Optimization of Traffic Count Locations for Estimation of Stochastic Origin-Destination Demands under uncertainty with Sensor Failure

Hao FU [a], William H.K. LAM [b] and Hu SHAO [c]
[a] *Department of Civil and Environmental Engineering,*
The Hong Kong Polytechnic University, China
Email: hao.fu@connect.polyu.hk
[b] *Department of Civil and Environmental Engineering,*
The Hong Kong Polytechnic University, China
Email: william.lam@polyu.edu.hk
[c] *Department of Mathematics, School of Sciences,*
China University of Mining and Technology, China
Email: h.shao@163.com

ABSTRACT

Stochastic OD demands are usually estimated from the link flows observed by traffic counting sensors over time. Unavoidably, traffic counting sensors located in the road network are subject to failure such that these links with failed sensors are not capable to obtain the link flows. This paper addresses the traffic count location optimization problem considering sensor failure to estimate mean and covariance of OD demands. The information loss of stochastic OD demands due to failed sensors can be quantified by the proposed criteria. Based on these criteria, the traffic count locations are optimized to minimize the information loss of stochastic OD demand estimates considering the uncertainty of sensor failure. To solve the proposed integer programming model, the Genetic Algorithm (GA) is used. Numerical examples are presented to demonstrate the effects of sensor failure on the estimation accuracy of stochastic OD demands.

Keywords: Sensor locations, Stochastic OD estimation, Sensor failure, Covariance

1 INTRODUCTION

1.1 Traffic count location problem

As the increasing attention on big data, traffic counting sensors have been and will be located in the road network for various purposes, such as OD estimation, link flow observation, travel time estimation.

Some previous studies considered the traffic count location problem in which traffic count locations were directly determined based on the network structure/topology but without explicit use of prior OD information (Yang and Zhou, 1998). There are some studies considering both traffic count location and OD estimation problems in which the traffic count locations were determined together with the estimation of OD demands. The objectives of these traffic count location models are to minimize the uncertainty of posterior OD demands (Zhou and List, 2010).

1.2 Failure of sensors

Concerning the failure of sensors as a significant factor that can affect the expected information gain from sensors, we identified limited studies that take into account the sensor failure consideration for flow observability and OD demand estimation purposes. As one of the studies in this area, Li and Ouyang (2011) identified the location of AVI sensors in traffic to maximize the route flow information gain considering sensors failure. Danczyk et al. (2016) studied the deployment of counting sensors on a freeway to minimize the overall freeway performance monitoring errors with respect to sensors failure. In another study, Salari et al. (2019) addressed the full/partial link flow observability problem trying to minimize the adverse effect of sensors failure on the link flow inference process. However, the consideration of sensor failure for stochastic OD demand estimation is absent in current research studies pertaining to the traffic count location problems.

1.3 Covariance relationship between OD pairs

Most of the existing studies on Network Sensor Location Problems (NSLP) focus on the optimization of traffic count locations for the best estimates of mean OD demand. However, it has been pointed out by some scholars that not only the mean OD demands but also the covariance of OD demand should be considered for OD estimation. Variation of the expected total travel time in the stochastic road network is mainly attributed to the covariance of OD demand.

An increasing attention has been received for the estimation of both the mean and covariance of OD demands. Yang et al. (2018) preserved all structural information obtained from different sensors to estimate stochastic travel demands. For OD demand covariance estimation, Singhal and Michailidis (2007) used a latent variable model to show sufficient conditions of n-order moments.

In this paper, we attempt to address the multi-type sensor location problem for stochastic OD demand and link travel time estimation with explicit consideration of sensor failure. The main contributions of this study can be viewed from two aspects: theoretical and methodological sides. For the theoretical aspect, the covariance relationships of traffic flows between different OD pairs and of travel time between different links have been considered for more accurate estimates. In addition, the failure rate of sensors is also considered in the multi-type sensor location problem to enhance the reliability of the road networks equipped with sensors. For the methodological aspect, we proposed a mathematical model that considers the effect of sensor failure on the multi-type sensor location problem for stochastic OD demand and link travel time estimation.

2 MODEL FORMULATION

2.1 Relationship between link flows and OD demands

The traffic flows on a link are observed by a traffic counting sensor during the same peak hour period (say 8:00 am - 9:00 am) from day to day. The link with (without) traffic sensor is called “observed link” (“unobserved link”) represented by $\tilde{\mathbf{A}}$ throughout this paper. Because of the daily fluctuations in travel demand, the link flow is not deterministic.

Then, the mean of observed link flow can be obtained by the following equation:

$$v_a = E[V_a] = E\left[\sum_{w\in W} p_{a,w} Q_w\right] = \sum_{w\in W} p_{a,w} q_w \qquad \forall a \in \tilde{\mathbf{A}} \tag{1}$$

Where $p_{a,w}$ is the link choice proportion, which is the possibility of link *a* traversed by OD pair *w*. q_w is the OD demand to be estimated on OD pair *w*.

The covariance between V_{a1} and V_{a2} can be deduced as:

$$\sigma^v_{a_1,a_2} = \mathrm{cov}\left[V_{a_1}, V_{a_2}\right] = \sum_{w_1\in\mathbf{W}}\sum_{w_2\in\mathbf{W}} p_{a_1,w_1} p_{a_2,w_2} \sigma^q_{w_1,w_2} \qquad \forall a_1, a_2 \in \tilde{\mathbf{A}} \tag{2}$$

Where $\sigma^q_{w_1,w_2}$ is the covariance between OD demands w_1 and w_2.

2.2 Stochastic OD demand estimation

Due to that the OD demands will change during the iteration, the stochastic link choice proportions and stochastic link flows are also updated during determining traffic count locations. The Bayes method for estimating the stochastic OD demands is formulated as follows:

Suppose that observed link flows follow a multivariate normal distribution, $V\,|\,Q \sim MVN(PQ, \Sigma^v)$, where V is an $m\times 1$ vector of observed link flow, Q is an $n\times 1$ parameter vector of estimated OD demands, P is an $m\times n$ given matrix of link choice proportion, and Σ^v is an $m\times m$ observed covariance matrix of link flow. In addition, suppose the prior distribution of OD demand is also multivariate normal, $Q \sim MVN(Q^{prior}, \Sigma_q^{prior})$, where Q^{prior} is an $n\times 1$ parameter vector of prior OD demands, Σ_q^{prior} is an $n\times n$ covariance matrix of prior OD demands. From Bayes method, we can get the posterior distribution of OD demands Q as follows:

$$Q\,|\,V \sim MVN(Dd, D) \tag{3}$$

Where

$$D^{-1} = P'\Sigma^{v-1}P + (\Sigma_q^{prior})^{-1} \tag{4}$$

$$d = P'\Sigma^{v-1}V + (\Sigma_q^{prior})^{-1}Q^{prior} \tag{5}$$

Therefore, the mean OD demands can be estimated.

$$\mathbf{q} = E(Q\,|\,V) = Dd \tag{6}$$

The posterior covariance matrix of OD demand can be expressed as follows:

$$\mathbf{\Sigma^q} = Cov(Q\,|\,V) = D \tag{7}$$

Based on the estimation of mean and covariance OD demands, stochastic link choice proportions together with the stochastic link flows are then updated by an adapted traffic flow simulator. The initial method can be referred to Lam and Xu (1999). We assume that OD demands follow a multivariate normal distribution. By sampling the OD demand from the overall population, SUE assignment is used to obtain mean and covariance of link flows and link choice proportions. Stochastic OD demands will be updated according to the proposed Bayes method.

2.3 Information loss for each OD pair

If a sensor installed on a link fails, the traffic flow information on this particular link cannot be observed. In this paper, we assume that the probability of missing the link flow observability of any observed link equals to the probability of failure of the sensor installed on that link. The failure of traffic sensors installed on links affects the estimation accuracy of stochastic OD demand. We are intended to quantify the information loss of OD demand estimates due to the failure of a sensor. The following equation measures the mean OD demand information loss for each OD pair:

$$L_w(z) = q_w f_a^{\sum_{a\in A}\varphi_{a,w} z_a} \qquad \forall w \in W \tag{8}$$

Where L_w is denoted as the mean OD demand information loss for OD pair *w* considering the sensor failure. f_a ranged from 0 to 1 is the failure probability of a traffic counting sensor located on link *a*. $\varphi_{a,w}$ is the link-OD incidence that signifies if OD pair *w* traverses link *a* ($\varphi_{a,w} = 1$) or not ($\varphi_{a,w} = 0$). z_a is the sensor location variable that signifies if a traffic counting sensor is located on link *a* ($z_a = 1$) or not ($z_a = 0$). $\sum_{a\in A}\varphi_{a,w} z_a$ calculates the number of observed links traversed by OD pair *w*. It can be seen from Equation (8) that the more observed links traversed by the OD pair, the less information loss for this OD demand with sensor failure.

Regarding the OD demand covariance which relates to a pair of OD demands, the covariance of OD demand information loss for each OD pair can be defined as follows:

$$L_{w_1,w_2}(z) = \left|\sigma^q_{w_1,w_2}\right| \left(f_a^{\sum_{a\in A}\varphi_{a,w_1} z_a} + f_a^{\sum_{a\in A}\varphi_{a,w_2} z_a} \right) \qquad \forall w_1, w_2 \in W \tag{9}$$

Where L_{w_1,w_2} is denoted as the covariance of OD demand information loss for OD pair *w* considering the sensor failure. In Equation (9), the more observed links traversed by either the OD pair, the less information loss for the OD demand covariance.

2.4 Optimization of traffic count locations with sensor failure

The aim of this paper is to enhance the reliability of the estimated OD demands by locating traffic counts in the road network considering the sensor failure. As the L_w and L_{w_1,w_2} can be regarded as two measures of the information loss for the estimated mean and covariance OD demands with sensor failure, the smaller the L_w and L_{w_1,w_2} are, the higher the reliability of the traffic count location scheme is. Then, minimizing L_w and L_{w_1,w_2} can enhance the reliability of the traffic count location schemes. Thus, a weighted-sum model is proposed for improving the estimation reliability of the OD mean and covariance matrices. As discussed above, the objective function can be expressed as follows:

$$\min \ \alpha \sum_{w\in W} L_w(z) / \sum_{w\in W} q_w + (1-\alpha) \sum_{w_1\in W}\sum_{w_2\in W} L_{w_1,w_2}(z) / \sum_{w_1\in W}\sum_{w_2\in W} \left|\sigma^q_{w_1,w_2}\right| \tag{10a}$$

However, there exist some general constraints that should be employed to reduce the unreasonable traffic count location schemes. These constraints include budget constraint, OD covering rule and link independence rule. The mathematical expressions if these adopted constraints are shown as follows, respectively.

$$\sum_{a\in\mathbf{A}} c_a z_a \le B \tag{10b}$$

$$\sum_{a\in A} \varphi_{a,w} z_a \ge 1 \qquad \forall w \in \mathbf{W} \tag{10c}$$

$$rank(\tilde{\mathrm{P}}) = \tilde{m} \tag{10d}$$

Where c_a is the installing and maintaining budget of one traffic counting sensor. B is the total budget.

It should be noted that in the objective function, the weighting parameter α implies the level of importance of mean and covariance of OD demands. Specifically, if the managers have more interests on the mean OD demand than its covariance, the weighting parameter α on information loss for mean OD demand estimates should be larger, vice versa.

3 SOLUTION ALGORITHMS

The proposed traffic count location optimization in this study is a non-convex problem with binary decision variables that represent the traffic count locations. This problem is NP-hard such that heuristic solution algorithm (e.g. Genetic algorithm) is used in this study.

Chromosome generation and representation

In the GA, a chromosome length equals the number of links in a network. The cells associated with the randomly selected links from each set of new links, which are considered as the unobserved links, should be set to zero. The population size in this solution algorithm is set to 30 in each iteration.

Fitness function

Eqn. (10a) is used as the fitness function in this solution algorithm.

Mutation and crossover procedures

The crossover procedure is applied to the proposed GA to increase the chance of reproduction of the chromosomes that stand in a higher rank considering the fitness function. Moreover, the mutation procedure is employed to keep the genetic diversity in the generation of a population.

Stopping criteria

The generational process in GA is repeated until a termination condition is reached.

4 NUMERICAL EXAMPLES

To demonstrate the performance of the proposed traffic count location optimization model considering sensor failure, Tuen Mun Road in Hong Kong has been chosen for analysis in this paper. The network consists of 3 zones, 4 nodes and 10 links as shown in Fig. 1. The observed link flows by traffic counting sensors are assumed based on the data provided by Annual Traffic Census (ATC) in Hong Kong. The prior and true OD demands are also assumed.

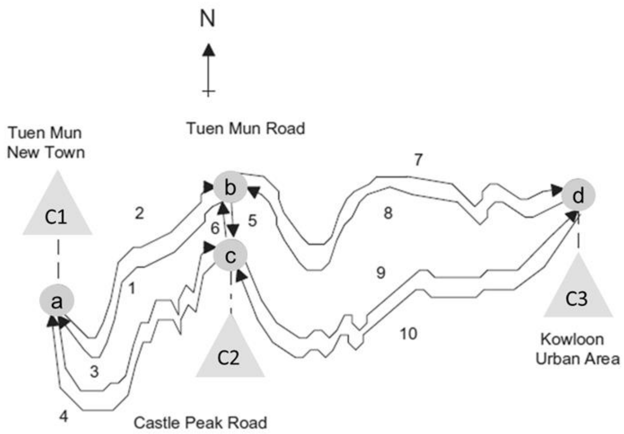

Fig. 1 Tuen Mun Road corridor network

The numerical examples presented in this section are to illustrate (i) the effects of sensor failure on the estimation accuracy of stochastic OD demands and (ii) the effects of mean and covariance of OD demands on the traffic count locations under uncertainty of the sensor failure.

4.1 Effects of sensor failure on the estimation accuracy

In this example, the weighting parameter α is set to 0.5 that means we treat the mean and covariance of OD demands as the same importance. For simplicity, the probability of sensor failure is assumed to be a constant 0.1. That is to say the failure probability is always 10%. Given the number of sensors, the optimal traffic count location schemes are shown as follows

Table 1 Effects of number of sensors on the information loss

Number of sensors	Optimal traffic count locations	Information loss		
		Mean OD demands	Covariance of OD demands	Average (Eqn. (13a))
1	{6}	6.67	5.83	6.25
2	**{6,7}**	6.52	5.75	**6.14**
3	**{6,7,9}**	6.02	5.2	**5.61**
4	**{5,6,7,9}**	5.95	5.08	**5.52**
5	{2,5,6,7,9}	5.16	4.31	4.74

It can be seen from Table 1 that the information loss for both mean and covariance of OD demand estimation decreases with the increasing number of sensors. The more sensor located on the road network, the more reliable of the sensor location scheme for stochastic OD demand estimation. The results shown in Table 1 can also demonstrated that links traversed by more OD demands have priority to be installed with traffic counting sensors.

From Table 1, it can also be observed that the difference of average information loss between schemes 2 and 3 is much larger than that between schemes 3 and 4. To clarify this observation, we note from the OD-link incidence matrix that the link 7 and link 9 are both traversed by three OD pairs, while link 6 by four OD pairs. In other words, the level of importance for the traffic counting sensors on link 7 or link 9 is similar so that the different of average information loss between scheme 3 and 4 is not obvious. However, if the sensors on both these two links fail, the information loss increases significantly.

4.2 Effects of stochastic OD demands on the traffic count locations under uncertainty of the sensor failure

To further demonstrate the contributions of this paper, we would like to understand how the mean and covariance of OD demands affect the traffic count locations under uncertainty of the sensor failure. Therefore, two different scenarios are implemented: (i) Only the information loss of mean OD demands is considered if $\alpha = 1$; and (ii) Only the information loss of covariance of OD demands is considered if $\alpha = 0$. The results of optimal traffic count locations and average information loss for different numbers of sensors are shown in Table 2.

Table 2 Effects of mean and covariance of OD demands

Number of sensors	Optimal traffic count locations		Average information loss	
	Only mean ($\alpha = 1$)	Only covariance ($\alpha = 0$)	Only mean ($\alpha = 1$)	Only covariance ($\alpha = 0$)
2	**{5,6}**	**{6,9}**	6.41	5.72
3	{5,6,7}	{5,6,9}	5.96	5.31
4	{5,6,7,9}	{5,6,7,9}	5.69	5.01

As can be seen from Table 2 that the optimal traffic count locations are different with consideration of the information loss of mean or covariance of OD demands. Different traffic count location schemes result in different information loss of mean and covariance of OD demands. If only the information loss of mean OD demands is considered, the links traversed by more OD pairs have priority to be located with sensors. If only the information loss of covariance of OD demands is considered, the links with more **common** traversed OD pairs have priority. For instance, when there are two sensors needed to be located on the road network, the optimal link set is {5,6} if only information loss of mean OD demands is considered; while the optimal link set becomes {6,9} if only information loss of OD demand covariance is considered. From the road network shown in Fig. 1, we can find that link 5 and link 6 are both traversed by four OD pairs, while link 9 is only traversed by three OD pairs. However, the common traversed OD pairs for links 5 and 6 are only two, but for links 6 and 9 are three. Therefore, the optimal link set for information loss of mean OD demands is links {5,6}, while for information loss of OD demand covariance is links {6,9}.

5 CONCLUSIONS AND FURTHER STUDIES

The existing studies focused on traffic count location problems for various purposes, such as OD demand estimation and link flow observability. However, with the uncertainty of sensor failure, different sets of observed links will result in different probabilities of information loss of the estimated stochastic OD demands in a network. This paper investigates traffic count location problems considering sensor failure to obtain more reliable results on stochastic OD demand estimates. The proposed model is applicable for the road network to locate traffic counting sensors so as to minimize the influence of a failed sensor on stochastic OD demand estimation.

In this paper, the possibility of sensor failure has been considered into the traffic count location problem so that the influence of a failed sensor on the estimation reliability can be minimized. In the numerical examples, the effects of sensor failure on stochastic OD demand estimation have been examined. It can be concluded from the numerical example that the information loss for both mean and covariance of OD demand estimation decreases with the increasing number of sensors. To reduce the influence of sensor failure on mean OD demand estimation, links traversed by more OD demands have priority to be installed with traffic counting sensors. On the other hand, for OD demand covariance estimation considering uncertainty of sensor failure, it can be concluded from the numerical examples that the links with more common traversed OD pairs have priority.

Further research should be carried out to investigate the sensor location optimization problems to estimate both OD demands and link travel times considering sensor failure. In addition, it would be worth exploring to extend the proposed model to consider the effects of multi-user classes and their covariance on the optimal traffic count location problem in a multi-modal traffic networks, when the information on vehicle composition and occupancy are available.

Acknowledgements

The work described in this paper was financially supported by grants from National Natural Science Foundation of China (Project Nos. 71671184), together with grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Nos. PolyU 152628/16E and R5029-18).

REFERENCES

Danczyk, A., Di, X. and Liu, H. X. (2016) A probabilistic optimization model for allocating freeway sensors, *Transportation Research Part C: Emerging Technologies*, 67, pp. 378–398.

Ehlert, A., Bell, M. G. H. and Grosso, S. (2006) The optimisation of traffic count locations in road networks, *Transportation Research Part B: Methodological*. Pergamon, 40(6), pp. 460–479.

Lam, W. H. K. and Xu, G. (1999) A traffic flow simulator for network reliability assessment, *Journal of Advanced Transportation*. John Wiley & Sons, Ltd, 33(2), pp. 159–182.

Li, X. and Ouyang, Y. (2011) Reliable sensor deployment for network traffic surveillance, *Transportation Research Part B: Methodological*. Pergamon, 45(1), pp. 218–231.

Salari, M., Kattan, L., Lam, W. H. K., Lo, H. P. and Esfeh, M. A. (2019) Optimization of traffic sensor location for complete link flow observability in traffic network considering sensor failure, *Transportation Research Part B: Methodological*. Pergamon, 121, pp. 216–251.

Singhal, H. and Michailidis, G. (2007) Identifiability of flow distributions from link measurements with applications to computer networks, *Inverse Problems*. IOP Publishing, 23(5), pp. 1821–1849.

Yang, H. and Zhou, J. (1998) Optimal traffic counting locations for origin–destination matrix estimation, *Transportation Research Part B: Methodological*. Pergamon, 32(2), pp. 109–126.

Yang, Y., Fan, Y. and Wets, R. J. B. (2018) Stochastic travel demand estimation: Improving network identifiability using multi-day observation sets, *Transportation Research Part B: Methodological*. Pergamon, 107, pp. 192–211.

Zhou, X. and List, G. F. (2010) An information-theoretic sensor location model for traffic origin-destination demand estimation applications, *Transportation Science*. INFORMS, 44(2), pp. 254–273.