ignores the covariance between link measurements. The effect of such correlations will be essential when more links are in use. Therefore, the proposed LRVR is more suitable to the cases when the budget is only for few sensors in comparison to the total number of links.

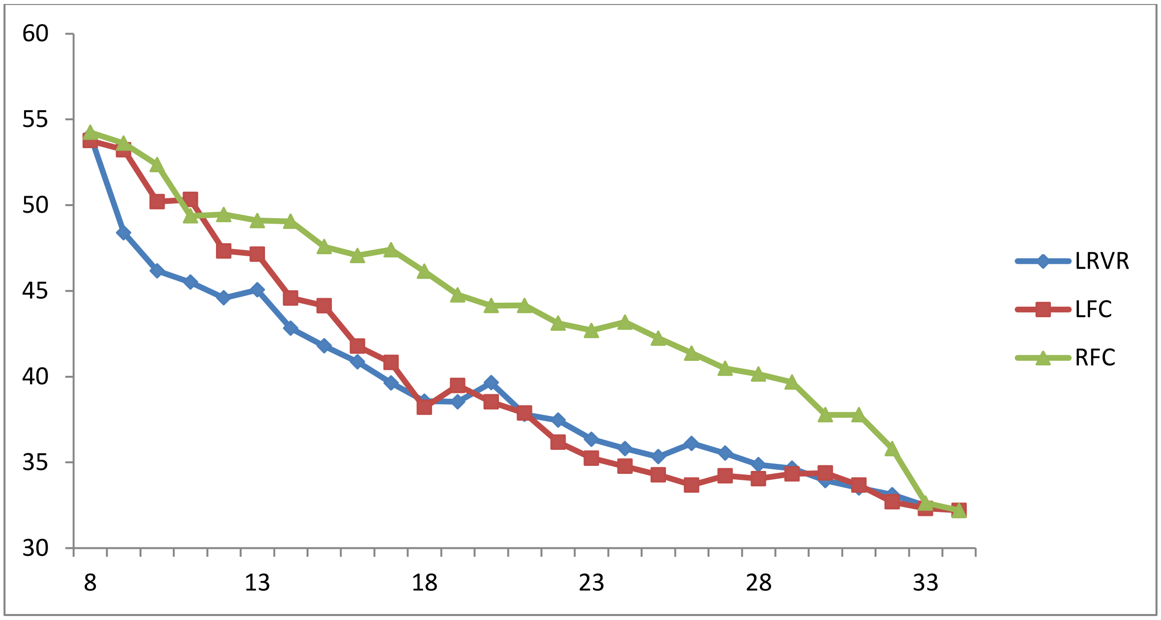

Figure 22 Posterior SSEs Estimated using MCMC Method with Good Prior for Φ

In the objective function Eq. (6.13), we use the value of the scale parameter $\mathbf{\Phi_0}$ for route flow variance-covariance $\mathbf{\Phi}$ to roughly estimate the total variance reduction. Because the prior distribution for the route flow variance-covariance $\mathbf{\Phi}$ is inverse-Wishart distribution, which requires another parameter *d* for the degree of freedom, the formulation of LRVR does not capture the information of the confidence about the scale parameter $\mathbf{\Phi_0}$. Figure 23 plots the SSE results for the average of 10 randomly generated scenarios (priors and observations) when setting a small degree of freedom *d* for the inverse-Wishart prior distribution (equals to the number of routes in the network), which represents the large uncertainty for the prior knowledge of route flow variance-covariance $\mathbf{\Phi}$. From Figure 23, the performance of LRVR is between that of LFC and RFC. This result indicates that the proposed approximation of variance reduction is less effective when a very vague prior knowledge for route flow variance-covariance is used.

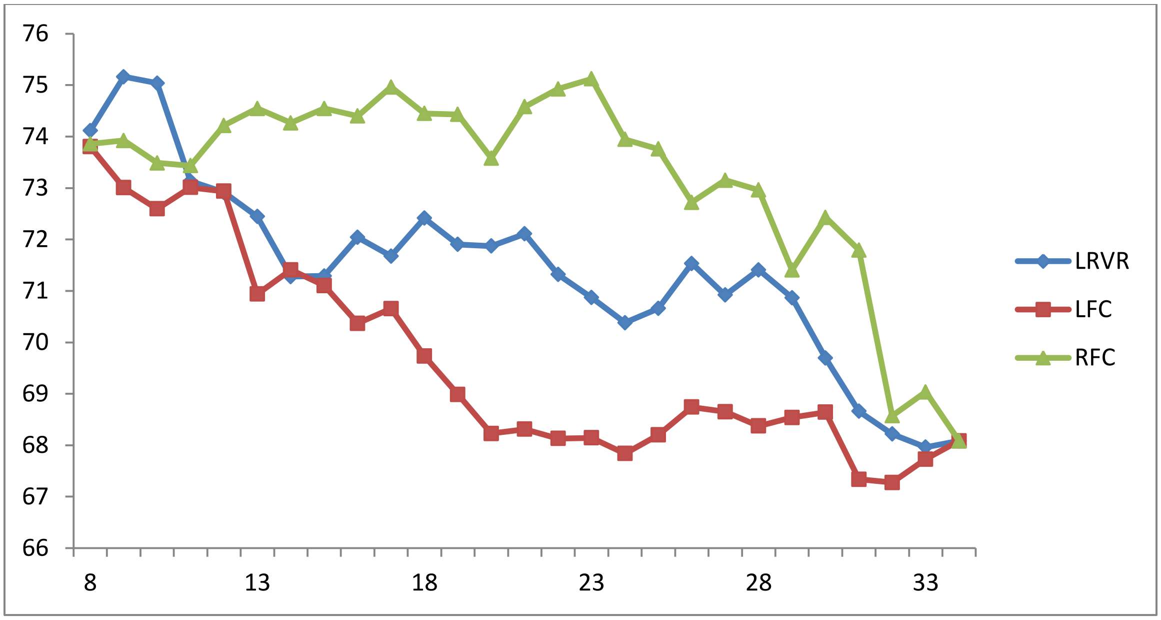

Figure 23 Posterior SSEs Estimated using MCMC Method with Bad Prior for Φ

6.6 Chapter Conclusions

In this chapter, a hierarchical linear Bayesian model was proposed for the stochastic route flow estimation. In the hierarchical model, the stochastic nature of the route flows in the network is modeled as a multivariate normal distribution with two parameters: mean and variance-covariance matrix. The variance-covariance matrix was assumed to be unknown and its prior knowledge was described as an inverse-Wishart distribution. Because the posterior distribution for the mean of route flows has no closed form in proposed model, MCMC was used to draw samples and estimate the posterior quantities, such as mean and standard deviation. In order to build a stationary Markov chain, all routes must be observed by at least one of the links so that the route coverage constraints are satisfied in the sensor location model. A sensor location model was proposed, the objective of which is to maximize the total variance reduction, with the route coverage constraint and the budget constraint. By replacing the value of scale parameter $\mathbf{\Phi_0}$ into the conditional posterior variance $\text{var}(\boldsymbol{\gamma} \mid \mathbf{\Phi}, \mathbf{Y})$, the variance reduction is approximated in order to solve the no closed form issue of marginal posterior variance $\text{var}(\boldsymbol{\gamma} \mid \mathbf{Y})$. Then the total variance reduction by multiple links is approximated by the summation of variance reduction by all the corresponding links individually. Therefore, an integer linear model LRVR was formulated to select the subset of links by optimizing the linear approximation of total variance reduction of route flow mean $\boldsymbol{\gamma}$.

The proposed model LRVR was tested with the experiment framework on a grid network. The experiment framework includes the procedures to generate the stochastic network scenarios, decide the optimal subset of links with sensors by solving the sensor location models (LRVR, LFC and RFC), and estimate the mean of posterior distribution for route flow mean $\boldsymbol{\gamma}$. The experiment results show the advantage of proposed integer linear model LRVR compared with traditional sensor location models LFC and RFC, especially when observing only a small number of links in the network. The limitation of LRVR model is as follows:

(1) Because the objective function of LRVR model is a linear approximation by assuming the total variance reduction by multiple links is the summation of that by each of the single links, the approximation does not count the effect of covariance between measurements on pairs of links; this approximation may be too much when the correlation among link measurements cannot be neglected.

(2) The degree of freedom in the prior distribution of the route flow variance-covariance matrix has not been used in the sensor location model. Hence, when the prior knowledge for route flow variance-covariance matrix is too vague (when d is small), the estimation for the marginal posterior variance $\text{var}(\boldsymbol{\gamma} \mid \mathbf{Y})$ is not

very accurate when substituting the value of the scale parameter $\mathbf{\Phi_0}$ in the conditional posterior variance $\mathrm{var}(\boldsymbol{\gamma} \mid \mathbf{\Phi}, \mathbf{Y})$. In this case, the traditional sensor location model, for example LFC, works better than LRVR.

Chapter 7

SUMMARY AND FUTURE RESEARCH

7.1 Summary of Research Results

Improving the quality of Origin-Destination (OD) demand estimates increases the effectiveness of design, evaluation and implementation of traffic planning and management systems. The quality of estimated OD demands from link counts depends on several factors, such as (1) the route-choice and traffic loading assumptions, (2) the quality of observed data from sensors, (3) the dependencies between link flows due to network topology and traffic loading, (4) the choice of OD estimation methods, and (5) where the sensors are located. The last two factors reveal two important research questions:

- How to compute the best estimates of the flows of interest by using anticipated data from given candidate sensors location;
- How to decide on the optimum subset of links where counting sensors are to be located.

The aim of this dissertation is to develop a decision framework to obtain the high-quality OD volume estimates in traffic networks. Three major contributions were made in this dissertation:

(1) It developed the decision framework for the recently defined Sensor Location Flow-Estimation problem. The developed framework is an integration of several well-defined problems in traffic modeling, such as (a) a traffic assignment model to load the OD traffic volumes on routes in a known route set, (b) a sensor location model to decide on which subset of links to locate counting sensors to

observe traffic volumes, and (c) an estimation model to obtain best estimates of OD or route flow volumes.

(2) The proposed decision framework is compatible with estimates of both deterministic and stochastic demands (route flows) in traffic networks. Four new location models and several algorithms were developed to locate both noise-free and noisy sensors in such deterministic and stochastic scenarios.

(3) This dissertation developed an experimental environment which can handle the evaluation and comparison of different sensor location models in terms of OD estimation qualities.

Chapter 3 first introduced a model to obtain good estimates of deterministic route flows in a traffic network using observation data from links, given apriori knowledge of route flows and their uncertainties; the uncertainties in the case were in terms of confidence intervals. A linear integer programming model was developed for location decisions of limited number of "perfect" (or noise-free) sensors. A greedy algorithm was proposed to optimally and efficiently solve the linear integer model.

In chapter 4, noisy sensors were considered which included error terms that were normally distributed. Also the uncertainties in prior knowledge of flows were modeled with a multivariate normal distribution. A Bayesian statistical procedure was used to produce the posterior route flow estimates by combining the likelihood (that is the anticipated distribution of observed values given the distribution of route flows) and the prior distribution of route flows. A sensor location model was proposed to obtain optimal location of these "noisy" sensors. A sequential procedure was developed to solve the sensor location model in polynomial time which was shown to be optimal when error measurements were statistically independent.

In chapter 5, the problem of estimating stochastic route flows was addressed. The stochastic real route flows are assumed to be generated from a Multivariate Normal distribution with two parameters: “mean” and “variance-covariance matrix”, and the “variance-covariance matrix” parameter was assumed to be known. A three-stage hierarchical linear Bayesian model was developed: the first hierarchy describes the relationship between link flow observations and route flows to be estimated; the second hierarchy defines the distribution of stochastic route flows; and the final hierarchy introduces the prior distribution of the “mean” parameter of stochastic route flows. A Bayesian *A*-optimal design was developed to choose the sensors’ location in order to maximize the pre-posterior expected reduction of uncertainties. Three solution methods, SEQ, GRE and DSEQ, were proposed for both independent and dependent measurements. First, the SEQ method, for the cases with independent link flow observations, sequentially selects one link at a time providing largest value of additional variance reduction, and updates the posterior variance for next iteration. The second method (GRE) calculates variance reduction by all single links, and selects first *N* links with largest values. GRE is proposed for the cases with independent link measurements in order to overcome the computational burden of SEQ. The last method (DSEQ) was designed for the cases when the measurements are dependent. Starting from an empty link set, DSEQ sequentially adds one link at a time, which maximizes the variance reduction with respect to posterior distribution due to all the selected links.

In chapter 6, the “variance-covariance matrix” parameter of the route flows was assumed to be unknown and its prior knowledge was described by an inverse-Wishart distribution. The three-stage hierarchical linear Bayesian model in chapter 5 was extended by introducing the prior distribution for the “variance-covariance matrix” parameter of the route flows in the final hierarchy. Because the posterior distribution for the “mean” parameter of route flows has no closed form in the Bayesian model, Markov Chain Monte Carlo approach was used to estimate the aposteriori quantities. For stability of the solution approach, it was required that all routes have to be observed by at least one sensor in order to obtain a stationary Markov chain. The sensor location model was developed to maximize an approximate variance reduction function.

In all cases the objectives of the sensor location models were the maximization of uncertainties reduction or the maximization in the reduction of variances of the posterior distribution of the estimates of the route flow volumes. Developed models were compared with other available models in the literature. The comparison showed that the models developed in this research performed better than available models in the literature.

7.2 Directions of Future Research

This dissertation addressed the recently defined *Sensor Location Flow-Estimation* problem and developed sensor location models for four specific scenarios. A number of future research problems can be solved to generalize the developed framework and enhance the ability to obtain the high-quality OD volume estimates in traffic networks.

The limitations that we have identified with the model in Chapter 6 (estimate stochastic demand with unknown variance and noisy sensors) are (i) the neglect of the correlations among link measurements and (ii) the effect of poor quality of the prior knowledge of “variance-covariance matrix” parameter for the stochastic route flow, on the for sensor location decisions. One direction for future research would be to derive a more accurate prediction of posterior variances before taking observations. Such prediction of posterior variances can be used as the criterion to choose the sensor locations. The modified model would incorporate the correlations among link measurements (such as the parameter to describe the variance-covariance matrix for link measurement errors) and the uncertainties of “variance-covariance matrix” parameter (such as the term to define the degree of freedom of the prior distribution).

Another possibility would be to develop OD estimation methods which would allow models outside of the conjugate families. The models in this dissertation assumed the traffic demand is generated from multivariate normal (*MVN*) distribution, which is a probability distribution in a conjugate family and easily produces mathematical tractable results of posterior quantities. Because the traffic demands and observations are concerned with counts (number of trips), the true distribution is probably some multivariate form of the Poisson distribution (Maher, 1983). However, *MVN* provides an accurate approximation for counts with means which are not too small (say, greater than 1000 vehicles). It would be very valuable to extend the framework and reformulate the sensor location problem under the Poisson assumption for OD demands. The aim would be to analyze the sensitivity of assumed distribution to the quality of estimated OD demands, and enhance the ability of proposed framework with different amount of volumes.

The third possible direction for future research would be to examine if equilibrium is reachable for the given route-choice set. If the given route-choice set could result traffic equilibrium in the network, instead of comparing route flow estimates with real route flows, the proposed framework would be extended with the comparison of OD estimates (summation of route flow estimates from all the alternative routes connecting each OD pair) to real OD demands, the result of which directly demonstrate the quality of demand estimates at the OD trips matrix level.

This dissertation focused on estimating OD demand (the lower level of *Sensor Location Flow-Estimation* problem) by using the Bayesian approach. Based on the literature review in Chapter 2, there are several other statistical methods, such as Maximum Likelihood and Generalized Least Squares, which could be used to provide OD estimates. Hence, a future research direction could be investigate with other OD estimation methods at the lower level of *Sensor Location Flow-Estimation* problem, other than Bayesian approach.. The sensitivity of OD estimation methods to the quality of demand estimates could be analyzed and the ability of proposed framework to obtain high quality demand estimates could be further enhanced.

Last but not least, the framework would be applied to some actual large scale networks (like Houston network tested in this dissertation) to compare with existing OD estimates and provide additional insights to traffic planners and the traffic modeling community.

REFERENCES

FHWA. (2011). Our Nation's Highways: 2011. Federal Highway Administration, US. Department of Transportation. Retrieved Jan. 2, 2012, from http://www.fhwa.dot.gov/policyinformation/pubs/hf/pl11028/.

(2011). 2011 Urban Mobility Report. Retrieved Jan. 2, 2012, from http://mobility.tamu.edu/ums/report/.

Abrahamsson, T. (1998). Estimation of Origin-Destination Matrices Using Traffic Counts-A Literature Survey. International Institute for Applied Systems Analysis, Laxenburg, Austria. Interim Report IR-98-021 / May.

Beckmann, M. J., McGuire, C. B., & Winston, C. B. (1956). *Studies in the Economics of Transportation*. Connecticut , USA: Yale University Press.

Bell, M. (1991). The estimation of origin-destination matrices by constrained generalized least squares, *Transportation Research Part B*, 25, 13-22.

Bikowitz, E., & Ross, S. (1985). Evaluation and Improvement of Inductive Loop Detectors. *Transportation Research Record*, 1010, Transportation Research Board, Washington, D.C., 76-80.

Bruynooghe, M., Gilbert, A., & Sakarovich, M. (1968). Une methode d'affectation du traffic. *Proceedings,4th International Symposium on the Theory of Road Traffic Flow*, Karlsruhe, West Germany.

Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator, *Transportation Research Part B*, 18, 289-299.

Castillo, E., Menéndez, J. M., & Jiménez, P. (2008). Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations, *Transportation Research Part B*, 42, 455–481.

Chaloner, K., & Verdinelli, I. (1995). Bayesian Experimental Design: A Review. *Statistical Science*, 10, 273-304.

Chen, A., Pravinvongvuth, S., Chootinan, P., Lee, M., & Recker, W. (2007). Strategies for selecting additional traffic counts for improving o–d trip table estimation, *Transportmetrica*, 3, 191–211.

Chootinan, P., Chen, A., & Yang, H. (2005). A bi-objective traffic counting location problem for origin-destination trip table estimation, *Transportmetrica*, 1, 65–80.

Clark, J. S., & Gelfand, A. E. (2006). *Hierarchical modeling for the environmental sciences : statistical methods and applications*. Oxford , UK: Oxford University Press.

Clyde, M. A. (2001) Experimental design: A Bayesian perspective, *International Encyclopedia of Social and Behavioral Sciences*, 8, 5075 –5081.

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

Daganzo, C. F. (1982). Unconstrained Extremal Formulation of Some Transportation Equilibrium Problems. *Transportation Science 16,* 332-360.

Daganzo, C. F., & Sheffi, Y. (1977). On Stochastic Models of Traffic Assignment, *Transportation Science*, 11, 253-274.

Dial, R. (1997). Bicriterion traffic assignment: Efficient algorithms plus examples, *Transportation Research Part B*, 31, 357-379.

Ehlert, A., Bell, M. G. H., & Grosso, S. (2006). The optimisation of traffic count locations in road networks, *Transportation Research Part B*, 40, 460–479.

Fisk, C.S. (1988). On combining maximum entropy trip matrix estimation with useroptimal assignment, *Transportation Research Part B*, 22, 69-73.

Fisk, C.S. (1989). Trip matrix estimation from link traffic counts: the congested network case, *Transportation Research Part B,* 23, 331-336.

Fisk, C.S., & Boyce, D.E. (1983). A note on trip matrix estimation from link traffic count data, *Transportation Research Part B,* 17, 245-250.

Hazelton, M. (2000). Estimation of origin-destination matrices from link flows on uncongested networks, *Transportation Research Part B,* 34, 549-566.

Hodgson, M. J. (1990). A flow-capturing location-allocation model, *Geographical Analysis*, 22, 270–279.

Gan, L., & Yang, H. (2001). Integer programming model for optimal selection of screen lines in road networks. *Proceedings of 6th Conference of Hong Kong Society for Transportation Studies,* Hong Kong, 122–130.

Gan, L., Yang, H., & Wong, S. C. (2005). Traffic counting location and error bound in origin-destination matrix estimation problems, *Journal of Transportation Engineering*, 131, 524–534.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science,* 7: 457-551.

Gentili, M. (2002). *New Models and Algorithms for the Location of Sensors on Traffic Networks*, PhD Dissertation, University of Rome La Sapienza, Department of Statistics Probability and Applied Statistics.

Gentili, M., & Mirchandani, P. B. (2005). Locating active sensors on traffic networks, *Annals of Operations Research*, 136, 229–257.

Gentili, M., & Mirchandani, P. B. (2011). Survey of models to locate sensors to estimate traffic flows. *Transportation Research Record*, 2243, 108-116.

Gentili, M., & Mirchandani, P. B. (2012). Locating sensors on traffic networks: Models, challenges and research opportunities. *Transportation Research C*, 24, 227-255.

Gilks, W.R., & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 41: 337-348.

Jörnsten, K., & Nguyen, S. (1979). On the estimation of a trip matrix from network data, , Centre de Recherche sur les Transports, Université de Montréal, Montreal, Publication No. 153.

Kawakami, S., Lu, H., & Hirobata, Y. (1992). Estimation of origin-destination matrices from link traffic counts considering the interaction of the traffic modes, *Papers in Regional Science*, 71, 139-151.

Klein, L. A., Mills, M. K., & Gibson, D. R. (2006). Traffic Detector Handbook: Third edition--Volume I. U.S. Department of Transportation, publication No. FHWA-HRT-06-108.

Lam, W. H. W., & Lo, H. P. (1990). Accuracy of O-D estimates from traffic counts. *Traffic Engineering and Control*, 31, 358-367.

Larsson, T., Lundgre, J. T., & Peterson, A. (2010). Allocation of Link Flow Detectors for Origin-Destination Matrix Estimation-A Comparative Study. *Computer-Aided Civil and Infrastructure Engineering*, 25, 116-131.

LeBlanc, L.J., & Farhangian, K. (1982). Selection of a trip table which reproduces observed link flows, *Transportation Research Part B*, 16, 83-88.

Lindley, D. V. (1972). *Bayesian Statistics-A Review*. Philadelphia, USA: SIAM.

Lo, H.P., Zhang, N., & Lam, W.H.K. (1996). Estimation of an origin-destination matrix with random link choice proportions: A statistical approach, *Transportation Research Part B*, 30, 309-324.

Maher, M. (1983). Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach, *Transportation Research Part B*, 17, 435-447.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. London, UK: Academic Press.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087-1091.

Mirchandani, P. B., & Soroush, H. (1987) Generalized Traffic Equilibrium with Probabilistic Travel Times and Perceptions. *Transportation Science*, 21, 133-152.

Mirchandani, P. B., Gentili, M., & He, Y. (2009). Location of vehicle identification sensors to monitor travel-time performance. *Intelligent Transport Systems IET*, 3, 289–303.

Murchland, J. D. (1969). Road Traffic Distribution in Equilibrium. *Mathematical Methods in the Economic Sciences*, Mathematiches Forschungsinstitut, Oberwolfack, West Germany.

Nguyen, S. (1977). Estimation of an OD matrix from network data: A network equilibrium approach. Centre de recherche sur les Transports, Université de Montreal, Publication 87.

Oppenheim, N. (1995). *Urban Travel Demand Modeling: from Individual Choices to General Equilibrium*. New York, USA: Wiley-Interscience.

Ortúzar, J. D., & Willumsen, L.G. (1994). *Modelling Transport.* West Sussex, England: John Wiley & Sons Ltd.

Peeta, S., & Ziliaskopoulos, A. (2001). Foundations of Dynamic Traffic Assignment: the Past, the Present and the Future. *Networks and Spatial Economics*, 1:2, 233-266.

Peterson, A. (2007). *The OD MAtrix Estimation Problem--Analysis and Computations.* Ph.D Dissertation, Linköping University, Linköping, Sweden.

Roberts, G. O. & Smith, A. F. M. (1993). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications* 49, 207-216.

Sheffi, Y., & Powell, W. B. (1982). An Algorithm for the Equilibrium Assignment Problem with Random Link Times. *Networks*, 12, 191-207.

Sheffi, Y. (1985). *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Englewood Cliffs , New Jersey: Prentice-Hall, Inc.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). *WinBUGS User Manual.* Cambridge, UK: MRC Biostatistics Unit.

Spiess, H. (1987). A maximum-likelihood model for estimating origin-destination matrices, *Transportation Research Part B*, 21, 395-412.

Tamin, O.Z., & Willumsen, L.G. (1989). Transport demand model estimation from traffic counts, *Transportation*, 16, 3-26.

Tebaldi, C., & West, M. (1998). Bayesian Inference of Network Traffic Using Link Count Data, *Journal of the American Statistical Association*, 93, 557–573.

Van Zuylen, H.J., & Willumsen, L.G. (1980). The most likely trip matrix estimated from traffic counts, *Transportation Research Part B*, Vol. 14 (3), 281-293.

Wardrop, J. G. (1952). Some theoretical aspects of road traffic research, *Proceedings, Institute of Civil Engineers, Part II*, 1, 325–378.

Wang, N., Gentili, M., & Mirchandani, P. B. (2012). A Model to Locate Sensors for Estimating Static OD Volumes given Prior Flow Information, *Transportation Research Record: Journal of the Transportation Research Board*, 67-73.

Wang, N., & Mirchandani, P. B. (2013). A sensor location model to optimize OD Estimation using a Bayesian statistical procedure. Accepted by *Transportation Research Record: Journal of the Transportation Research Board*.

Yang, H., Iida, Y., & Sasaki, T. (1991). An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts, *Transportation Research Part B*, 25, 351–363.

Yang, H., Sasaki, T., Iida, Y., & Asakura, Y. (1992). Estimation of origin-destination matrices from link traffic counts on congested networks, *Transportation Research Part B*, 26, 417-434.

Yang, H., & Zhou, J. (1998). Optimal traffic counting locations for origin-destination matrix estimation, *Transportation Research Part B*, 32, 109–126.

Yim, P. K. N., & Lam W. H. K. (1998). Evaluation of count location selection methods for estimation of o–d matrices, *Journal of Transportation Engineering*, 124, 376–383.

Zhou, X. (2004). *Dynamic Origin-Destination Demand Estimation and Prediction for Off-line and On-line Dynamic Traffic Assignment Operation.* Ph.D Dissertation, The University of Maryland, College Park, Maryland, USA.

Zhou, X. & List, G.F. (2010). An Information-Theoretic Sensor Location Model for Traffic Origin-Destination Demand Estimation Applications. *Transportation Science*, 44, 254-273.