EFFECTIVENESS OF DATA COMPOSITION ON OD MATRIX ESTIMATION

BERNARD FRIEDRICH, YUN-PANG WANG
Professor, Ph.D. Student
Institute for Transport, Road Engineering and Planning,
University of Hannover, Appelstr. 9A, 30167 Hannover, Germany
friedrich@ivh.uni-hannover.de; wang@ivh.uni-hannover.de

SUMMARY

In this paper the effectiveness of two types of information with different number of detector deployed intersections on OD matrix estimation is investigated. These two types of information are namely turning flows and inbound link flows at intersections. To do that an index is proposed to decide the priority of installing detectors at intersections, and three existing indices are adopted to check the performance of the proposed index on OD matrix estimation. The result shows that the information containing in turning flows contribute more to the accuracy of OD matrix estimation than that in inbound link flows at intersections, and the more the detector deployed intersections, the least important the knowledge of turning flows. It also indicates that the proposed index is slightly better than the others with the consideration of the operational simplicity.

INTRODUCTION

Considering the extensive cost of building an OD matrix, many models with the usage of traffic link counts have been proposed and widely applied for updating an old matrix over the last decades. Although with the dramatic advancement of innovative technologies traffic data can be completely and more realistically retrieved from electrical devices, comprehensive real time traffic information in networks is not easily available as it is costly to install and maintain these infrastructures. Thus, it is important to know which kind of flow information in networks is valuable so that detectors can be then deployed efficiently and effectively for OD matrix estimation.

The previous related studies are mainly focused on how to decide detector locations for achieving better OD matrix estimation. They can be divided into two groups. One group is to select the adequate traffic counting locations in order to infer the flow information for all links in networks. The respective OD matrix estimation can then be better estimated with complete link flow information. Binaco et al (1) proposed a two-stage approach to derive more flow information on links for improving the accuracy of OD matrix estimation. In stage one, the minimum number and location of sensors will be determined by solving a sensor location problem (SLP) in order to derive the link flows over the entire network. The measurement costs of sensors are considered and the turning ratios at all network nodes are given as a basic assumption. To solve the SLP two iterative heuristics finding a lower bound and an upper bound are also proposed with the consideration of selecting the nodes crossed by the largest number of OD paths. In stage two, the complete estimated link flow information and the link choice proportions of OD pairs from the prior matrix will be used for OD matrix estimation.
estimation. Gentili and Mirchandani (2) applied the complexity analysis searching for the adequate number of image sensors at intersections to collect necessary link flows and turning ratios in order to add new specific constraints for the unique solution of the OD estimates.

The other group is based on some rules/indices to select the important links for obtaining accurate OD estimates. The necessary number and location of links can be decided, when the minimal difference of the estimated and the actual OD matrices reaches. Hogberg (3) pointed out that links with expected high traffic flows and links passed by OD pairs and have yet to be counted on other links are more valuable than the others. Willumsen (3) applied the theory of entropy maximization to derive an equation for calculating the contribution of each new traffic count with the assumption that the old and present matrices are sufficiently close. A rough estimation for the new traffic counts is required in advance. This estimate is suggested to be obtained by road characteristics or 6-minute counts. Yang and Zhou (4) proposed four rules for selecting traffic counting points in order to enhance the reliability of the estimated OD matrix and then developed a heuristic algorithms and an integer linear programming model to determine the counting locations. These four rules are defined as: (a) OD Covering Rule; (b) Maximal flow fraction rule; (c) Maximal flow-intercepting rule; (d) Link independence rule. Based on these rules, links will be chosen to deploy detectors, when they are independent from another, can intercept as many flows as possible and can collect the large portion of trips for any OD pair. Yim and Lam (5) proposed the maximal OD selection method (MOD) with the consideration of eliminating the redundant information. This method contains two rules; (a) Maximal net OD captured rule, and (b) Maximal total OD captured rule. The link with the highest net OD flow will be chosen at each iteration until all the available links are checked. If two or more links have the same amount of net OD flow, the one with the greatest total OD flow will be selected, since more independent information can enhance the reliability of the OD estimates. Lotz (6) proposed a rule with the consideration of the accuracy of OD estimates. Based on the link flow information, the average deviation between the estimated and the actual OD flows for each OD pair is first calculated. The link, which contributes less deviation to OD estimates, will be primarily chosen to deploy with detectors. Tamin and Suyuti (7) pointed out three major factors for determining adequate number and location of detectors: (a) proportion ratio of the trips for each OD pair at each link; (b) independence and inconsistency conditions of traffic flows; (c) physical conditions of links. Based on the above three factors a multi-criteria evaluation procedure is developed. The procedure is to prioritize the independent links which are used by many OD pairs and at the same time serve a large portion of OD flows. The degree of saturation and the side friction factor of links are then taken into consideration to decide the priority list of links in networks. Lastly, the relationship between the level of accuracy of the estimated matrices compared to the actual one and the number of selected traffic counts is analyzed to determine the best number and location of traffic counts. Kim et al (8) developed a link-based model with the application of three algorithms, Greedy Adding (GA), Greedy Adding and Substitution (GAS) and Branch and Bound (BB) Algorithm, to determine the set of optimal links, which achieve the minimum cost of traffic counting and at the same time covering all OD pairs. The main rule adopted in these three algorithms is to select the links which serve a large number of OD pairs until all the OD pairs are observed.

From the above studies it is apparent that the captured number of OD pairs, the captured amount of OD flows and the redundant information among links are considered as the main factors, when applying rules to select important traffic counting points. It also shows that these studies are mainly focused on selecting the links for collecting link flows, not for collecting turning flows at intersections. Matschke and Friedrich (9) have examined and
concluded that the more information from link flows and turning flows, the better the OD estimates. Matschke and Friedrich (10) also proposed that the turning rates at intersections can be inferred from the combination of detection time and phase status. By this procedure, the local traffic flow variations can not only be detected, but also be used to infer the traffic flows for the surrounding links without detectors, which helps improve the accuracy of OD matrix estimation. However, extra detectors at the outbound links of intersections need to be installed when applying this method and the related cost for installing and maintaining detectors at each intersection will thus be incurred. It is thus necessary to verify the effectiveness of the information from turning flows on OD matrix estimation to assure the benefits.

To verify that effectiveness flow information in networks is classified into two groups with the consideration of information independency in this research. One group is the information from turning flows and the other one is that from inbound link flows at intersections. The influence of the different number of detector deployed intersections on the effectiveness of these two types of flow information on OD matrix estimation is also investigated. To do that an index is proposed to prioritize the traffic counting intersections with the consideration of the above mentioned important factors and the operational simplicity. In addition, three indices from the above mentioned researches are adopted to check the performance of the proposed index on OD matrix estimation. Other then that, the relationship between the number of detector-deployed intersections and the accuracy of OD matrix estimation will be also examined to check if a certain percentage of installing detectors at intersections in networks can be drawn. To achieve the above objectives test networks will be built with the elaboration of the turning movements at intersections as showed in figure 1.

Figure 1 Layout of test network with the elaboration of turning movements at intersections
SELECTION OF TRAFFIC COUNTING INTERSECTIONS

In the following the proposed index will first be introduced. The three adopted indices are chosen with the consideration of the operational simplicity consistent with the principle of Index FW. Each index will be also interpreted below, on how to apply for selecting traffic counting spots from intersections instead of from links.

Index FW

The fundamental constraint used in models for estimating OD matrices from traffic counts is

$$\sum_{g} T_{y} \cdot p_{y}^{a} = V_{a}^{obs} \quad [1]$$

Where $T_{y}$ = the estimated OD flow from origin i to destination j

$V_{a}^{obs}$ = the observed traffic flow for link a

$p_{y}^{a}$ = the link choice proportion of OD flow from origin i to destination j passing link a.

It interprets that the summation of the estimated OD flow from origin i to destination j timing the link choice proportion of OD flow from j to destination j using link a should be consistent with the observed traffic volumes for link a. Thus with the estimated $p_{y}^{a}$, based on a historic matrix and the given traffic flows, the respective OD matrix can be estimated by adopting one of OD matrix estimation models. It is thus explicit to know that $p_{y}^{a}$ is one of the most important parameters. The way to obtain $p_{y}^{a}$ is generally by assigning the respective representable historic OD matrix with an adequate assignment model onto the investigated network.

With the consideration of operational simplicity index FW is proposed, with the key parameter $p_{y}^{a}$ and the relative cost of detector deployment. Its mathematic form is defined as

$$FW = \sum_{a} \sum_{y} (p_{y}^{a}) \cdot \left( \frac{N_{A}}{N_{eff}} \right) \quad \forall \ a, \text{ and } a \in A \quad [2]$$

Where $p_{y}^{a}$ = the link choice proportion of OD flow from origin i to destination j passing link a

$N_{A}$ = the number of possible turning movements or inbound traffic movements at intersection A. It presents the number of detectors for collecting complete information.

$N_{eff}$ = the number of turning movements or inbound traffic movements, which are used at least by one OD pair, at intersection A. It presents the number of necessary detectors.

$N_{A}/N_{eff}$ = the relative cost saving ratio of detector deployment at intersection A

By this index the amount of valuable information, presented by the total amount of $p_{y}^{a}$ value containing at each intersection, is calculated and weighted with the relative cost saving ratio of detector deployment, based on the number of traffic movements used by OD flows. A great FW value represents the respective intersection containing significant amount of valuable information with less cost of detector deployment for OD matrix estimation.
Index L

The concept of index L is proposed by Lotz (6) and is defined as

\[ L = \sqrt{\frac{T}{T} \sum_{i,j} (T_{ij} - \tilde{T}_{ij})^2} \quad \forall \ A \]  \hspace{1cm} [3]

Where \( T_{ij} \) = the estimated OD flow from origin i to destination j based on the flow information at intersection A
\( \tilde{T}_{ij} \) = the actual OD flow from origin i to destination j
\( T \) = the number of OD pairs in the investigated network, which does not include the intra-zonal OD pairs, e.g. from origin i to destination i.

Based on the flow information (turning flows or link flows) at each single intersection, the average deviation between each estimated and actual OD flows is firstly calculated for each single intersection. The intersection with smaller deviation has more contribution on OD matrix estimation and will be primarily chosen to deploy with detectors.

Index KCC

It is based on the rule proposed by Kim et al (8) and is defined as

\[ KCC = \sum_{a \in A} \sum_{i,j} (N_{ij}^a) \quad \forall \ a, \text{ and } a \in A \]  \hspace{1cm} [4]

Where \( N_{ij}^a \) = the number of OD pair from origin i to destination j passing link a, which belongs to intersection A

According to this index the number of OD pairs passing the links at each intersection will be cumulated. The intersection with the largest number of OD pairs will be firstly selected to deploy detectors.

Index YL

It is based on the rules proposed by Yim and Lam (5) and is defined as

\[ YL = \sum_{a \in A} (F_{ij}^{max}) \quad F_{ij}^{max} = Max(F_{ij}(a)) \quad \forall \ a, \text{ and } a \in A, \text{ and } F_{ij}(a) = p_{ij}^a \cdot t_{ij} \quad \forall \ a, \text{ and } a \in A \]  \hspace{1cm} [5]

Where \( p_{ij}^a \) is defined as above
\( t_{ij} \) = the historical OD flow from origin i to destination j
\( F_{ij}^{max} \) = the maximal OD flow from origin i to destination j passing link a at intersection A

According to the index the maximal OD flow for each OD pair from i to j among links at each intersection will be firstly calculated and the maximal OD flows for all OD pairs at each intersection will be summed up as the index value. The intersection with great YL value will be foremost chosen to install detectors. When two or more intersections have the same YL value, the value of the total OD flow at each intersection will be the second criteria and the one with greater total OD flow will be selected.
EXPERIMENTAL STUDY

General Description

To examine the effectiveness of data composition on OD matrix estimation two networks containing nine and sixteen intersections are built respectively, as shown in figure 2 and 3. The OD matrix for each network is also established as the comparison base with the respective estimated matrices. Simulation software VISUM is also applied to assign traffic into the networks. The resultant traffic flows is then treated as the traffic data collected from detectors. To analyze the effectiveness of the information from turning flows on OD matrix estimation, two test networks are further built with the elaboration of turning movements. The structure is presented in figure 1.

Figure 2 Layout of test network 1
Figure 3 Layout of test network 2

One of the common models derived by Van Zuylen (11), is adopted for this research for conducting the OD matrix estimation. It is derived on the basis of the information measure proposed by Brillouin in 1952. In this model the information containing in each link is treated independently and is added up in order to obtain the most likely OD matrix, with the satisfaction of the constraints of the available traffic flows. Its equation is defined as:

\[
T_{ij} = t_{ij} \cdot \prod_a X_a^{\frac{t_{ij}}{s_{ij}}} , \quad g_{ij} = \sum_a p_a^{ij} , \quad \text{subject to} \quad \sum_g t_{ij} \cdot p_g^{ij} = V_a^{obs}
\]

Where \(T_{ij}\) = the estimated OD flow from origin i to destination j
\(t_{ij}\) = the historic OD flow or a prior guess of the OD flow from origin i to destination j
\(p_a^{ij}\) = the link choice proportion of OD flow from origin i to destination j passing link a

\[X_a^{s+1} = X_a^{s} \times \frac{V_a^{obs}}{V_a^{s}}\] , which is the adjustment factor for link a by using the ratio of the observed traffic flow \(V_a^{obs}\) and estimated traffic flows \(V_a^{s}\) for link a in each iteration.
In addition, the available traffic flow data will not only be used as constraints for solving the solutions, it is also used as quasi link capacities to calculate the respective link travel cost for getting more accurate $p_{ij}^a$ values. The cost function used here is based on the traffic assignment manual of the U.S. Bureau of Public Roads. It is defined as:

$$t_{\text{link}} = \frac{d}{s}(1+x^3)$$  \[7\]

Where $d$ = link distance

$s$ = travelling speed

$x$ = the ratio of flow and capacity or the ratio of flow and the detected traffic flow

For the performance evaluation root mean square error (RMSE) is adopted here to measure how closely the estimated flow of each OD pair tracks the actual one. It is defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{I} \sum_{j=1}^{J} (T_{ij} - \hat{T}_{ij})^2}{T}}$$  \[8\]

Where $T_{ij}$ = the estimated OD flow from origin i to destination j

$\hat{T}_{ij}$ = the actual OD flow from origin i to destination j

$T$ = the number of OD pairs in the investigated network, which does not include the intra-zonal OD pairs, e.g. from origin i to destination i.

Test Results

According to the aforementioned selection indices, index L, index KCC, index YL, index FW and the random selection, 10 priority lists for installing detectors at intersections are built with the information from turning flows and from inbound link flows in test network 1 and 2 respectively. The accuracy of the OD matrix estimation with different types of flow information and with different number of detector-deployed intersections for each priority list is then analyzed.

Effectiveness of Data Composition from Turning Flows and from Link Flows at Intersections on OD Matrix Estimation

Figures 4 and 5 show that better OD estimates can be obtained when more information are available in both small and larger networks. Moreover it shows that turning flows at intersections contribute in general more valuable information than inbound link flows at intersections. This is because the information from turning flows provides more specific flow constraints not only for solving $X_g$ in equation 6, but also for providing adequate link capacities. However it is also revealed in both test networks that when the number of detector deployed intersections increases, the accuracy difference between the OD matrix estimations with information from turning flows and from link flows becomes less. It is partially because of the accumulation of the unavoidable estimated $p_{ij}^a$ error from the assignment model when more information is collected from the turning flows at intersections. It thus implies that the more intersections equipped with detectors, the least importance of the information from turning flows.
Besides it appears that the improvement degree of the estimating accuracy in relation to the number of the detector employed intersections depends on the size of the network. In the test network 1 due to the less number of intersections the information at each intersection contributes more than that in test network 2. The improvement curves are thus steeper in test network 1 and smoother in test network 2. In both test networks the adequate percentage of detector-equipped intersections for good OD estimates cannot be inferred, since each intersection has certain amount of contributions on OD matrix estimation.

Figure 4 Accuracy of OD matrix estimation with the information from turning flows and from link flows based on different indices in test network 1
Figure 5  Accuracy of OD matrix estimation with the information from turning flows and from link flows based on different indices in test network 2

Performance of the Selection Indices

The accuracy of the OD matrix estimation based on the selected indices is further investigated and the result in figure 6 shows that the improvement degree with index FW is slightly more efficient than the other indices in test network 1. It is especially when the information from turning flows at intersections is available, since the $p_{ij}$ values are more accurate with more flow constraints in a smaller network. In network 2 the improvement degree with index L becomes more efficient when the number of detector deployed intersections is less than 12. When it is more than 12, the improvement degree with index FW becomes more efficient again. Generally, index F performs better than the others, taking into consideration the operational simplicity.

Figures 6-2, 6-3 and 6-4 indicate that the deviation of the accuracies of the OD matrix estimations based on these indices are quite small, especially when the network is larger. The differences between the maximal and minimal RMSEs with different quantity of detector equipped intersections is principally less than 15 % in test network 2.
CONCLUDING REMARKS

To verify the effectiveness of the information from turning flows on OD matrix estimation this paper applies four indices to prioritize counting intersections according to their contribution to OD matrix estimation instead of adopting the linear programming approach. Based on the selection results the accuracy of OD matrix estimation with the information from turning flows and from inbound link flows is examined. The result shows that turning flows contribute in general more information to OD matrix estimation than inbound link flows at intersections. However it is also revealed that the more the number of detector deployed intersections, the least important the knowledge of turning flows.

We also inspect the performance of index FW, index L, index KCC, index YL, and the random selection on OD matrix estimation. Although the accuracy difference of OD matrix estimation among these indices is not significant, index F is slightly more efficient for obtaining better OD estimates when considering the operational simplicity. The close result between the OD estimates from the random selection and from the other indices may be due to the small size of the test networks. It is therefore still better to use indices for systemically selecting traffic counting spots to assure the quality of OD estimates. Also note that, adequate

Figure 6 Deviation of the estimated RMSE error of OD matrix estimations based on the priority lists from Indices F, G, KCC, YL and the random selection
percentage of detector-equipped intersections for obtaining appropriate OD estimates cannot be inferred in this research, since each intersection has certain amount of contribution on OD matrix estimation.

In the future work, the effectiveness of combining the information partial from turning flows and link flows on OD matrix estimation should be further examined. The adequate percentage of detector deployed intersections and links in networks also deserve to be continuously explored.

REFERENCES