



USDOT Region V Regional University Transportation Center Final Report

NEXTRANS Project No. 109PUY2.1

Field Data Based Data Fusion Methodologies to Estimate Dynamic Origin-Destination Demand Matrices from Multiple Sensing and Tracking Technologies

By

Senlai Zhu, Ph.D. Candidate

School of Transportation, Southeast University
Si Pai Lou #2, Nanjing, China, 210096
Tel: +86-15850602236
Fax: +86-025-83795649
Email: zhusenlai@163.com

Yuntao Guo, Ph.D. Candidate

Lyles School of Civil Engineering/NEXTRANS Center
Purdue University
3000 Kent Avenue
West Lafayette, IN 47906
Tel: +1-765-496-9768
Email: guo187@purdue.edu

Hong Zheng, Ph.D.

Lyles School of Civil Engineering/NEXTRANS Center
Purdue University
3000 Kent Avenue
West Lafayette, IN 47906
Tel: +1-765-496-9768
Email: zheng225@purdue.edu

Srinivas Peeta, Ph.D., Professor

Lyles School of Civil Engineering
Purdue University
550 Stadium Mall Drive
West Lafayette, IN 47907
Tel: +1-765-494-2209



Fax: +1-765-496-7996
Email: peeta@purdue.edu

Gitakrishnan Ramadurai, Ph.D., Assistant Professor

Department of Civil Engineering
BSB 229, Transportation Engineering Division
Indian Institute of Technology, Madras
Chennai, India
Tel: +91-44-22574298
Email: gitakrishnan@iitm.ac.in

Jian Wang, Ph.D. Student

Lyles School of Civil Engineering/NEXTRANS Center
3000 Kent Avenue
West Lafayette, IN 47906
Tel: +1-765-496-9768
Email: wang2084@purdue.edu

DISCLAIMER

Funding for this research was provided by the NEXTRANS Center, Purdue University under Grant No. DTRT12-G-UTC05 of the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (OST-R), University Transportation Centers Program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.



USDOT Region V Regional University Transportation Center Final Report

TECHNICAL SUMMARY

NEXTRANS Project No.

Final Report, 31st July 2016

Title

Field Data Based Data Fusion Methodologies to Estimate Dynamic Origin-Destination Demand Matrices from Multiple Sensing and Tracking Technologies

Introduction

Modern technologies can use various types of sensors to collect traffic data; these include GPS, blue tooth, video, automatic vehicle identification (AVI), plate scanning, etc. Based on the characteristic of data collected by sensors, sensors can be categorized as follows. (a) Counting sensors: these sensors can count vehicles on a single lane or a set of lanes in the network. (b) Image/video sensors: these sensors can take images or videos of moving flows. (c) Vehicle-ID sensors: these sensors can be used to identify vehicle IDs in the network. Those sensors/technologies can get variant traffic data including link counts, intersection turning movements, flows and travel time on links and partial paths. This research seeks to propose a Bayesian method and tries to synthesize these multiple sources of data together to estimate dynamic O-D demand, thereby filling a key gap in the current dynamic O-D demand estimation literature.

Findings

The proposed Bayesian method can synthesize multiple sources of data well and provide good estimation. It was shown that the source-specific deviations between estimated and observed traffic counts are all small. It also implies that more traffic counts can lead to smaller variance of the dynamic O-D demand, which means each added traffic count can reduce the uncertainty in the O-D estimation. The proposed Bayesian statistical method can provide not only point estimates of dynamic O-D demand, but also the corresponding statistical information of the estimates. These statistical information can quantize the stochastic of the O-D estimates.

Recommendations

The proposed Bayesian method can effectively synthesize multiple sources of data and estimate dynamic O-D demands with fine accuracy. The proposed model and algorithm can be used to analyze the impacts of various sensor types for dynamic O-D demand estimation. Future, this research serves as a foundational methodology for urban transportation applications.

ACKNOWLEDGMENTS

The authors would like to thank the NEXTRANS Center, the USDOT Region V Regional University Transportation Center at Purdue University, for supporting this research.

TABLE OF CONTENTS

	Page
Table of contents	ii
List of figures	iii
CHAPTER 1. A BAYESIAN METHOD FOR DYNAMIC ORIGIN-DESTINATION DEMAND ESTIMATION SYNTHESIZING MULTIPLE SOURCES OF DATA	1
1.1 Introduction	1
1.2 Literature review	3
1.3 Methodology	6
1.4 Methodology	7
CHAPTER 2. OPTIMAL HETEROGENEOUS SENSOR DEPLOYMENT STRATEGY FOR DYNAMIC ORIGIN-DESTINATION DEMAND ESTIMATION	22
2.1 Introduction	22
2.2 Objectives	25
2.3 Methodology	26
2.4 Results and discussions	37
CHAPTER 3. CONCLUSIONS	47
REFERENCES	49

LIST OF FIGURES

Figure 1.1 Bayesian statistical method	12
Figure 1.2 The Nguyen-Dupuis network	13
Figure 1.3 O-D variance after updating each traffic count in one iteration	14
Figure 1.4 Measures of performance after each iteration	15
Figure 1.5 95% confidence intervals for O-D demand estimates	16
Figure 2.1 Test network	41
Figure 2.2 Traces of covariance matrices of dynamic O-D demand estimation after updating each identified sensor location with different given number of node sensors, without considering the time duration constraint.	42
Figure 2.3 Traces of covariance matrices of dynamic O-D demand estimation after updating each identified sensor location with different given number of node sensors, considering the time duration constraint.	43
Figure 2.4 Traces of covariance matrices of the dynamic O-D demand estimates for the two models after updating the sensor locations under the same time duration constraint.	44

CHAPTER 1. A BAYESIAN METHOD FOR DYNAMIC ORIGIN-DESTINATION DEMAND ESTIMATION SYNTHESIZING MULTIPLE SOURCES OF DATA

1.1 Introduction

Time-dependent origin–destination (O-D) matrices are essential inputs for dynamic traffic assignment (DTA) models. Reliable time-dependent O-D demand matrices have become useful for many real-time traffic planning and management applications, such as the online evaluation of Intelligent Transportation System (ITS) strategies and real-time route guidance. In general, O-D demand matrices can be obtained either from household surveys or estimated by traffic counts. O-D demand obtained by household surveys is not only costly but also vulnerable to become outdated. Thus the use of traffic counts to estimate O-D demand becomes attractive because it is cheap, ease to collect data and to implement.

There is a rich body of literatures estimate static or time-dependent O-D demand using link counts. As the number of independent link counts is usually far less than the number of time-dependent O-D pairs, it results in an underspecified (or degenerate) problem that has no unique solution (Hazelton 2001). Therefore, additional information is needed to acquire a unique solution for the O-D estimation problem. Modern technologies can use various types of sensors to collect traffic data; these include GPS, blue tooth, video, automatic vehicle identification (AVI), plate scanning, etc. Those sensors/technologies can get variant traffic data including link counts, intersection turning movements, flows and travel time on links and partial paths.

Based on the characteristic of data collected by sensors, sensors can be categorized as follows. (a) Counting sensors: these sensors can count vehicles on a single lane or a set of lanes in the network, including inductive loop detectors, magnetic detectors, etc. One can use vehicle count data to measure traffic characteristics such as speed, density, occupancy, and flow rates. (b) Image/video sensors: these sensors can take images or videos of moving flows. For example, a fixed camera or video can be used to measure flows at an intersection. By processing the images or videos, it is able to collect positions of moving vehicles in the scene. (c) Vehicle-ID sensors: these sensors can be used to identify vehicle IDs in the network. For instance, license plate readers which use camera images, or Automatic Vehicle Identification (AVI) readers which use radio-frequency identification (RFID) tags or bar-codes, can be deployed over lanes or on the

roadside to detect vehicles. GPS can be deployed on vehicles to track its route. With image based and vehicle-ID based sensors, full or partial path information such as path flows or travel times can be inferred.

Although various sources of data can be used in O-D demand estimation, most literatures tend to use just one source of data, or combine it with link counts; very few studies synthesize multiple sources of data together to estimate O-D demand. Synthesizing multiple sources of data is difficult because they are correlative and complementary each other thus cannot be simply combined together. Hence, the statistical correlations among them should be analyzed. Moreover, most literatures do not make use of travel times (travel times on links or partial paths) in O-D estimation. This is because the connections between travel time and O-D demand cannot be measured directly. However, travel time data can often be collected much more easily (e.g., by Vehicle-ID sensors) than the volume data (especially along a path or a partial path). Therefore, it is worthwhile to consider the travel time data in O-D demand estimation.

To bridge the gap above, this paper tries to synthesize multiple sources of data together, mainly including link counts, time-varying flows and travel time along partial observed paths, and turning movements at intersections, to estimate dynamic O-D demand. Specifically, we treat each time-dependent O-D demand as a random variable satisfying normal distribution, and propose a Bayesian statistical model to estimate dynamic O-D demand by synthesizing these multiple sources of data. By solving the dynamic user equilibrium (DUE) problem based on an assumed prior O-D demand, the prior distribution (including an vector of expected values and a variance-covariance matrix) of all considered variables is estimated. The relationships among all variables are analyzed by variance-covariance matrices. By updating the assumed prior distribution of all variables using traffic counts, we establish the posterior distributions of all variables, based on which point estimation and probability confidence intervals are inferred to measure the intrinsic uncertainty. In the proposed Bayesian statistical model, we convert the observed sub-path travel time to several sub-path flows so as to incorporate sub-path travel time information in O-D estimation. Specifically, for a sub-path with a given departure time, we sample the normal distributed sub-path travel time to get arrival time for each user, and the mean of all the normal distributed sub-path travel times is equal to the observed sub-path travel time. By this sampling method we convert the sub-path travel time information to sub-path flows which is more appropriately analyzed in O-D estimation.

1.2 Literature review

1.2.1 *Static O-D demand estimation*

O-D demand estimation was studied extensively for static case. It usually uses a prior matrix (or seed matrix) in order to obtain a unique solution. Overall, these methods can be classified as:

(1) Least squares (Cascetta 1984; Cascetta and Nguyen 1988; Doblaz and Benitez 2005) and generalized least squares (GLS) (Bell 1991; Nie and Zhang 2010; Caggiani et al. 2013) methods. These methods are usually bi-level problems. The upper level is to minimize the weighted distances between the target and estimated OD demands, and between the measured and estimated traffic volumes; the lower level model is a static user equilibrium problem.

(2) Entropy concept based methods (Van and Willumsen 1980; Xie et al. 2011). These methods maximize the entropy subject to a set of constraints. The entropy concept measures how reasonable and close to reality of an estimated O-D matrix. Subject to the prior O-D matrix, the probability distribution of O-D demand which best represents the current state of knowledge is the one with the maximum entropy.

(3) Maximum likelihood methods (Spiess 1987; Parry and Hazelton 2012). These methods maximize the likelihood of the prior O-D matrix and the observed traffic counts conditional on the estimated O-D matrix. The elements of the prior O-D matrix are assumed to be obtained from a set of random variables with given probability distribution.

(4) Bayesian inference (Maher, 1983; Hazelton 2001; Hazelton 2010; Perrakis et al. 2012; Wei and Asakura 2013) and Bayesian network (Tebaldi and West 1998; Castillo et al. 2008b, c; Cheng et al. 2014) methods. These methods treat traffic flow as random variables. Using observed traffic counts to update the assumed prior distribution, the posterior distribution of all variables is built based on Bayes theorem.

1.2.2 *Dynamic O-D demand estimation*

Estimating time-dependent O-D demand is more complicated than the static O-D case due to its time-varying characteristic. Some studies straightforwardly extend methods of static O-D estimation to the dynamic case using time-varying link counts. For example, Cascetta et al. (1993) proposed a GLS method to estimate dynamic O-D demand based on a simplified

assignment model on a general network. Following Cascetta et al. (1993), Sherali and Park (2001) proposed a constrained least squares (CLS) formulation but solved for path flows rather than O-D demand. Based on the bi-level formulation proposed by Fisk (1988) for the static O-D estimation, Tavana and Mahmassani (2001) proposed a bi-level generalized least squares optimization model with an iterative solution framework to estimate the dynamic O-D demand. Based on a least square modeling approach, Bierlaire and Crittin (2004) proposed an algorithm for sparse least squares that is computationally efficient to favor real-time estimation and prediction of dynamic O-D demand. Okutani (1987) first introduced state space model into dynamic O-D estimation with the state vector indicating the unknown O-D flows. Since then, the state-space model is further studied by Ashok and Ben-Akiva 1993, Zhou and Mahmassani (2007), Cho et al. (2009), etc.

Note that most existing methods for dynamic O-D estimation problem are characterized by a bi-level optimization structure. The upper-level problem is to minimize two deviation functions: (1) the deviation between observed and estimated traffic counts over all time intervals, and (2) the deviation between the target or historical demand and estimated demand matrices. The lower-level problem is the DTA problem, which determines a time-dependent network flow pattern that satisfies dynamic user equilibrium (DUE) condition. For example, Kattan and Abdulhai (2006) proposed a non-iterative approach to dynamic O-D estimation based on a machine-learning technique using advanced parallel evolutionary algorithms. Balakrishna et al. (2008) and Cipriani et al. (2011) introduced gradient approximation methods within a simultaneous perturbation stochastic approximation framework in order to reduce the number of simulation runs when calculating numerical derivatives or gradients. Huang et al. (2012) developed an approach to estimate travel demand in a large-scale microscopic traffic simulation model based on a Guided Genetic algorithm with a distributed implementation to improve computational efficiency and reduce memory requirements. Recently, Tympakianaki et al. (2015) applied a Cluster-wise simultaneous perturbation stochastic approximation algorithm to the dynamic O-D estimation.

Meanwhile, single-level formulations have also been proposed for the dynamic O-D demand estimation problem. For instance, Nie and Zhang (2008) formulated a novel single-level formulation based on variational inequalities (VI), which utilizes the dynamic link-path incidence relationships in a generic projection-based VI solution framework. Based on a single

level reformulation, Lundgren and Peterson (2008) proposed a heuristic algorithm to address the dynamic O-D demand estimation problem, which is an adaptation of the projected gradient method.

1.2.3 O-D demand estimation using multiple source data

Due to recent advances in real-time sensing technologies, a variety of sensors can be used to collect different types of traffic data, such as GPS, Bluetooth, video camera, AVI, plate scanning, etc. With these sensing technologies, one can process the data to obtain traffic data measurements on links, turning movement counts at nodes, and full or partial vehicle trajectories along the path.

Turning movements at intersections are normally detected by Image/video sensors. Intersection turning movements provide more information on users' travel behavior and usage of network topologies than link counts. Several studies estimate O-D demand by using turning movements at intersections. For instance, Yang et al. (1998) proposed a neural network approach to estimate dynamic O-D demand using node-based traffic counts. Alibabai and Mahmassani (2009) presented a dynamic O-D estimation model based on a bi-level optimization method that utilized both turning movement volumes and link volumes. Lu et al. (2014) proposed a Kalman filter approach to estimate dynamic O-D demand using link counts and observed turning movements.

To estimate O-D demand, path-based information is more desirable because they can fully reflect users' route choice behavior and network topology. However, path information cannot be fully detected, so researchers normally make use of observed flows or data on partial path, which can be captured by GPS, mobile phone, plate scanning, AVI, etc. For instance, to estimate static O-D demand, Parry and Hazelton (2012) proposed a likelihood-based statistical model combining link counts and sporadic path data. Castillo et al. (2013) presented a Bayesian method based on plate scanning. Recently, Hu et al. (2015) proposed link-based and path-based models to estimate O-D demand based on traffic counts by vehicle detector sensors and license plate recognition. In dynamic case, Dixon (2000) proposed a three-stage procedure to estimate O-D demand from transponder-based AVI data. Recognizing low identification rates associated with license-plate based AVI data, Van der Zijpp (1997) proposed a constrained optimization formulation to estimate the unknown O-D demand and identification rates jointly. Antoniou et al.

(2004) introduced path-flow proportion matrices that relate O-D demand to sub-path tag counts, and extended Ashok's framework (1996) to estimate and predict tagged vehicular O-D demand. Zhou and Mahmassani (2006) proposed a nonlinear ordinary least-squares model to systematically combine AVI counts, link counts, and historical O-D demand information.

In real application, observing partial path or sub-path flow could be difficult and costly. However, path-based travel time can be observed much easily and thus can be used in O-D estimation. For example, based on the transponder tag data collected from a freeway corridor in Houston, Dixon and Rilett (2002) applied the framework developed by Cascetta et al.(1993) to calculate the link-flow proportions based on the observed travel time from AVI counts. By adapting the analytical approach of Ghali and Smith (1995) for evaluating the local link marginal travel times, Qian and Zhang (2011) incorporated the travel time gradients into the single-level O-D estimation framework proposed by Nie and Zhang (2008) in order to utilize travel time measurements.

Except for the above measurements, O-D demand is also estimated by using other types of traffic information. For instance, Chang and Wu (1994) made use of flow counts across screen lines and cordon lines in dynamic O-D demand estimation. Since speed and density provide the best representation of traffic congestion, some researchers made use of these traffic measures to estimate dynamic O-D demand (see, for example, Balakrishna 2006; Lu et al. 2013). Additionally, Iqbal et al. (2014) and Lauren et al (2015) made use of mobile phone data to infer O-D trips.

1.3 Methodology

To estimate the dynamic OD demand in Chennai, a Bayesian statistical method is used. And we do the following assumptions:

Assumption 1 : The traffic demands between all time-dependent OD pairs are assumed to follow multivariate normal (MVN) distributions.

Assumption 2: It is assumed that the time-dependent path flow is the product of path choice proportion and the time-dependent OD matrices, where the path choice proportion is a deterministic variable, which can be obtained from solving the dynamic user equilibrium problem. And the time-dependent choice proportion of sub-path, turning movement at intersection and link can be derived from the path choice proportion.

Based on assumption 2, the whole sets of random variables involved in our model are related by the linear expression:

$$\begin{pmatrix} \mathbf{D} \\ \mathbf{F} \\ \mathbf{F}_{sub} \\ \mathbf{S} \\ \mathbf{V} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{P} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\varphi}\mathbf{P} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\phi}\mathbf{P} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \boldsymbol{\Psi}\mathbf{P} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{D} \\ \boldsymbol{\varepsilon} \\ \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{pmatrix} \quad (1)$$

where \mathbf{D} , \mathbf{F} , \mathbf{F}_{sub} , \mathbf{S} , \mathbf{V} are vectors of time-dependent OD demand, path flows, sub-path flows, intersection turning movements and link flows, respectively. \mathbf{P} , $\boldsymbol{\varphi}$, $\boldsymbol{\phi}$ and $\boldsymbol{\Psi}$ are the corresponding choice proportion matrices, which can be obtained by solving the dynamic traffic assignment problem. $\boldsymbol{\varepsilon}$, $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are the error terms and their expected values are all zero.

Then according to assumption 1, the prior distribution of all variables can be derived from the historical data. Then based on the traffic counts, the posterior distribution can be obtained by using the following updating formula (Maher, 1983):

$$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}=\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{Y}} + \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}}) \quad (2)$$

$$\boldsymbol{\Sigma}_{\mathbf{YZ}|\mathbf{X}=\mathbf{x}} = \boldsymbol{\Sigma}_{\mathbf{YZ}} - \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XZ}} \quad (3)$$

where \mathbf{Y} and \mathbf{Z} both refer to the sets of all variables; \mathbf{X} is the set of observed variables, respectively; $\boldsymbol{\mu}_{\mathbf{X}}$ and $\boldsymbol{\Sigma}_{\mathbf{XX}}$ are the mean vector and covariance matrix of the observation \mathbf{X} ; $\boldsymbol{\Sigma}_{\mathbf{YZ}}$ is the covariance matrix of \mathbf{Y} and \mathbf{Z} ; $\boldsymbol{\Sigma}_{\mathbf{XZ}}$ is the covariance matrix of \mathbf{Z} and \mathbf{X} ; $\boldsymbol{\mu}_{\mathbf{X}}$ and $\boldsymbol{\Sigma}_{\mathbf{XX}}$ are the mean vector and covariance matrix of the observation \mathbf{X} ; $\boldsymbol{\Sigma}_{\mathbf{YX}}$ is the covariance matrix of \mathbf{Y} and \mathbf{X} ; \mathbf{x} is the actual observed value of \mathbf{X} .

To simplify the calculation, we updated the traffic counts one by one. In this case, matrix inversion is avoided. Figure 1.1 shows the whole process of our method.

1.4 Methodology

We demonstrate the proposed method using Nguyen–Dupuis network, as shown in Figure 1.2. It consists of 13 nodes, 38 bidirectional links. Six nodes {12, 1, 4, 8, 2, 3} are terminal nodes, which could be either origins or destinations. Vehicles can travel from left to right (from origins {12, 1, 4} to destinations {8, 2, 3}) or from right to left (from origins {8, 2, 3} to destinations {12, 1, 4}). So in total there are 18 O-D pairs. The O-D matrix is time-dependent with 15-minute intervals and the number of time intervals is six. Demand for each O-D pair at each departure time is 30 in the seed matrix. We suppose the “true” O-D demand matrix is

known, which is generated from the seed matrix randomly. The observed data is assumed to be collected by sensors in the network. Specifically, we assign the “true” matrix in the network by DUE method and place sensors in the network to obtain the sensor data, which collects flows on 4 sub-paths, time-dependent turning movements at 18 intersections and time-dependent link counts on 30 links. These sensor data, as tabulated in Table 1, serve as our observed data. In such a manner the observed data is consistent with the “true” matrix and assignment method in the model. We then try to estimate time-dependent O-D reversely from the observed data to match the “true” matrix. The DUE method is a standalone procedure in the model, which can be solved by off-the-shelf traffic software. In this paper we used a dynamic assignment and simulation model - DYNASMART-P 1.3.0 - to solve DUE.

To measure the performance of the proposed Bayesian method and the algorithm, three aggregate measures were used: the percentage root-mean-square error (%RMSE), the mean absolute error (MAE) and Theil’s inequality coefficient \mathbf{U} (Toledo and Koutsopoulos, 2004) for traffic counts, to measure the fit between estimated and observed traffic counts:

$$\%RMSE = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N (y_n^{est} - y_n^{obs})^2}}{\frac{1}{N} \sum_{n=1}^N y_n^{obs}} \times 100\% \quad (4)$$

$$MAE = \frac{1}{N} \sum_{n=1}^N |y_n^{est} - y_n^{obs}| \quad (5)$$

$$\mathbf{U} = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N (y_n^{est} - y_n^{obs})^2}}{\sqrt{\frac{1}{N} \sum_{n=1}^N (y_n^{est})^2 + \frac{1}{N} \sum_{n=1}^N (y_n^{obs})^2}} \quad (6)$$

where N is the number of measurements, y_n^{est} is the estimated measurement, and y_n^{obs} is the observed measurement. Note that the value of \mathbf{U} is between zero and one. $\mathbf{U} = 0$ implies a perfect fit between the estimated and observed measurements, while $\mathbf{U} = 1$ indicates the worst possible fit.

Similarly, to measure the fit between estimated and “true” O-D demand, three measures were used as following:

$$\%OD_RMSE = \frac{\sqrt{\frac{1}{M} \sum_{m=1}^M (d_m^{est} - d_m^{obs})^2}}{\frac{1}{M} \sum_{m=1}^M d_m^{obs}} \quad (7)$$

$$OD_MAE = \frac{1}{M} \sum_{m=1}^M |(d_m^{est} - d_m^{obs})| \quad (8)$$

$$OD_U = \frac{\sqrt{\frac{1}{M} \sum_{m=1}^M (d_m^{est} - d_m^{obs})^2}}{\sqrt{\frac{1}{M} \sum_{m=1}^M (d_m^{est})^2 + \frac{1}{M} \sum_{m=1}^M (d_m^{obs})^2}} \quad (9)$$

where M is the number of O-D pairs, d_m^{est} is the estimated O-D demand, and d_m^{obs} is the “true” O-D demand.

Start from the seed matrix and the sensor data, the time-dependent O-D demand is estimated by the procedure introduced in Section 5. The value of α in step 1 is 0.5, λ in step 4.1 is 1.0 and ρ in step 7 is 0.1.

The total O-D variance is the sum of variance of each time dependent O-D demand. Figure 3 shows how the total O-D variance changes within one iteration after the traffic count is updated one by one. It shows that the total O-D variance is decreasing with each added and updated traffic count. Smaller variance means lower uncertainty in the estimation, so updating each traffic count can improve accuracy of O-D estimation. Figure 1.3 also indicates that more traffic counts can lead to lower variance of the dynamic O-D demand estimation, since more updated information can be used to improve the O-D estimation. Because traffic counts are updated one by one in the proposed algorithm, when new traffic data comes in, no need to resolve the Bayesian statistical model from scratch but just need to continue to update the procedure with the additional traffic data. In real-world applications, we can measure the quality of traffic data by analyzing the resultant variance of the dynamic O-D demand estimation, so as to determine whether to add additional traffic data in the procedure or not.

Figure 1.4 illustrates how the measures of traffic count performance change at each iteration. It shows that three measures of performance have similar trends and they are all decreasing by iterations (although there are small fluctuations). This indicates that the proposed method normally can identify a solution that reduces the total error of traffic counts compared to

that of last iteration. There are 30 iterations shown in Figure 4, till which the three measures of performance become flat. Noticeably, after 30 iterations, %RMSE has been reduced from around 36% to 6%, MAE has been reduced from around 51.00 to 7.00, and Theil's inequality coefficient U has been reduced substantially from the initial value 0.25 to 0.035. These results demonstrate the high quality of the proposed Bayesian method for dynamic O-D estimation.

Tables 1.1, 1.2 and 1.3 show the relative errors between the estimated and observed sub-path travel time, node turning movements and link flows respectively. Note that the relative errors of almost 50% traffic counts are less than 5%, and the relative errors of 85% traffic counts are less than 10%. Note that relative errors of traffic counts with high values are small. Specifically, for the sub-path travel time, the relative errors are all small and less than 5%. For the node turning movement, relative errors of about 80% estimation are less than 10%. For the link counts, about 80% of the relative errors are less than 10% and over 50% of them are less than 5%. These results indicate that not only the total error of the estimation obtained by the proposed Bayesian method is reduced significantly (as also shown by Figure 3), but also the errors for each type of traffic counts are small. This demonstrates that the proposed Bayesian method can synthesize multiple sources of data well and provide good estimation.

The proposed method can provide not only the point estimates (i.e., expected values), but also the variances, which represent the associated uncertainty for the corresponding O-D demand. Based on the expected values and variances, we can obtain the posterior distribution of the time-dependent O-D demand. According to the posterior distribution, the confidence intervals for each O-D demand can be identified. In summary, the proposed Bayesian statistical method can provide not only point estimates of dynamic O-D demand, but also the corresponding statistical information of the estimates.

Based on the posterior distribution of the time-dependent O-D demand, Figure 1.5 shows the 95% confidence intervals for each time-dependent O-D demand estimates. Note that because most variances are small, the lengths of the 95% confidence intervals are also small, which means low uncertainty involved in the estimated O-D demand. It can be seen that the length of intervals for some O-D demand estimates are much smaller than lengths of others. This is because according to the traffic assignment proportions of traffic counts (i.e., φ , ϕ and Ψ in Eq. (31)), some O-D pairs have much more traffic counts related to them. In such a case, the corresponding variance and length of confidence interval could be small since it has more

information to update them and reduce the variability. In fact, if we have more observed traffic counts, the variances will be even smaller and the resultant O-D demand estimates have even lower uncertainty (as demonstrated by Figure 1.2). This gives a hint to determine which links, nodes and/or paths need to be observed when estimating traffic flows by the Bayesian method, that is, the network sensor location problem. We can locate sensors on a set of links, nodes and/or paths which lead to lower uncertainty of the dynamic O-D demand estimation. Since this problem is out of the scope of this paper, we leave this problem for future research.

Table 1.4 shows the values of three measures (as shown in Eqs. (38-40)) related to the performance of O-D demand estimates. According to the three measures, it can be seen that the total error between the estimated and “true” O-D demands is relatively small. For example, the resultant OD_U is around 0.1. It also can be seen that the errors between the estimated and “true” O-D demands in some time intervals are larger than errors in other time intervals. This is because in those time intervals with larger errors, only a few traffic counts are related to the O-D pairs in the considered time intervals. Thus the precision of the O-D demand estimates is much lower.

To further study the impact of the number of traffic counts on the precision of the O-D demand estimates, take the O-D demand estimates in time interval 2 for example. Table 1.5 compares performance of O-D demand estimates with different number of traffic counts related to the considered time interval. For comparison purpose, only traffic counts on links are considered in Table 1.2. When users from an O-D pair in time interval 2 use a link collecting traffic counts, traffic count on this link is treated as related to the O-D pair in time interval 2. From Table 1.5, it can be seen that when the number of related traffic counts increases, the error between the estimated and “true” O-D demands decreases, as demonstrated by the changes of the three measures’ values. Thus, if we have more observed traffic counts, the errors will be even smaller and the resultant O-D demand estimates have even higher accuracy. In summary, according to Figure 1.4 and Table 1.4, more observed traffic counts can lead to lower uncertainty and higher precision of O-D demand estimates.

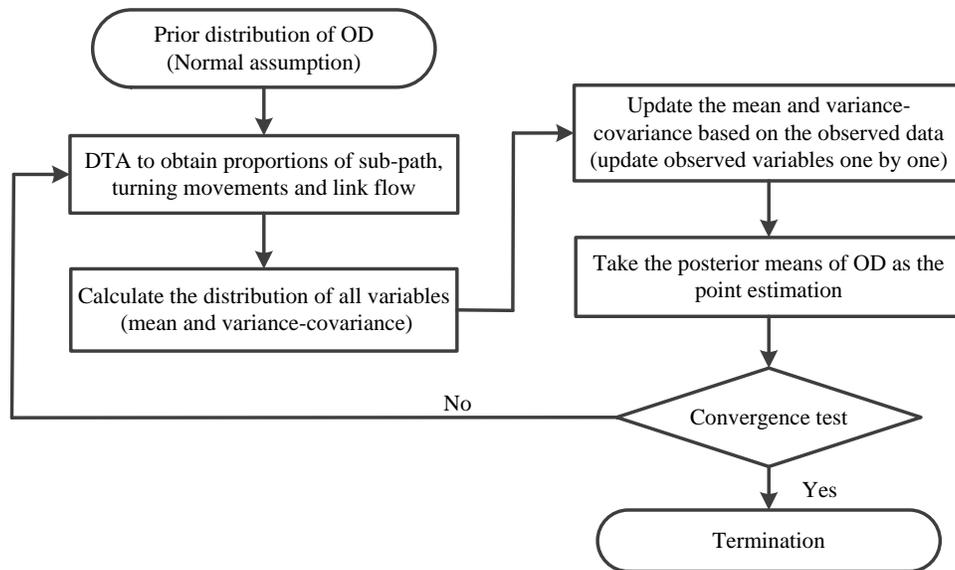


Figure 1.1 Bayesian statistical method

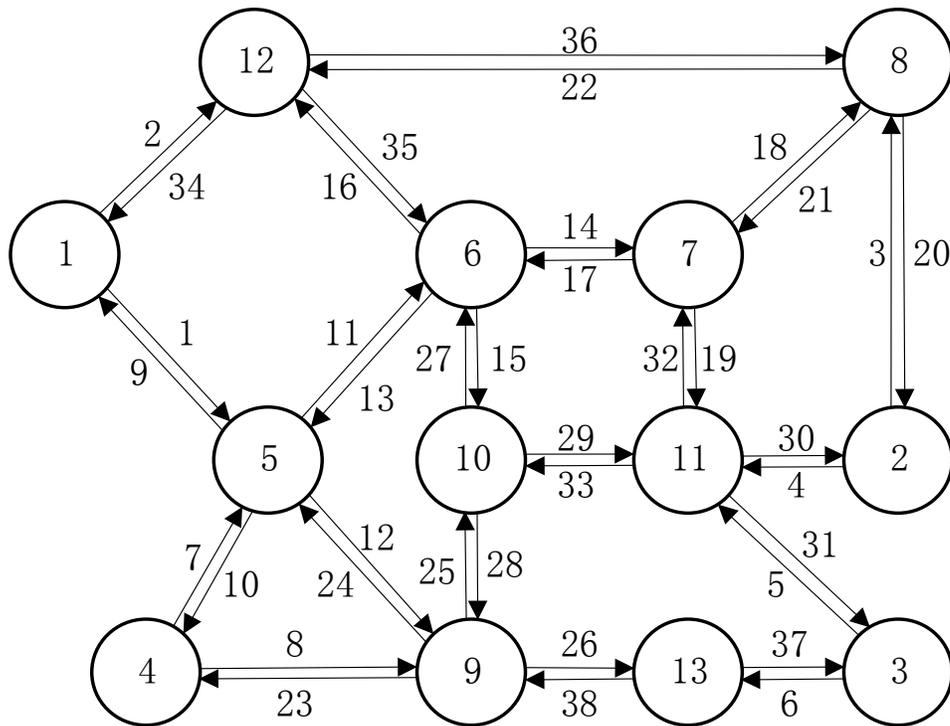


Figure 1.2 The Nguyen-Dupuis network

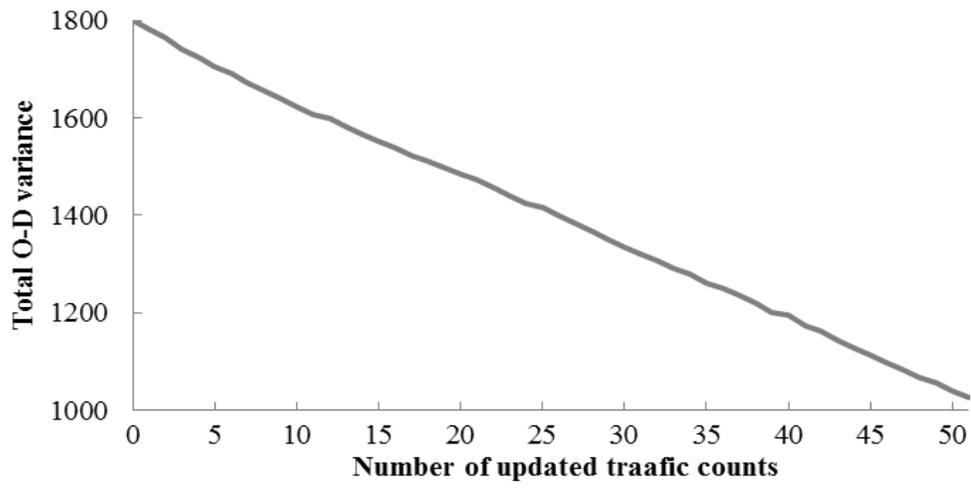


Figure 1.3 O-D variance after updating each traffic count in one iteration

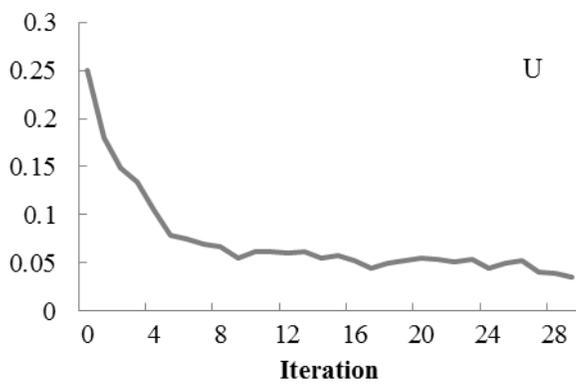
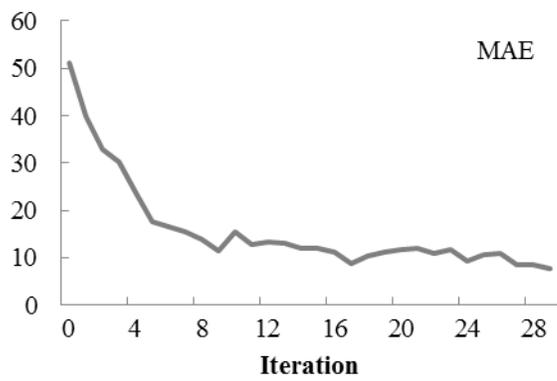
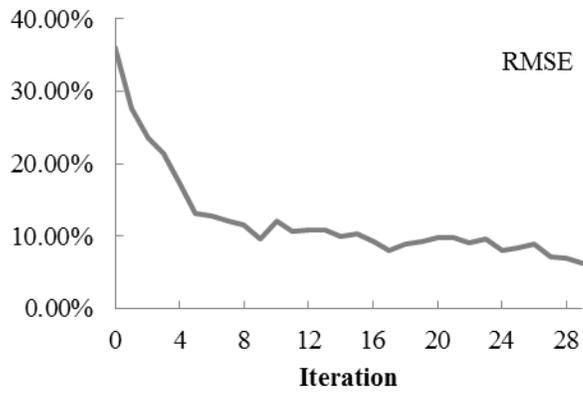


Figure 1.4 Measures of performance after each iteration

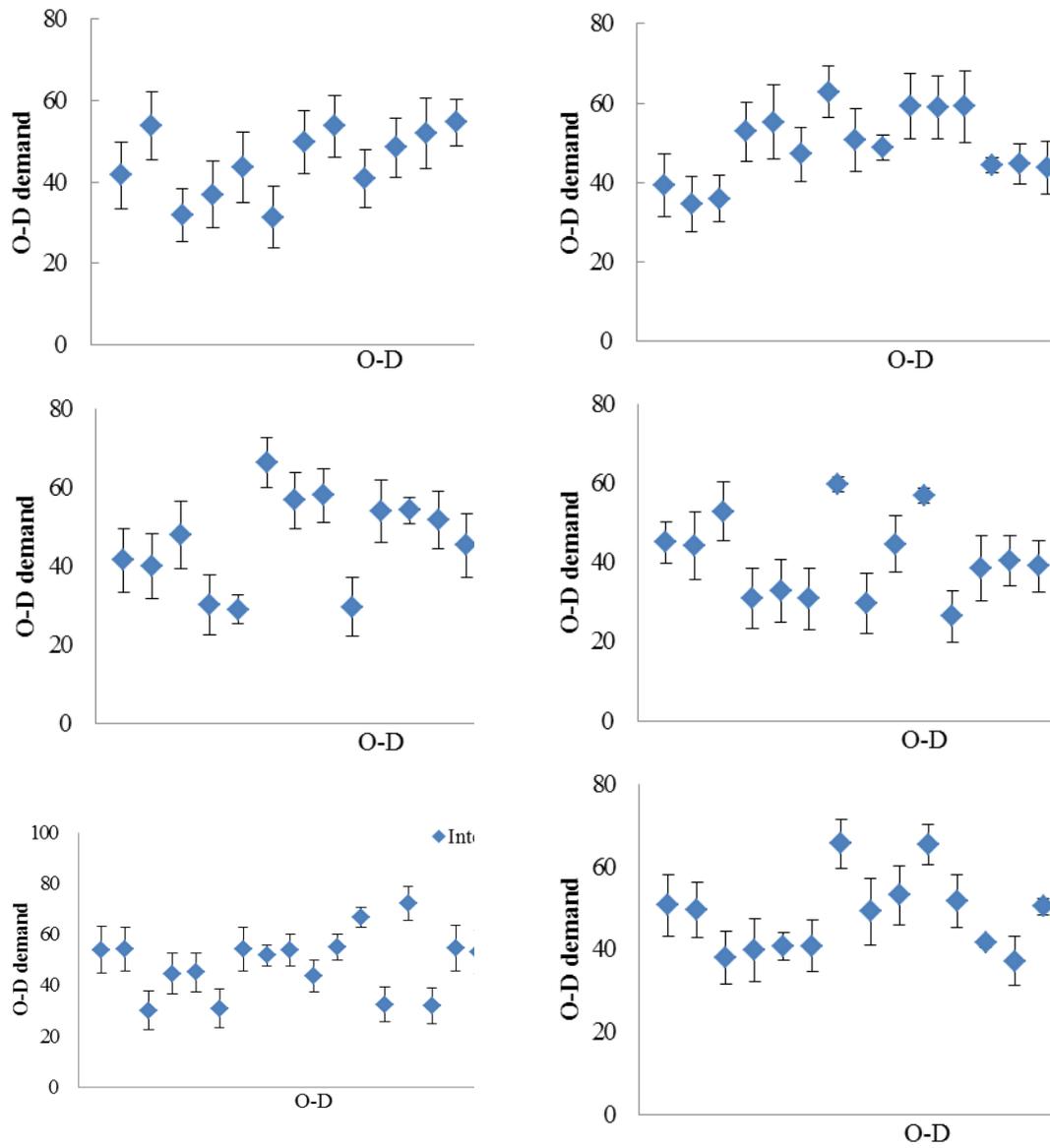


Figure 1.5 95% confidence intervals for O-D demand estimates

Table 1.1 Observed and estimated sub-path travel time

Sub-Path	Departure time	Observed	Estimated	Relative error (%)	Sub-Path	Departure time	Observed	Estimated	Relative error (%)
5-6-7-8	1	7.5	7.45	0.67	5-9-10-11	2	7.5	7.34	2.13
5-6-7-8	3	5	4.83	3.4	5-9-10-11	4	7	7.1	1.43

Table 1.2 Observed and estimated node turning movements

Turning movement	Departure time	Observed	Estimated	Relative error (%)	Turning movement	Departure time	Observed	Estimated	Relative error (%)
2-11-10	0	63	68	8.00	7-6-5	3	36	27	25.00
10-6-12	1	41	51.21	24.90	7-6-12	4	40.72	44	8.06
3-11-10	1	91	82.07	9.81	7-6-5	4	127.26	120.02	5.69
2-11-10	1	124	132	6.45	2-11-7	4	57.92	63	8.77
7-6-5	2	39	40	2.56	2-11-10	4	73.42	74.34	1.25
7-11-3	2	50	47	6.00	7-6-12	5	70.86	75.7	6.83
3-11-10	2	35	32.67	6.66	7-6-5	5	108.25	105.87	2.20
2-11-10	2	138	141	2.17	2-11-10	5	52	53.66	3.19
10-6-12	3	35	36	2.86	7-6-12	6	34.63	32.75	5.43

Table 1.3 Observed and estimated link flows

Link	Departure time	Observed	Estimated	Relative error (%)	Link	Departure time	Observed	Estimated	Relative error (%)
8-12	0	143.06	138.28	3.34	3-13	3	140.09	147.18	5.06
5-9	0	161	163.11	1.31	3-13	4	142.35	150.56	5.77
3-13	0	139	163.42	17.57	5-6	4	140	142	1.43
4-5	0	137	146.5	6.93	6-5	4	179.26	176.59	1.49
12-8	1	154.26	161.94	4.98	7-8	4	167.25	153	8.52
8-12	1	189.81	192.39	1.36	8-7	5	159.23	166.01	4.26
5-9	1	285	282.8	0.08	12-8	5	216.47	194.13	10.32
9-5	1	240	260.61	8.59	8-12	5	186.88	189.59	1.45
8-12	2	147.23	147.96	0.50	1-12	5	195.59	172.38	11.87
12-1	2	136	160.04	17.68	3-13	6	93.34	96.43	3.31
5-9	2	339.94	337.66	0.67	12-8	6	141.67	121.21	14.44
9-5	2	322.75	311.92	3.35	8-12	6	128.86	138.79	7.71
1-12	3	99.67	88.73	10.98	12-1	6	107.41	111.69	3.99
5-9	3	284.45	309.13	8.68	5-6	6	21	25	19.05
9-5	3	100.96	97.64	3.29	9-13	6	35.11	34	3.16

Table 1.4 Comparison between estimated and true O-D demand

Time interval	%OD_RMSE	OD_MAE	OD_U
0	22.17%	8.36	0.11
1	34.97%	13.00	0.18
2	31.42%	11.53	0.16
3	29.22%	11.77	0.15
4	15.07%	5.53	0.07
5	23.20%	9.29	0.11
Total	25.30%	9.42	0.12

Table 1.5 Performance of O-D demand estimates with different number of traffic counts (time interval = 2)

Number of related traffic counts	%OD_RMSE	OD_MAE	OD_U
4	33.69%	12.87	0.18
8	31.55%	11.66	0.17
12	21.04%	8.09	0.11
16	18.95%	7.76	0.10

CHAPTER 2. OPTIMAL HETEROGENEOUS SENSOR DEPLOYMENT STRATEGY FOR DYNAMIC ORIGIN-DESTINATION DEMAND ESTIMATION

2.1 Introduction

O-D demand has been estimated using traffic data measurements which can be obtained through sensor installation, as introduced in chapter 1. However, in real-world applications, sensors cannot be installed on every link, node and/or path in the network due to limited budget. This motivates the need to optimally determine the locations of sensors in the network. As the locations of sensors in the network can significantly affect the accuracy and reliability of the O-D demand estimates, the network sensor location problem (NSLP) has received a lot of attention in recent years.

Most existing NSLP models are designed for static O-D demand estimation, and sensors are usually located on links only. Because the actual O-D demand is usually unknown in determining the sensor locations, indirect quality measures that do not need knowledge of the exact O-D demand are used in most existing NSLP models. For example, Lam and Lo (1990) proposed to use traffic flow volume and O-D coverage criteria to determine priorities for locating sensors. Yang et al. (1991) introduced the maximum possible relative error (MPRE) criterion to calculate the most possible deviation of the estimated O-D demand from the unknown true O-D demand. Yang and Zhou (1998) further proposed four basic location rules, namely the maximal flow fraction rule, the O-D covering rule, the maximal flow interception rule, and the link independence rule. Yim and Lam (1998) evaluated several of these rules on a large-scale network. Bianco et al. (2001) proposed an iterative two-stage procedure and several priority-based greedy heuristics to cover the O-D demand and reduce the MPRE value. Gan et al. (2005) introduced a modified MPRE formulation, termed the expected relative error (ERE), to represent

the expected error between the true and estimated O-D demands. Bierlaire (2002) proposed the total demand scale (TDS) measure to calculate the difference between the maximum and minimum possible total demand estimates, which can be used for both static and dynamic O-D demand estimation. Chen et al. (2012) extended the TDS measure to consider the quality measure at different spatial levels, and labeled it the generalized demand scale (GDS) measure. Simonelli et al. (2012) proposed a synthetic dispersion measure (SDM), which is related to the trace of the covariance matrix of the posterior demand estimates conditional upon a set of sensor locations. Based on the trace of the covariance matrix of the posterior traffic flow estimates, Zhu et al. (2014, 2015) proposed a stepwise method to identify sensor locations for the traffic flow estimation, including the O-D demand, path flows and the unobserved link flows.

Estimating time-dependent (dynamic) O-D demand is substantially more complicated compared to the static O-D estimation problem due to the time-varying characteristic. Hence, the corresponding NSLP is also difficult to address. Eisenman et al. (2006) proposed a conceptual framework for the sensor location problem to minimize the error in the real-time O-D demand estimates. Fei et al. (2007) extended Eisenman et al.'s (2006) approach to examine the NSLP under two different scenarios (with and without budget constraints). The TDS measure proposed by Bierlaire (2002) to locate sensors for the static O-D estimation problem was also used in their study to estimate the time-dependent O-D demand.

The NSLP has been investigated to handle different types of measurements such as AVI readers and license plate recognition techniques, mainly for the static O-D estimation problem. For instance, Chen et al. (2004) proposed a multi-objective model for locating AVI readers on the network, which was extended by Chen et al. (2010) to accommodate different travel demand patterns. Minguez et al. (2010) sought to optimize the traffic plate scanning locations for O-D

demand and route flow estimation under budget constraints. The traffic plate scanning location problem was also studied by Castillo et al. (2013). Yang et al. (2006) and Chen et al. (2007) proposed addressing the screen line-based traffic counting location problem. Hu et al. (2015) proposed a bi-level optimization model to solve the NSLP and determine an optimal deployment strategy for heterogeneous sensors (vehicle detector sensors and license plate recognition). For the dynamic O-D demand estimation problem, Asakura et al. (2000) provided an off-line least-squares model to simultaneously determine the O-D demand and the identification rates of AVI data based on the locations of the AVI readers. Zhou and List (2010) proposed a model for locating a limited set of traffic counting stations and AVI readers in a network so as to maximize the expected information gain for the dynamic O-D demand estimation problem. Zhu et al. (2016) proposed a NSLP model for dynamic O-D demand estimation to determine optimal heterogeneous sensor locations (link sensors and node sensors).

In addition to its use in the traditional O-D demand estimation problem, NSLP has been used in other related domains. For example, Viti et al. (2008) solved the sensor location problem for the travel time estimation problem. Xing et al. (2013) proposed an information-theoretic sensor location model to minimize total travel time uncertainty. Bianco et al. (2014) applied a genetic algorithm approach to identify sensor locations for estimating all link flows in the network. Castillo et al. (2008a) and Viti et al. (2014) discussed the observability problem, to identify the set of sensor locations that would enable full O-D demand observability. He (2013) proposed a graphical approach to locate sensors for link flow inference. Hu et al. (2009) suggested a procedure that does not require any prior O-D demand matrix but entails explicit path enumeration, for the identification of all link flows using traffic data measurements on an

optimal subset of links under a budget constraint. Ng (2012) relaxed the assumption of explicit path enumeration by using a node-based formulation for the Hu et al. (2009) problem context.

In the literature, most NSLP models are restricted to one type of sensor, especially for the dynamic O-D demand estimation problem, where only traffic data measurements collected by link sensors has been considered. In addition, past studies have not considered the impact on the optimal sensor deployment strategy due to the time duration for which traffic data measurements are available. For example, past studies require traffic data measurement for a link for the entire time period of interest (such as the peak period, which can be in the order of hours) related to determining the dynamic O-D demand to identify the optimal sensor deployment strategy. However, in practice, traffic data measurements may be available for that link for only a much shorter time period (for example, one hour); the NSLP models proposed in past studies are restrictive in that they cannot be used to determine the optimal sensor deployment strategy without traffic data measurements for the entire time period. The optimal sensor deployment strategy may change when the time duration for which traffic data measurements are available is limited.

2.2 Objectives

To address the aforementioned gaps, this paper proposes a NSLP model to identify the optimal heterogeneous sensor deployment strategy to maximize the quality of dynamic O-D demand estimates, or to minimize the variability of dynamic O-D demand estimates, under a limited budget. The sensor deployment strategy is in terms of the numbers of link (counting) and node (video/image) sensors and their installation locations. We assume that each time-dependent O-D demand is a random variable, and the variability is measured by the trace of the covariance

matrix of the posterior O-D demand estimates. In the proposed model, counting sensors are assumed to be located on links to measure link flows, while video/image sensors are assumed to be located at nodes to measure turning movements. We assume that the cost of a counting sensor is cheaper than that of a video/image sensor. To factor the time duration for which traffic data measurements are available, we add a time duration constraint to the proposed NSLP model that specifies the time duration for which traffic data measurements are available. To study the impact of time duration constraint on the optimal sensor deployment strategy, we compare the optimal sensor deployment strategy using traffic data measurements for the entire time period to the strategy using traffic data measurements for a shorter time period.

A sequential sensor location algorithm that avoids matrix inversions is introduced to solve the proposed NSLP model. Since the costs of link and node sensors are different, the optimal numbers of link and node sensors cannot be simultaneously identified using a simple approach. The proposed algorithm first assumes the number of node sensors in the network as given, and under this scenario selects the sensor deployment strategy under a budget constraint with the lowest variability in the dynamic O-D demand estimates by sequentially adding one sensor (including sensor type and location) at a time to avoid matrix inversions and simplify the computation. The process is repeated for other scenarios with different given number of node sensors in the network. It then compares the selected sensor deployment strategies for various scenarios with given number of node sensors, and selects the optimal sensor deployment strategy as the one with the lowest variability in the dynamic O-D demand estimates.

2.3 Methodology

2.3.1 *Relationships among variables considered in the NSLP model*

Consider the following flow conservation equation:

$$f_{i,k,t} = p_{i,k,t} d_{i,t} \quad (1)$$

where $d_{i,t}$ is the flow of O-D pair i with departure time t , $f_{i,k,t}$ is the number of users between O-D pair i choosing path k with departure time t , and $p_{i,k,t}$ is the proportion of users between O-D pair i with departure time t choosing path k .

Define matrices \mathbf{D} , \mathbf{F}_i , and \mathbf{F} as follows:

$$\mathbf{D} = [d_{1,1}, d_{1,2}, \dots, d_{1,t}, \dots, \underbrace{d_{i,1}, d_{i,2}, \dots, d_{i,t} \dots}_{\text{time-dependent demand of O-D pair } i}, \dots]^T \quad (2)$$

$$\mathbf{F}_i = [f_{i,1,1}, f_{i,1,2}, \dots, f_{i,1,t}, \dots, \underbrace{f_{i,k,1}, f_{i,k,2}, \dots, f_{i,k,t} \dots}_{\text{time-dependent flow of path } k \text{ from O-D pair } i}]^T \quad (3)$$

$$\mathbf{F} = [\mathbf{F}_1^T, \mathbf{F}_2^T, \dots, \mathbf{F}_i^T, \dots]^T \quad (4)$$

where \mathbf{D} is the vector of all considered time-dependent O-D demands, \mathbf{F}_i is the vector of all time-dependent path flows between O-D pair i and \mathbf{F} is the vector of all time-dependent path flows.

Define a $\mathbf{m} \times \mathbf{s}$ matrix $\mathbf{P}_{i,t}$, where \mathbf{m} is the total number of paths for O-D pair i with departure time t , and \mathbf{s} is the dimension of \mathbf{D} . The (k, j) th element $\mathbf{P}_{i,t}(k, j)$ of $\mathbf{P}_{i,t}$ is defined as:

$$\mathbf{P}_{i,t}(k, j) = \begin{cases} p_{i,k,t} & \text{if } j = (i - 1) * |\mathbf{T}| + t \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where $|\mathbf{T}|$ is the number of time intervals.

Define matrix \mathbf{P} as follows:

$$\mathbf{P} = [\mathbf{P}_{1,1}^T, \mathbf{P}_{1,2}^T, \dots, \mathbf{P}_{2,1}^T, \mathbf{P}_{2,2}^T, \dots, \mathbf{P}_{i,1}^T, \mathbf{P}_{i,2}^T, \dots]^T \quad (6)$$

Then, the path flows satisfy the following flow conservation condition:

$$\mathbf{F} = \mathbf{P}\mathbf{D} \quad (7)$$

The time-dependent path flow is the product of the path choice proportion and the time-dependent O-D demand. Note that the path choice proportion is a deterministic variable, which can be obtained by solving the dynamic user equilibrium (DUE) problem. Then, the time-dependent node turning movements and link flows can be derived from the path flows.

Define v_j as the flow on link j , and $\varphi_{i,k,t,j}$ as the time-dependent link-path incidence indicator. $\varphi_{i,k,t,j} = 1$, if link j belongs to path k of O-D pair i with departure time t , and $\varphi_{i,k,t,j} = 0$, otherwise.

The link flow can be derived from the time-dependent path flows as follow:

$$v_j = \sum_i \sum_k \sum_t \varphi_{i,k,t,j} f_{i,k,t} \quad (8)$$

Define matrices $\boldsymbol{\varphi}_{i,j}$, $\boldsymbol{\varphi}_j$, $\boldsymbol{\varphi}$ and \mathbf{V} as follows:

$$\boldsymbol{\varphi}_{i,j} = [\varphi_{i,1,1,j}, \varphi_{i,1,2,j}, \dots, \varphi_{i,k,1,j}, \varphi_{i,k,2,j} \dots] \quad (9)$$

$$\boldsymbol{\varphi}_j = [\boldsymbol{\varphi}_{1,j}, \boldsymbol{\varphi}_{2,j}, \dots, \boldsymbol{\varphi}_{i,j}, \dots] \quad (10)$$

$$\boldsymbol{\varphi} = [\boldsymbol{\varphi}_1^T, \boldsymbol{\varphi}_1^T, \dots, \boldsymbol{\varphi}_j^T \dots]^T \quad (11)$$

$$\mathbf{V} = [v_1, v_2, \dots, v_j, \dots]^T \quad (12)$$

where \mathbf{V} is the vector of all considered link flows and $\boldsymbol{\varphi}$ is the corresponding incidence indicator vector. $\boldsymbol{\varphi}_{i,j}$ is the vector of incidence indicators for each time-dependent path of O-D pair i that includes link j . $\boldsymbol{\varphi}_j$ is the vector of incidence indicators related to each time-dependent path of each O-D pair that includes link j .

By considering an error term $\boldsymbol{\varepsilon}$, the following linear relationship can be expressed between the path flow vector \mathbf{F} and the link flow vector \mathbf{V} :

$$\mathbf{V} = \boldsymbol{\varphi}\mathbf{F} + \boldsymbol{\varepsilon} = \boldsymbol{\varphi}\mathbf{P}\mathbf{D} + \boldsymbol{\varepsilon} \quad (13)$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots)$ are mutually independent random variables with zero mean.

Define $S_{j_{a,b},\tau}$ as the number of users traveling from upstream node a to downstream node b connected by node j with departure time τ , and $\psi_{i,k,t,j_{a,b}}$ is the incidence indicator (i.e., $\psi_{i,k,t,j_{a,b}} = 1$, if the sub-path made up of nodes a, j and b belongs to path k of O-D pair i with departure time t , and $\psi_{i,k,t,j_{a,b}} = 0$, otherwise). Here, $a \in \mathbf{N}_u$ and $b \in \mathbf{N}_d$, where \mathbf{N}_u is the set of upstream nodes of node j and \mathbf{N}_d is the set of downstream nodes of node j .

Define a column vector \mathbf{S}_j as the set of all the turning movements at node j and a row vector $\boldsymbol{\Psi}_{i,k,t,j}$ as the set of incidence indicators related to path k of O-D pair i with departure time t choosing each turning movement at node j . Then, vectors $\boldsymbol{\Psi}_{i,j}$, $\boldsymbol{\Psi}_j$, $\boldsymbol{\Psi}$ and \mathbf{S} are defined as follow:

$$\boldsymbol{\Psi}_{i,j} = [\psi_{i,j_{i,1,1,j}}, \psi_{i,j_{i,1,2,j}}, \dots, \psi_{i,j_{i,k,1,j}}, \psi_{i,j_{i,k,2,j}}, \dots, \psi_{i,j_{i,k,t,j}}] \quad (14)$$

$$\boldsymbol{\Psi}_j = [\boldsymbol{\Psi}_{1,j}, \boldsymbol{\Psi}_{2,j}, \dots, \boldsymbol{\Psi}_{i,j}, \dots] \quad (15)$$

$$\boldsymbol{\Psi} = [\boldsymbol{\Psi}_1^T, \boldsymbol{\Psi}_2^T, \dots, \boldsymbol{\Psi}_j^T, \dots]^T \quad (16)$$

$$\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_j, \dots]^T \quad (17)$$

where \mathbf{S} is the vector of all considered turning movements and $\boldsymbol{\Psi}$ is the corresponding incidence indicator vector. $\boldsymbol{\Psi}_{i,j}$ is the vector of incidence indicators related to each time-dependent path of O-D pair i choosing each turning movement at node j . $\boldsymbol{\Psi}_j$ is the vector of

incidence indicators related to each time-dependent path of each O-D pair choosing each turning movement at node j .

By considering an error term $\boldsymbol{\eta}$, the following linear relationship can be expressed between the path flow vector \mathbf{F} and the turning movement vector \mathbf{S} :

$$\mathbf{S} = \boldsymbol{\Psi}\mathbf{F} + \boldsymbol{\eta} = \boldsymbol{\Psi}\mathbf{P}\mathbf{D} + \boldsymbol{\eta} \quad (18)$$

where $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots)$ are mutually independent random variables with zero mean.

According to Eqs. (7), (13) and (18), the random variables in the proposed model can be expressed through the following linear relationships:

$$\begin{pmatrix} \mathbf{D} \\ \mathbf{F} \\ \mathbf{V} \\ \mathbf{S} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{P} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Phi}\mathbf{P} & \mathbf{I} & \mathbf{0} \\ \boldsymbol{\Psi}\mathbf{P} & \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{D} \\ \boldsymbol{\varepsilon} \\ \boldsymbol{\eta} \end{pmatrix} \quad (19)$$

In this study, since all of the considered variables are treated as random variables, the traffic demands between all time-dependent O-D pairs are assumed to follow multivariate normal (MVN) distributions. This is because these random variables are the outcome of a large number of independent Bernoulli experiments in which the users decide where to travel and which routes to choose. This assumption is similar to assumptions made in past studies (33-34).

Specifically, \mathbf{D} are multivariate normal random variables with mean $E(\mathbf{D})$ and variance $\boldsymbol{\Sigma}_{\mathbf{D}}$, $\boldsymbol{\varepsilon}$ are mutually independent normal random variables with mean $E(\boldsymbol{\varepsilon})$ and variance $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$, and $\boldsymbol{\eta}$ are mutually independent normal random variables with mean $E(\boldsymbol{\eta})$ and variance $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$. Based on Eq. (19), the prior variance-covariance matrix $\boldsymbol{\Sigma}_{(\mathbf{D},\mathbf{F},\mathbf{V},\mathbf{S})}$ is:

$$\boldsymbol{\Sigma}_{(D,F,V,S)} = \begin{pmatrix} \boldsymbol{\Sigma}_D & \boldsymbol{\Sigma}_D \mathbf{P}^T & \boldsymbol{\Sigma}_D (\boldsymbol{\varphi} \mathbf{P})^T & \boldsymbol{\Sigma}_I \\ \mathbf{P} \boldsymbol{\Sigma}_D & \mathbf{P} \boldsymbol{\Sigma}_D \mathbf{P}^T & \mathbf{P} \boldsymbol{\Sigma}_D (\boldsymbol{\varphi} \mathbf{P})^T & \mathbf{P} \boldsymbol{\Sigma}_I \\ \boldsymbol{\varphi} \mathbf{P} \boldsymbol{\Sigma}_D & \boldsymbol{\varphi} \mathbf{P} \boldsymbol{\Sigma}_D \mathbf{P}^T & \boldsymbol{\varphi} \mathbf{P} \boldsymbol{\Sigma}_D (\boldsymbol{\varphi} \mathbf{P})^T + \boldsymbol{\Sigma}_\varepsilon & \boldsymbol{\varphi} \mathbf{P} \boldsymbol{\Sigma}_I \\ \boldsymbol{\Psi} \mathbf{P} \boldsymbol{\Sigma}_D & \boldsymbol{\Psi} \mathbf{P} \boldsymbol{\Sigma}_D \mathbf{P}^T & \boldsymbol{\Psi} \mathbf{P} \boldsymbol{\Sigma}_D (\boldsymbol{\varphi} \mathbf{P})^T & \boldsymbol{\Psi} \mathbf{P} \boldsymbol{\Sigma}_I \end{pmatrix} \quad (20)$$

Based on the variance-covariance matrix, the correlations among the considered variables can be derived, and these correlations can be used to derive the NSLP model as shown in the next section.

2.3.2 Formulation of the NSLP models

The prior distribution $\phi_{\mathbf{D}}$ (i.e., $E(\mathbf{D})$ and $\boldsymbol{\Sigma}_D$ in Eq. (20)) of time-dependent O-D demand is assumed to be obtained using historical O-D data. From Eq. (20), we can derive the prior distribution of all variables. In turn, let \mathbf{Z} be a sensor deployment strategy whose counted flows are known to be $\mathbf{Z} = \mathbf{z}$ (including observed link flows and node turning movements). Then, $\phi_{\mathbf{D}|\mathbf{Z}=\mathbf{z}}$, obtained by updating the counted flows, is the posterior distribution of the time-dependent O-D demand.

In this study, the trace $Tr(\boldsymbol{\Sigma}_{\mathbf{X}})$ of the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$ is adopted as a measure of variability related to the random vector \mathbf{X} . Therefore, we adopt the trace $Tr(\boldsymbol{\Sigma}_{\mathbf{D}|\mathbf{Z}=\mathbf{z}})$ to represent the variability of the time-dependent O-D demand conditional on the counted flows $\mathbf{Z} = \mathbf{z}$. Since $Tr(\boldsymbol{\Sigma}_{\mathbf{D}|\mathbf{Z}=\mathbf{z}})$ generally depends on the counted flows $\mathbf{Z} = \mathbf{z}$, the variability of the posterior random vector $\mathbf{D}|\mathbf{Z} = \mathbf{z}$ can be defined as the average of $Tr(\boldsymbol{\Sigma}_{\mathbf{D}|\mathbf{Z}=\mathbf{z}})$, as follows:

$$E[Tr(\boldsymbol{\Sigma}_{\mathbf{D}|\mathbf{Z}=\mathbf{z}})] = \int_{\Omega_{\mathbf{Z}}} Tr(\boldsymbol{\Sigma}_{\mathbf{D}|\mathbf{Z}=\mathbf{z}}) \phi_{\mathbf{Z}} \partial \mathbf{z} \quad (21)$$

where $\phi_{\mathbf{Z}}$ is the density function of the random variable \mathbf{Z} and $\Omega_{\mathbf{Z}}$ is its domain.

The NSLP can be formulated as the problem of finding the optimal sensor deployment strategy \mathbf{Z}^* , which minimizes the variability of the posterior random vector $\mathbf{D}|\mathbf{Z} = \mathbf{z}$ in the domain $\Omega_{\mathbf{z}}$, subject to budget constraints. Then, the NSLP model for dynamic O-D demand estimation is proposed as follow:

$$\min \int_{\Omega_{\mathbf{z}}} Tr(\Sigma_{\mathbf{D}|\mathbf{Z}=\mathbf{z}}) \phi_{\mathbf{z}} \partial \mathbf{z} \quad (22)$$

$$\text{s.t. } b_v L + b_s S \leq b_{max} \quad (23)$$

where b_v is the cost of a link sensor, b_s is the cost of a node sensor, b_{max} is the overall available budget, L is the cardinality of the identified link set, *and* S is the cardinality of the identified node set.

By optimizing (22) and (23), the resultant optimal heterogeneous sensor deployment strategy depends on the variability of both prior and posterior O-D demand estimates. The posterior estimates are related to the relationships among the considered variables, which depend on the network topology and the users' travel behaviors. In summary, the optimal heterogeneous sensor deployment strategy obtained by solving the proposed NSLP model incorporates the variability of prior O-D demand estimates, the network topology, and the users' travel behaviors.

As we have obtained the prior means and variance-covariance matrix of all the variables as shown in Eq. (20), the mean and the covariance matrix of the variables can be updated based on some observed variables using the following equations (Maher 1983) under the assumption of normal distribution:

$$\Sigma_{\mathbf{XY}|\mathbf{Z}=\mathbf{z}} = \Sigma_{\mathbf{XY}} - \Sigma_{\mathbf{XZ}} \Sigma_{\mathbf{ZZ}}^{-1} \Sigma_{\mathbf{YZ}} \quad (24)$$

where \mathbf{X} and \mathbf{Y} both refer to the components of $(\mathbf{D}, \mathbf{F}, \mathbf{V}, \mathbf{S})$; Σ_{ZZ} is the covariance matrix of the observation \mathbf{Z} ; Σ_{XZ} is the covariance matrix of \mathbf{X} and \mathbf{Z} ; Σ_{YZ} is the covariance matrix of \mathbf{Y} and \mathbf{Z} ; Σ_{XY} is the covariance matrix of \mathbf{Y} and \mathbf{X} ; and $\Sigma_{XY|Z=z}$ is the posterior covariance matrix of \mathbf{Y} and \mathbf{X} .

Under the assumption of multivariate normal distribution, it can be shown that the conditional variance $\Sigma_{XY|Z=z}$ does not depend on the specific counted values of \mathbf{Z} (i.e. \mathbf{z}). Therefore, the trace $Tr(\Sigma_{D|Z=z})$ does not depend on the actual values of the traffic data measurements, but just on the random sensor deployment strategy \mathbf{Z} .

Hence, the optimization problem (22) and (23) can be rewritten as:

$$\min \int_{\Omega_Z} Tr(\Sigma_{D|Z=z}) \phi_Z \partial z = Tr(\Sigma_D - \Sigma_{DZ} \Sigma_{ZZ}^{-1} \Sigma_{DZ}) \quad (25)$$

$$\text{s.t. } b_v L + b_s N \leq b_{max} \quad (26)$$

where Σ_{DZ} and Σ_{ZZ} are derived from Eq. (20).

In real-world applications, the time duration for which traffic data measurements are available is usually limited to a shorter time period compared to the entire time period of interest related to determining the dynamic O-D demand. To factor this, the proposed model includes a time duration constraint:

$$\min \int_{\Omega_Z} Tr(\Sigma_{D|Z=z}) \phi_Z \partial z = Tr(\Sigma_D - \Sigma_{DZ} \Sigma_{ZZ}^{-1} \Sigma_{DZ}) \quad (27)$$

$$\text{s.t. } b_v L + b_s N \leq b_{max} \quad (28)$$

$$\tau_i \in [\tau_{min}, \tau_{max}], \forall i \in \mathbf{Z} \quad (29)$$

where τ_i is the departure time of user at sensor location i ; τ_{min} is the start time of the considered time duration for which traffic data measurements are available and τ_{max} is the end time.

Due to the time duration constraint, incidence indicators $\varphi_{i,k,t,j}$ and $\psi_{i,k,t,j,a,b}$ are not necessarily 0 or 1 for deriving Σ_{DZ} and Σ_{ZZ} using Eq. (20). In this case, $\varphi_{i,k,t,j}$ is equal to the proportion of users of path k of O-D pair i with departure time t choosing link j with departure time τ_j which is in the time period $[\tau_{min}, \tau_{max}]$. $\psi_{i,k,t,j,a,b}$ is equal to the proportion of users of path k of O-D pair i with departure time t traveling from upstream node a to downstream node b connected by node j with departure time τ_j included in the time period $[\tau_{min}, \tau_{max}]$. Given the prior O-D demand, these proportions can be obtained by solving the DUE problem.

We denote the model without the time duration constraint as NSLP-NT model, and the model with the time duration constraint as NSLP-T model.

2.3.3 The sequential sensor location algorithm

As the costs for a link sensor and a node sensor are different, the optimal numbers of link sensors and node sensors cannot be determined simultaneously in a simple manner. Therefore, we first specify a given number of node (or link) sensors in the network, and select the sensor deployment strategy with the lowest variability in the dynamic O-D demand estimates. Then, we compare the selected sensor location strategy for different given number of node (or link) sensors, and choose the optimal strategy as the one with the lowest variability.

To solve the aforementioned NSLP models, calculating the inverse of Σ_{ZZ} (i.e., Σ_{ZZ}^{-1}) requires a large amount of computational effort, especially for large-scale networks since the dimension of Σ_{ZZ} is usually very large. Interestingly, if we sequentially update one observed

variable at a time in Eq. (24), it does not involve matrix inverse calculation, because in such a case, Σ_{DZ} is a column vector and Σ_{ZZ} is a scalar (i.e., $\Sigma_{ZZ} = \sigma_{ZZ}$). Hence, the proposed NSLP-NT and NSLP-T models can be solved using a sequential sensor location algorithm, summarized using the following eight steps:

Step 0: Initialization: Calculate the choice proportion matrices \mathbf{P} , Φ and Ψ by solving the DUE problem.

Step 1: Calculate the prior variance and covariance of all considered variables according to Eq. (20).

Step 2: Define the maximum number of node sensors n^{max} based on the budget constraint b_{max} , i.e., $n^{max} = \text{int}(b_{max}/b_s)$. Define m as the number of node sensors, whose initial value is 0. Define λ^{min} as the minimum value of the objective function, whose initial value is equal to the trace of the prior covariance matrix of the dynamic O-D demand, that is, $\lambda^{min} = \text{tr}(\Sigma_D)$. Define $\tilde{\mathbf{Z}}$ as the final (chosen) optimal heterogeneous sensor deployment strategy and \mathbf{Z} as the current optimal heterogeneous sensor deployment strategy; the initial values of $\tilde{\mathbf{Z}}$ and \mathbf{Z} are both set to be null sets.

Step 3: Identify the maximum number of link sensors l^{max} ($l^{max} = \text{int}[(b_{max} - b_s m)/b_v]$). Define l as the number of the identified link sensor locations, n as the number of the identified node sensor locations and set the initial values of l and n to be 0. Define μ as the current value of the objective function, whose initial value is $\mu = \text{tr}(\Sigma_D)$.

Step 4: Identify one sensor location and add it to the sensor deployment strategy \mathbf{Z} .

Define link or node Z^* as the next sensor location which minimizes the objective function, that is:

$$Z^* = \mathbf{arg\ min} Tr(\Sigma_{\mathbf{D}} - \Sigma_{\mathbf{DZ}}\Sigma_{\mathbf{ZZ}}^{-1}\Sigma_{\mathbf{DZ}}) \quad (30)$$

In this step, if $n < m$ and $l < l^{max}$, the new identified sensor can be a link sensor or a node sensor; if $n = m$ and $l < l^{max}$, the new identified sensor can only be a link sensor; and if $n < m$ and $l = l^{max}$, the new identified sensor can only be a node sensor.

Step 5: Update the variance, covariance and the current value μ of the objective function. According to Eq. (24), the variance and covariance of traffic flows can be updated using the following formula:

$$\Sigma_{\mathbf{XY}|Z^*} = \Sigma_{\mathbf{XY}} - \Sigma_{\mathbf{XZ}^*}\Sigma_{\mathbf{YZ}^*}/\sigma_{Z^*Z^*} \quad (31)$$

where \mathbf{X} and \mathbf{Y} both refer to the components of $(\mathbf{D}, \mathbf{F}, \mathbf{V}, \mathbf{S})$, Z^* is the new identified sensor location by Step 4, and $\sigma_{Z^*Z^*}$, which is a scalar, is the variance of Z^* .

Step 6 : If the new identified sensor location is a link sensor location, set $l = l + 1$; otherwise set $n = n + 1$. If $l = l^{max}$ and $n = m$, continue with Step 7. Otherwise, go to Step 4.

Step 7: Specify the optimal heterogeneous sensor deployment strategy \mathbf{Z} under the condition in which the defined number of node sensor locations is m . If $\mu < \lambda^{min}$, set $\lambda^{min} = \mu$, and $\tilde{\mathbf{Z}} = \mathbf{Z}$.

Step 8: If $m < n^{max}$, set $m = m + 1$ and initialize the variance and covariance of all variables (using the values based on Step 1); set \mathbf{Z} as a null set and go to Step 3. Otherwise, stop the algorithmic process and specify the optimal heterogeneous sensor deployment strategy $\tilde{\mathbf{Z}}$.

$\tilde{\mathbf{Z}}$ In this algorithm, given the number of node sensors (in Step 2), Steps 3-7 identify one sensor location at a time and then update the identified sensor locations. In Step 4, only one sensor location is identified. In this case, because Σ_{DZ} is a column vector and Σ_{ZZ} is a scalar, if the total number of considered links and turning movements in the network is a and the number of identified sensor locations is c , the number of calculations needed in this step is linearly related to the number of the remaining links and turning movements, i.e., $a - c$. Since most of the calculations in this algorithm are involved in Step 4, the computational time of solving the NSLP model is linear with respect to the number of links and turning movements in the network.

2.4 Results and discussions

We apply the proposed NSLP models and algorithm to the test network shown in Figure 2.1, which includes 11 traffic analysis zones (100 O-D pairs), 70 nodes, and 141 directed links. The O-D matrix is time-dependent, and specified in 15-minute intervals. The entire analysis period is two hours (from 8:00 a.m. to 10:00 a.m.); hence the time-dependent O-D matrix has 800 records (8 time intervals for 100 O-D pairs). The prior O-D demand is established from the historical data available for this network. The total available budget is assumed to be 300. The costs for a node sensor and link sensor are assumed to be 50 and 15, respectively. The numerical experiments were conducted using DYNASMART-P 1.3.0, and the prior traffic data

measurements for all links and nodes are obtained by solving the dynamic user equilibrium problem based on the prior O-D demand.

Figure 2.2 shows the plots of the traces of the covariance matrices of the dynamic O-D demand estimation for the NSLP-NT model using the proposed sequential algorithm under different given number of node sensors (labeled NS# in the figure). Figure 8 shows the corresponding plots for the NSLP-T model, in which the time duration for which traffic data measurements are available is set to be one hour, from 9:00 a.m. to 10:00 a.m. As shown in Figures 7 and 8, the traces decrease with each added sensor. More sensors imply that more observed information can be collected to update the variance-covariance matrix, which can reduce the variability of the dynamic O-D demand estimates. The traces decrease more rapidly when a node sensor is added rather than a link sensor, because a node sensor can detect several turning movements (for example, 12 turning movements at a traditional four-way intersection) and thus provide more updated information compared to a link sensor. Hence, as the NS# increases, the traces decrease more rapidly as each node sensor is added.

Table 2.1 shows the optimal sensor deployment strategies for the NSLP-NT model and Table 2.2 shows the optimal strategies for the NSLP-T model. In terms of the notation in Tables 2.1 and 2.2, for example, 7(n) represents a node sensor located at node 7; 16-17 represents a link sensor on link 16-17 with upstream node 16 and downstream node 17.

As illustrated by Tables 2.1 and 2.2, under the budget constraint, though a node sensor can provide more updated information compared to a link sensor, more node sensors cannot reduce the variability of the O-D demand estimates. Thereby, the number of node sensors in the optimal heterogeneous sensor deployment strategy for the NSLP-NT model is 3, though the maximum number of node sensors is 6 because it leads to the lowest objective function value

(bolded in the table). When there are 6 node sensors in the network, the allowed number of link sensors is 0. However, in the optimal heterogeneous sensor deployment strategy, the number of link sensors is 8, though the number of node sensors is only 3. This is because a link sensor is much cheaper than a node sensor, and more node sensors can imply fewer link sensors, leading to a tradeoff in terms of the desirable numbers of each sensor type. The number of node sensors in the optimal heterogeneous sensor deployment strategy for the NSLP-T model is 5. This higher value compared to the NSLP-NT case is because the more constrained context due to the time duration constraint under NSLP-T favors the use of more node sensors to elicit more information. In summary, the optimal numbers for both the link sensors and node sensors can be determined using the proposed algorithm.

In addition, from Tables 2.1 and 2.2, the optimal heterogeneous sensor deployment strategy obtained for the NSLP-T model is different from that of the NSLP-NT model. This implies that considering the time duration for which traffic data measurements are available can lead to a different optimal sensor deployment strategy. The objective function value of the NSLP-NT model is lower than that of the NSLP-T model as it is a less constrained problem. This implies that the posterior variance of variables in the NSLP-T model is larger than that in the NSLP-NT model. This is because in the NSLP-NT model the update information is collected for the entire time period of interest related to determining the dynamic O-D demands, while the update information in the NSLP-T model is collected for only the time duration specified. Hence, the NSLP-T model has less observed information to update the posterior variance which leads to a larger variability in the dynamic O-D demand estimates as more observed information implies lower variability of the O-D demand estimates (as illustrated in Figures 7 and 8).

To further compare the NSLP-T and the NSLP-NT models, we use the optimal sensor deployment strategy from the NSLP-NT model to collect traffic data measurements, but now with the time duration constraint used for the NSLP-T model. Figure 9 shows how the traces of the covariance matrices of the dynamic O-D demand estimates of the two models change after updating the identified sensor location one-at-a-time using the proposed sequential algorithm. Note that compared to the NSLP-NT model, the trace of the covariance matrix of the dynamic O-D demand estimation under the NSLP-T model decreases more rapidly and the optimal value is lower (the traces in NSLP-T and NSLP-NT models are 24308.59 and 25123.47, respectively). This demonstrates the superior accuracy and performance of the NSLP-T model. That is, the NSLP-T model optimal sensor deployment strategy performs better than the NSLP-NT model optimal sensor deployment strategy with the consideration of the same time duration constraint used in the NSLP-T model.

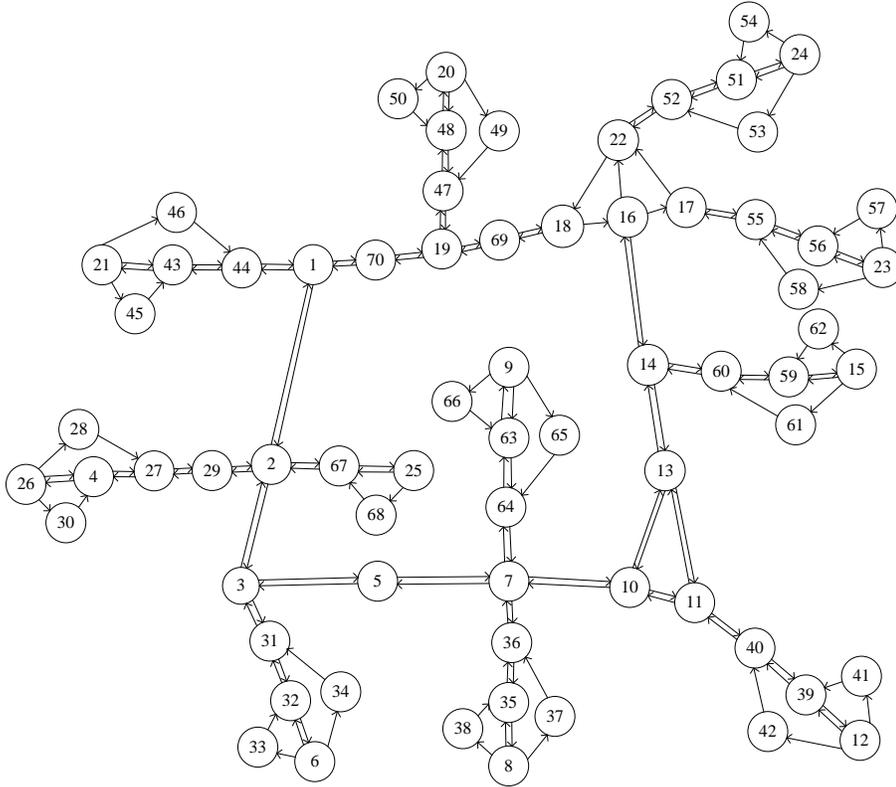


Figure 2.1 Test network

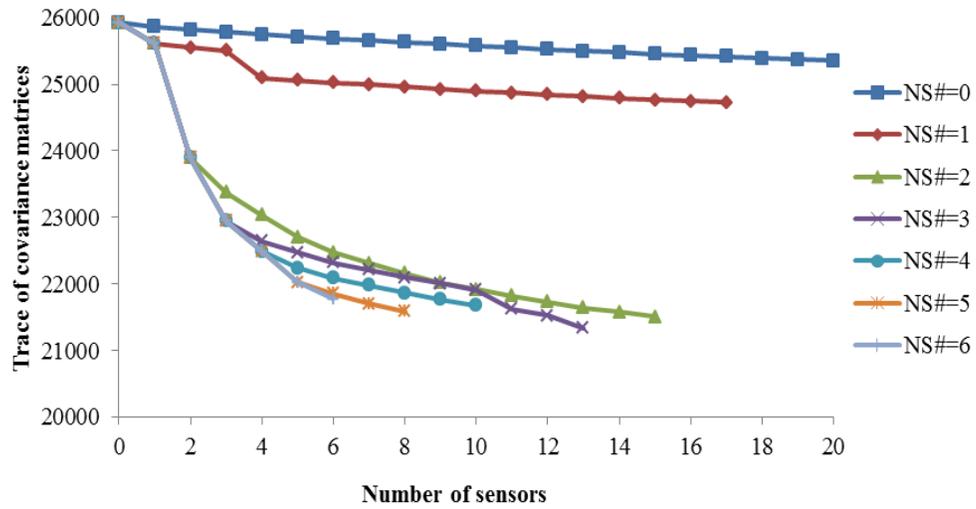


Figure 2.2 Traces of covariance matrices of dynamic O-D demand estimation after updating each identified sensor location with different given number of node sensors, without considering the time duration constraint.

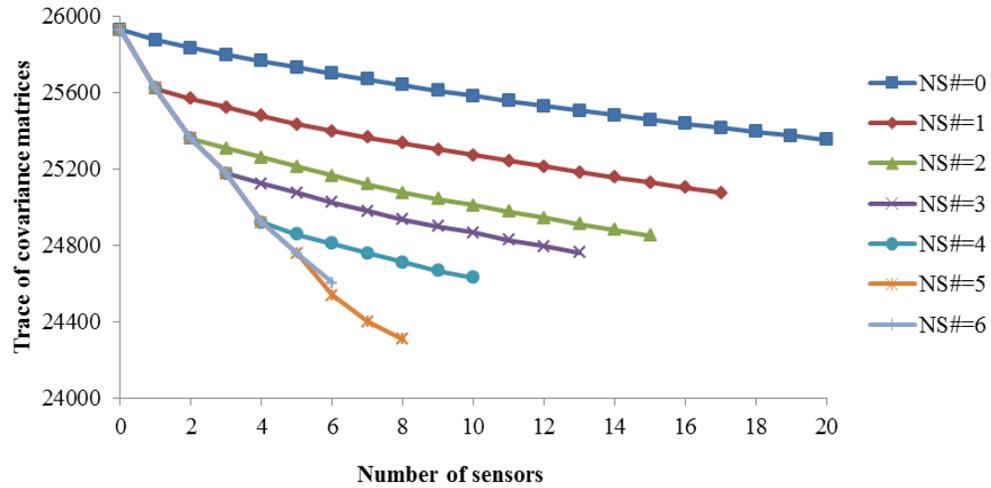


Figure 2.3 Traces of covariance matrices of dynamic O-D demand estimation after updating each identified sensor location with different given number of node sensors, considering the time duration constraint.

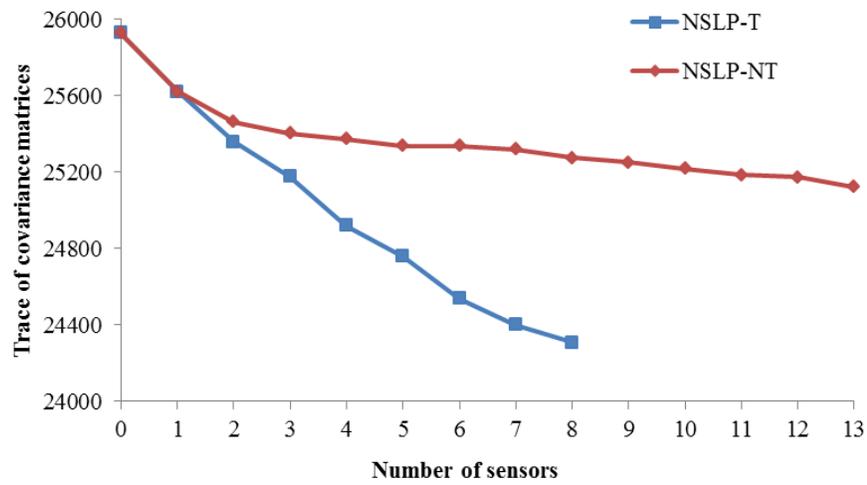


Figure 2.4 Traces of covariance matrices of the dynamic O-D demand estimates for the two models after updating the sensor locations under the same time duration constraint.

Table 2.1 Sensor Deployment Strategies for the NSLP-NT Model

Number of node sensors	Sensor deployment strategy	Objective function value
0	16-17, 22-18, 13-11, 31-3, 3-31, 11-13, 64-7, 29-2, 44-1, 47-19, 10-13, 67-2, 19-47, 7-64, 36-7, 7-36, 1-44, 60-14, 14-60, 10-11	25351.54
1	7(n), 16-17, 10-7, 7-36, 13-11, 22-18, 31-3, 10-11, 3-31, 11-13, 44-1, 47-19, 29-2, 10-13, 67-2, 19-47, 14-60	24724.28
2	7(n), 3(n), 1-2, 16-14, 19-1, 22-18, 18-16, 3-31, 11-13, 14-16, 31-3, 7-36, 18-19, 14-13, 10-11	21507.6
3	7(n), 3(n), 1(n), 15, 12, 3-31, 5-3, 7-36, 67-2, 19-18, 1-44, 19-1, 16-17	21338.33
4	7(n), 3(n), 1(n), 16(n), 18-16, 3-31, 5-3, 7-36, 22-18, 19-1	21682.16
5	7(n), 3(n), 1(n), 16(n), 2(n), 5-3, 3-31, 7-36	21591.57
6	7(n), 3(n), 1(n), 16(n), 2(n), 10(n)	21777.34

Table 2.2 Model Sensor Deployment Strategies for the NSLP-T Model

Number of node sensors	Sensor deployment strategy	Objective function value
0	16-17, 13-11, 31-3, 11-13, 22-18, 3-31, 1-44, 64-7, 44-1, 10-11, 19-47, 10-13, 29-2, 36-7, 7-36, 67-2, 60-14, 14-60, 47-19, 2-29	25352.05
1	7(n), 16-17, 64-7, 7-36, 13-11, 19-47, 1-44, 11-13, 36-7, 3-31, 60-14, 7-10, 5-7, 31-3, 44-1, 10-13, 22-18	25074.89
2	7(n), 2(n), 16-17, 2-67, 2-29, 64-7, 7-36, 13-11, 1-2, 67-2, 29-2, 11-13, 19-47, 36-7, 60-14	24851.41
3	7(n), 2(n), 14(n), 16-17, 2-67, 2-29, 64-7, 14-60, 11-13, 1-2, 60-14, 67-2, 22-18	24761.18
4	7(n), 2(n), 14(n), 16(n), 14-16, 2-67, 2-29, 64-7, 14-60, 16-17	24630.79
5	7(n), 2(n), 14(n), 16(n), 1(n), 44-1, 60-14, 1-2	24308.59
6	7(n), 2(n), 14(n), 16(n), 1(n), 10(n)	24603.31

CHAPTER 3. CONCLUSIONS

To estimate dynamic O-D demand, a Bayesian method is proposed. The method can synthesize multiple sources of data, including link counts, turning movements at intersections, sub-path flows and/or sub-path travel time. Demand between each O-D pair at each departure time is assumed to satisfy normal distribution. Results shows that the total variability of O-D demand decreases with each added traffic count. More traffic counts can lead to smaller variance of the dynamic O-D demand, which means updating each traffic count can reduce the uncertainty in the O-D estimation. Using the proposed algorithm, the total deviations between estimated and observed traffic counts is decreasing at each iteration, as supported by three measures of performance. After a few iterations, the three measures all decrease to small values and become flat, which implies a well fit between estimated and observed traffic counts. Moreover, the source-specific deviations between estimated and observed traffic counts are small too. This demonstrates the proposed Bayesian method can synthesize multiple sources of data well. It also implies that more traffic counts lead to lower uncertainty of the O-D demand estimates, resulting a better accurate estimation.

To study the sub-problem of dynamic O-D estimation problem, this research proposes a network sensor location problem model to determine the optimal heterogeneous sensor deployment strategy for the dynamic O-D demand estimation problem. By maximizing the quality or minimizing the variability of the O-D demand estimates under a given budget constraint, the proposed model can be used to determine the optimal link (counting) and node (video/image) sensors numbers and their installation locations. In the proposed model, the trace of the covariance matrix of the posterior O-D demand estimates is adopted as a measure of the variability of the O-D demand estimates. The time duration for which traffic data measurements

are available is constrained in the proposed model to factor that traffic data are usually collected for a shorter time period in practice rather than the entire time period of interest (such as the peak period) in the context of determining the optimal sensor deployment strategy. Results show that node sensors normally can be identified preferentially than links sensors and a tradeoff in terms of the desirable numbers of each sensor type can be obtained by the proposed algorithm. Results also illustrates that the optimal sensor deployment strategy can change significantly under constraint of time duration for which traffic data measurements are available.

REFERENCES

- Alibabai H, Mahmassani H (2009) Dynamic Origin-Destination Demand Estimation Using Turning Movement Counts. *Transportation Research Record: Journal of the Transportation Research Board* 2085:39 – 48.
- Antoniou C, Ben-Akiva M, Koutsopoulos H (2004) Incorporating Automated Vehicle Identification Data Into Origin-Destination Estimation. In: *Proceeding of the 83rd Annual Meeting in Transportation Research Board, TRB 2004*.
- Asakura Y., Hato E, Kashiwadani M (2000). Origin-Destination Matrices Estimation Model Using Automatic Vehicle Identification Data and its Application to the Han-Shin Expressway Network. *Transportation*, 27(4): 419-438.
- Ashok K (1996) Estimation and prediction of time-dependent origin-destination flows. Dissertation, Massachusetts Institute of Technology.
- Ashok K, Ben-Akiva ME (1993) Dynamic O-D matrix estimation and prediction for real-time traffic management systems. In: Daganzo CF (ed) *Transportation and traffic theory*. Elsevier, New York, pp 465 – 484
- Balakrishna R (2006) Off-line Calibration of dynamic Traffic Assignment Models. Dissertation, Massachusetts Institute of Technology.
- Balakrishna R, Ben-Akiva ME, Koutsopoulos HN (2008) Time-dependent origin-destination estimation without assignment matrices. In: Chung E, Dumont AG (ed) *Transport Simulation: Beyond Traditional Approaches*. EPFL Press, pp. 201 – 213.
- Bell MGH (1991) The estimation of origin – destination matrices by constrained generalized least squares. *Transportation Research Part B* 25 (1):13 – 23.
- Bianco L, Cerrone C, Cerulli R, Gentili M (2014). Locating Sensors to Observe Network Arc Flows: Exact and Heuristic Approaches. *Computers & Operations Research*, 46:12-22.

- Bierlaire M (2002). The Total Demand Scale: A New Measure of Quality for Static and Dynamic Origin-Destination Trip Table. *Transportation Research Part B*, 36(9): 837-850.
- Bierlaire M, Crittin F (2004) An efficient algorithm for real-time estimation and prediction of dynamic OD tables. *Operations Research* 52 (1):116–127.
- Caggiani L, Ottomanelli M, Sassanelli D (2013) A fixed point approach to origin – destination matrices estimation using uncertain data and fuzzy programming on congested networks. *Transportation Research Part C* 28:130 – 141.
- Cascetta E (1984). Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimation. *Transportation Research Part B* 18 (4 – 5):289 – 299.
- Cascetta E, Nguyen S (1988) A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B* 22(6):437 – 455.
- Cascetta E, Inaudi D, Marquis G (1993) Dynamic estimators of origin–destination matrices using traffic counts. *Transportation Science* 27 (4):363–373.
- Castillo E, Jiménez P, Menéndez JM, Conejo AJ (2008a). The Observability Problem in Traffic Models: Algebraic and Topological Methods. *IEEE Transactions on Intelligent Transportation Systems*, 9(2): 275-287.
- Castillo E, Menéndez JM, Sánchez-Cambronero S (2008b). Traffic estimation and optimal counting location without path enumeration using Bayesian networks. *Computer Aided Civil and Infrastructure Engineering* 23(3):189-207.
- Castillo E, Menéndez JM, Sánchez-Cambronero S (2008c) Predicting traffic flow using Bayesian networks. *Transportation Research Part B* 42(5):482-509.
- Castillo E, Jiménez P, Menéndez JM, Nogal M (2013) A Bayesian method for estimating traffic flows based on plate scanning. *Transportation* 40(1): 173-201.
- Chang GL, Wu J (1994) Recursive estimation of time-varying origin-destination flows from traffic counts in freeway corridors. *Transportation Research Part B* 28 (2):141–160.

- Chen A, Chootinan P, Pravinongvuth S (2004). Multiobjective Model for Locating Automatic Vehicle Identification Readers. *Transportation Research Record: Journal of the Transportation Research Board*, 1886: 49-58.
- Chen A, Chootinan P, Ryu S, Wong SC (2012). Quality Measures for Origin-Destination Estimation from Traffic Counts: A Review and a Generalized Demand Scale Measure. *Journal of Transportation Engineering*, 138(11):1340-1349.
- Chen A, Pravinongvuth S, Chootinan P (2010). Scenario-based Multiobjective AVI Reader Location Models under Different Travel Demand Patterns. *Transportmetrica*, 6(1): 53-78.
- Chen A, Pravinongvuth S, Chootinan P, Lee M, Recker W (2007). Strategies for Selecting Additional Traffic Counts for Improving O-D Trip Table Estimation. *Transportmetrica*, 3(3): 191-211.
- Cheng L, Zhu S, Chu Z, Chen J (2014). A Bayesian Network Model for Origin-Destination Matrices Estimation Using Prior and Some Observed Link Flows. *Discrete Dynamics in Nature and Society*, 2014.
- Cho HJ, Jou YJ, Lan CL (2009) Time dependent origin-destination estimation from traffic count without prior information. *Netw Spatial Econ* 9:145–170.
- Cipriani E, Florian M, Mahut M, Nigro M (2011) A gradient approximation approach for adjusting temporal origin – destination matrices. *Transportation Research Part C* 19:270 – 282.
- Dixon MP (2000) Incorporation of automatic vehicle identification data into synthetic OD estimation process. Dissertation, The University of Texas at Austin.
- Dixon MP, Rilett LR (2002) Real-time OD estimation using automatic vehicle identification and traffic count data. *Comput.-Aided Civil Infrastructure Engineering* 17(1):7–21.
- Doblas J, Benitez FG (2005) An approach to estimating and updating origin-destination matrices based upon traffic counts preserving the prior structure of a survey matrix. *Transportation Research Part B* 39 (7):565 – 591.

- Eisenman SM, Fei X, Zhou X, Mahmassani HS (2006). Number and Location of Sensors for Real-Time Network Traffic Estimation and Prediction: Sensitivity Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 1964: 253-259.
- Fei X, Mahmassani HS, Eisenman SM. Sensor Coverage and Location for Real-Time Traffic Prediction in Large-Scale Networks. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
- Fisk C (1988) On combining maximum trip estimation with user optimal assignment. *Transportation Research Part B* 22 (1):69–79.
- Ghali MO, Smith MJ (1995) A model for the dynamic system optimum traffic assignment problem. *Transportation Research Part B* 29 (3):155 – 170.
- Guo, Y., X. He, S. Peeta, J. Weiss, Internal Curing for Concrete Bridge Decks: Integrating a Social Cost Analysis in Evaluating the Long-Term Benefit, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2577, 2016, pp. 17-24.
- Guo, Y., and Peeta, S. Rail-Truck Multimodal Freight Collaboration: Truck Freight Carrier Perspectives in the United States. *Journal of Transportation Engineering*, Vol. 141, No. 11, 2015, 1-11.
- Guo, Y., Peeta, S., and Mannering, F. L. Rail-truck multimodal freight collaboration: a statistical analysis of freight-shipper perspectives. *Transportation Planning and Technology*, Vol. 39, No. 5, 2016a, 484-506.
- Guo, Y., Peeta, S., and Somenahalli, S. The impact of walkable environment on single family residential property values. *Journal of Transport and Land Use*, Vol. 10, No. 2, 2015, DOI: <http://dx.doi.org/10.5198/jtlu.2017.824>.
- Guo, Y., Agrawal, S., Peeta, S., and Somenahalli, S. Impacts of property accessibility and neighborhood built environment on single-unit and multiunit residential property values. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2568, 2016b, 103-112.

- Guo, Y., Wang, J., and Peeta, S. The impacts of household registration and head of household travel behavior on travel mode choice in China. In compendium of Transportation Research Board 95th Annual Meeting, Washington D.C., U.S., 2016c.
- Hazelton ML (2001) Inference for origin – destination matrices: estimation, prediction and reconstruction. *Transportation Research Part B* 35 (7):667 – 676.
- Hazelton ML (2010) Statistical inference for transit system origin-destination Matrices. *Technometrics* 52(2):221–230.
- He S (2013). A Graphical Approach to Identify Sensor Locations for Link Flow Inference. *Transportation Research Part B*, 51: 65-76.
- Hu SR, Peeta S, Chu CH (2009). Identification of Vehicle Sensor Locations for Link-based Network Traffic Applications. *Transportation Research Part B*, 43(8-9):873-894.
- Hu SR, Peeta S, Liou H-T (2015) Integrated Determination of Network Origin-Destination Trip Matrix and Heterogeneous Sensor Selection and Location Strategy. *IEEE Transactions on Intelligent Transportation Systems*, 17(1): 195-205.
- Huang S, Sadek AW, Guo L (2012) A computational-based approach to estimating travel demand in large-scale microscopic traffic simulation models. *ASCE Journal of Computing in Civil Engineering*, 27(1):78–86.
- Iqbal MS, Choudhury CF, Wang P, González MC (2014) Development of origin – destination matrices using mobile phone call data. *Transportation Research Part C* 40:63 – 74.
- Kattan L, Abdulhai B (2006) Non iterative approach to dynamic traffic origin-destination estimation using parallel evolutionary algorithms. *Transportation Research Record: Journal of the Transportation Research Board* 1964:201 – 210.
- Lam WHK, Lo HP (1990). Accuracy of O-D Estimates from Traffic Counting Stations. *Traffic Engineering and Control*, 31:358-367.

- Lam WHK, Shao H, Sumalee A (2008). Modeling Impacts of Adverse Weather Conditions on a Road Network with Uncertainties in Demand and Supply. *Transportation Research Part B*, 42(10): 890-910.
- Lauren A, Jiang S, Murga M, González MC (2015) Origin – destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C* 58:240-250.
- Lu CC, Zhou X, Zhang K (2013) Dynamic origin – destination demand flow estimation under congested traffic conditions. *Transportation Research Part C* 34:16 – 37.
- Lu Z, Rao W, Wu Y, Guo L, Xia J (2014). A Kalman filter approach to dynamic OD flow estimation for urban road networks using multi-sensor data. *Journal of advance transportation* 49(2):210-227.
- Lundgren JT, Peterson A (2008) A heuristic for the bilevel origin – destination-matrix estimation problem. *Transportation Research Part B* 42 (4):339 – 354.
- Maher MJ (1983) Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. *Transportation Research Part B* 17 (6):435–447.
- Minguez R, Sanchez-Cambronero S, Castillo E, Jimenez P (2010). Optimal Traffic Plate Scanning Location for OD Trip Matrix and Route Estimation in Road Networks. *Transportation Research Part B*, 44(2): 282-298.
- Ng M (2012). Synergistic Sensor Location for Link Flow Inference without Path Enumeration: A Node-based Approach. *Transportation Research Part B*, 46(6): 781-788.
- Nie YM, Zhang HM (2008) A variational inequality approach for inferring dynamic origin – destination travel demands. *Transportation Research Part B* 42 (7):635 – 662.
- Nie YM, Zhang HM (2008) A relaxation approach for estimating origin-destination trip tables. *Netw Spatial Econ* 10:147-172.

- Parry M, Hazelton ML (2012) Estimation of origin–destination matrices from link counts and sporadic routing data. *Transportation Research Part B* 46(1):175–188.
- Perrakis K, Karlis D, Cools M, Janssens D, Vanhoof K, Wets G (2012) A Bayesian Approach for Modeling Origin–Destination Matrices. *Transportation Research Part B*, 46(1):200–212.
- Qian Z, Zhang HM (2011) Computing individual path marginal cost in networks with queue spillbacks. *Transportation Research Record: Journal of the Transportation Research Board* 2263:9 – 18.
- Shao H, Lam WHK, Sumalee A, Chen A, Hazelton ML (2014). Estimation of Mean and Covariance of Peak Hour Origin – Destination Demands from Day-to-Day Traffic Counts. *Transportation Research Part B*, 68: 52-75.
- Sherali HD, Park T (2001) Estimation of dynamic origin–destination trip tables for a general network. *Transportation Research Part B* 35 (3):217–235.
- Simonelli F, Marzano V, Papola A, Vitiello I (2012). A Network Sensor Location Procedure Accounting for O-D matrix estimate variability. *Transportation Research Part B*, 46(10): 1624-1638.
- Spieß H (1987) A maximum likelihood model for estimating origin – destination matrices. *Transportation Research Part B* 21 (5):395 – 412.
- Tavan H (2001) Internally-Consistent Estimation of Dynamic Network Origin–Destination Flows from Intelligent Transportation Systems Data Using Bi-level Optimization. Dissertation, The University of Texas at Austin.
- Tebaldi C, West M (1998) Inference on network traffic using link count data. *Journal of the American Statistical Association* 93:557 – 576.
- Tympakianaki A, Koutsopoulos HN, Jenelius E (2015) c-SPSA: Cluster-wise simultaneous perturbation stochastic approximation algorithm and its application to dynamic origin–destination matrix estimation. *Transportation Research Part C* 55:231-245.

- Van der Zijpp NJ (1997) Dynamic OD-Matrix Estimation From Traffic Counts and Automated Vehicle Identification Data. *Transportation Research Record: Journal of the Transportation Research Board* 1607:87–94.
- Van Zuylen HJ, Willumsen L (1980) Network tomography: the most likely trip matrix estimated from traffic-counts. *Transportation Research Part B* 14 (3):291 – 293.
- Viti F, Rinaldi M, Corman F, Tampère CMJ (2014). Assessing Partial Observability in Network Sensor Location Problems. *Transportation Research Part B*, 70: 65-89.
- Viti F, Verbeke W, Tampère CMJ (2008). Sensor Locations for Reliable Travel Time Prediction and Dynamic Management of Traffic Networks. *Transportation Research Record: Journal of the Transportation Research Board*, 2049:103-110.
- Wei C, Asakura Y (2013) A Bayesian approach to traffic estimation in stochastic user equilibrium networks. *Transportation Research Part C* 36:446 – 459.
- Xie C, Kockelman KM, Waller ST (2011) A maximum entropy-least squares estimator for elastic origin – destination trip matrix estimation. *Transportation Research Part B* 45 (9):1465 – 1482.
- Xing T, Zhou X, Taylor J (2013). Designing Heterogeneous Sensor Networks for Estimating and Predicting Path Travel Time Dynamics: An Information-Theoretic Modeling Approach. *Transportation Research Part B*, 57: 66-90.
- Yang H, Akiyama T, Sasaki T (1998) Estimation of time-varying origin-destination flows from traffic counts: A neural network approach. *Mathematical and Computer Modelling* 27(9-11):323-334.
- 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(5): 351-363.
- Yang H, Yang C, Gan LP (2006). Models and Algorithms for the Screen Line-based Traffic-Counting Location Problems. *Computer & Operations Research*, 33: 836-858.

- Yang H, Zhou J (1998). Optimal Traffic Counting Locations for Origin-Destination Matrix estimation. *Transportation Research Part B*, 32(2):109-126.
- Yim KN, Lam WHK (1998). Evaluation of Count Location Selection Methods for Estimation of OD Matrices. *Journal of Transportation Engineering*, 124(4): 376-383.
- Zhou X, List GF (2010). An Information-Theoretic Sensor Location Model for Traffic Origin-Destination Demand Estimation Applications. *Transportation Science*, 44(2): 254-273.
- Zhou X, Mahmassani HS 2006. Dynamic OD demand estimation using automatic vehicle identification data. *IEEE Transactions on Intelligent Transportation Systems* 7 (1):105-114.
- Zhou X, Mahmassani HS (2007) A structural state space model for real-time traffic origin-destination demand estimation and prediction in a day-to-day learning framework. *Transportation Research Part B* 41 (8):823–840.
- Zhu S, Cheng L, Chu Z, Chen A, Chen J (2014). Identification of Network Sensor Locations for Traffic Flow Estimation. *Transportation Research Record: Journal of the Transportation Research Board*, 2443:32-39.
- Zhu S, Zheng H, Peeta S, Guo Y, Cheng L, Sun W (2016). Optimal Heterogeneous Sensor Deployment Strategy for Dynamic Origin-Destination Demand Estimation. *Transportation Research Record: Journal of the transportation research board*, 2567. (DOI: 10.3141/2567-03).
- Zhu S, Cheng L, Sun C, Chen J. Network Sensor Location Models Accounting for Variability of Traffic Flow Estimation. Presented at 94th Transportation Research Board Annual Meeting, Washington DC, 2015.

Contacts

For more information:

Srinivas Peeta
Purdue University
3000 Kent Ave, NEXTRANS Center
Ph: (765) 496 9726
Fax: (765) 807 3123
peeta@purdue.edu
<https://engineering.purdue.edu/~peeta/>

NEXTRANS Center
Purdue University - Discovery Park
3000 Kent Ave.
West Lafayette, IN 47906

nextrans@purdue.edu
(765) 496-9724

www.purdue.edu/dp/nextrans