EFFICIENT DATA COLLECTION STRATEGIES FOR THE ESTIMATION OF O-D MATRICES FROM TRAFFIC COUNTS

William H.K. LAM
Dept. of Civil &
Structural Engg.
Hong Kong Polytechnic
Hong Kong

Michael G.H. BELL
Transport Operations
Research Group
University of Newcastle
Upon Tyne, England

INTRODUCTION

Estimation of the origin-destination (O-D) matrix directly from traffic counts offers great economic advantages because it eliminates expensive interview surveys, burdensome data editing and subsequent analysis. This subject has received a lot of attention in the last two decades and several techniques have been developed using flow and other information.

These techniques can be mainly classified as a log-linear model group, including the entropy maximisation (EM) and information minimisation (IM) models (see Van Zuylen and Willumsen, 1980), or a linear model group, such as the generalised least square (GLS) method. Cascetta and Nguyen (1988) review a range of models covering the two groups. Two fundamental problems faced when applying these models in practice are how to determine (1) which links to survey, and (2) how large the survey should be to obtain a reliable estimate of the O-D matrix.

This paper investigates how to determine the count locations and the sample size (i.e. the link coverage or the percentage of links counted) required for O-D estimation, particularly when there is no prior O-D and link flow dataset. First, the basic underlying theories of the three models are briefly reviewed. There follow discussions of different strategies for selecting links, together with their difficulties in applications. A selection index is proposed and a systematic sampling method is introduced for choosing traffic count locations. The effects of count location and coverage for the estimation of O-D matrix from link flows are illustrated by using a numerical example. Examinations of the efficiency of the link counts for O-D estimation are presented. Finally, some conclusions are drawn and an efficient data collection strategy is recommended for the estimation of O-D matrices from traffic counts.

1. ESTIMATION OF O-D MATRIX FROM TRAFFIC COUNTS

The fundamental equation in the estimation of an O-D matrix from traffic counts is

$$\sum_i \sum_j P^k_{ij}\, t_{ij} = V^k, \qquad \text{for } k = 1, 2, \ldots, K$$

$$\text{for } i, j = 1, 2, \ldots, S \qquad (1)$$

where t_{ij} is the number of trips from origin i to destination j,

V^k is the traffic count on link k, K is the set of links for which counts are available and P_{ij}^k is the link choice proportion which is the proportion of flow from i to j via link k. When V^k on some links and the link choice matrix (P_{ij}^k) are given, the trip matrix (t_{ij}) can be estimated using EM, IM or GLS models. In uncongested networks, one may assume that the link choice proportions are independent of the O-D matrix. However, this is not in general true. In congested networks, a bilevel programming approach is reqired to allow for equilibrium assignment (see Inaudi and et al., 1991). We assume fixed link choice proportions in this paper.

Both of the EM and IM model tend to equalize the cell values to the average number of trips per O-D pair. There are some specific properties of the EM and IM models which may induce bias in the O-D estimation (see Maher, 1987).

The generalised least square (GLS) model for estimation of trip matrices from traffic counts was introduced by Cascetta (1984). It allows the combination of survey data relating directly to O-D travel and traffic count data, while taking into account the error variances of these two data sources.

The GLS model is an efficient non-iterative procedure in getting the O-D estimates that replicate the observed link flows accurately. In fact, the GLS approach approximates the EM method when the link flows are known to a high level of accuracy. The quality of approximation depends on the quality of the prior O-D information. However, the disadvantages of the GLS model are with negative estimates. A simple iterative algorithm has been introduced by Bell (1991) to solve the problem of negative GLS fitted values subject to inequality constraints.

Lam and Lo (1990) investigated the effects of alternative input data sets on the accuracy of O-D estimates using EM model applied to a real-life situation. A systematic approach (namely, stepwise selection method) for selecting links with flow information was proposed and has shown to be useful if a good prior O-D trip matrix was available. However, this stepwise selection method would be time consuming in searching the optimal solution.

In practice, the key issue in the estimation of an O-D matrix from traffic counts is how to determine a set of links for which flow information should be available, particularly when there is no prior O-D and link flow dataset.

2. DATA REQUIRED FOR O-D ESTIMATION FROM TRAFFIC COUNTS

To define a strategy for collecting link-count data from a network, three factors must be included : (1) the locations of the surveyed links, (2) the priority order of the counted links, and (3) the sample size required (i.e. the link coverage or the percentage of links counted). The methods for selecting links with flow information are discussed in Section 2.1. An investigation of a systematic sampling method for improving the O-D estimation from traffic counts is given in Section 2.2. A selection index is defined and a link selection process is introduced in Section 2.3, together with the measures for the estimation reliability of the O-D matrix and link flows. Lastly, a measure of the effectiveness of the link flow data is given.

2.1 Methods for Selecting Count Locations

In this section, the methods for selecting count locations for the O-D estimation are discussed. Han and Sullivan (1983) examined three methods for collecting link-count data from a network, namely, (i) 'random selection', (ii) 'major link selection', and (iii) 'geographic pattern schemes'. Logie and Hynd (1990) suggested an 'O-D coverage' method for grouping count sites into screenlines. They are discussed as below:

(1) The Random Selection (RS)

The counted links are randomly selected from all the links of the network. This may lead to the undesirable result that only a small proportion of O-D trips has been observed.

(ii)The Major Link Selection (MLS)

The major links can be defined as those links with highest traffic flow. In other words, all links are sorted in descending order according to traffic flows (i.e. V^k). However, since the links are interrelated in the sense that the inclusion of one link in the model may reduce the contributions to O-D estimation of some other links (especially the adjacent links), this would lead to the inclusion of some "redundant" link flow information. Moreover, this method requires a complete set of link flow information in order to sort out the links by the order of traffic flow.

(iii)The GeographicPattern Scheme (GPS)

Links selected from a GPS strategy would form a particular geographic pattern, e.g. a cordon or screenline, or a combination of these two. Generally, roadside interviews and license plate surveys are conducted by locating a series of survey stations along screenlines so as to collect representative O-D data. Results of Han and Sullivan (1983) revealed that the GPS approach produces the most accurate O-D matrix at various levels of link coverage and is the most cost effective of the above three strategies.

(iv) The O-D Coverage (ODC)

Another criterion that can be used is to group count sites into screenlines that balance the objectives of:

(a) maximizing the number of O-D pairs that have all routes passing through the screenlines; and
(b) minimizing the number of O-D pairs per screenline, as this maximizes the information value of the counts for the corresponding matrix cells.

A description of exactly how the sites be chosen is not provided. Nevertheless, both of the GPS and ODC methods require the identification of screenlines. This is usually undertaken by experienced traffic engineers through trial-and-error on the basis of study objectives. Hence, it is important to develop a systematic sampling method for selecting traffic count locations particularly when there is no prior O-D and link flow information available.

2.2 Systematic Sampling Method

In order to ensure the reliability of the O-D matrix estimated from traffic counts in practice, there are certain prerequisites. First, the initial sites for traffic counts must be located so that the trips between any O-D pair can be observed. Second, the estimation algorithm should be capable of considering the error variances of the prior O-D estimates and link flow data as supplementary count data may be incorporated later. Third, a useful selection index should be well-defined in order to determine the additional count locations required. Finally, measures of fit to describe the reliability of the estimated O-D matrix and the accuracy of the estimated link flows can be defined where complete information on the "true" O-D trip table and all "observed" link flows may not be available in general.

Basically, the systematic sampling method proposed in this paper involves two distinct parts or steps; estimation of O-D matrix from traffic counts and determination of additional count locations in order to improve the O-D estimation effectively. These parts should be solved iteratively until an acceptable result is reached. The method may be stated as follows :

1) Conduct traffic counts at all centroid links (in one direction only) so as to obtain the total trip production or attraction information (i.e. ΣO_i or ΣD_j); link flows and turning movements will be used for O-D estimation and accuracy check respectively.

2) Get a prior O-D matrix initially by setting all cell values equal to the average number of trips per cell ($\bar{T}$);

$$\text{where} \quad \bar{T} = \sum_i O_i \,/\, N \text{ or } = \sum_j D_j \,/\, N \qquad (2)$$

N = the number of cells to be estimated in the O-D matrix.
= $(S^2 - S)$ if the intrazonal trips would not be estimated.

3) Set n = 0 and t^n_{ij} = prior O-D matrix.

4) Estimate link choice proportions P^k_{ij} by using a proportional assignment procedure (e.g. all-or-nothing).

5) Set n = n + 1, and substitute P^k_{ij} into GLS model to obtain a new trip matrix t^n_{ij} based on the prior O-D matrix and the traffic count data available (such as link flows at zone centroids initially).

6) Test whether the estimation reliability of solution (such as the turning movements and/or partial "true" O-D, if available) is acceptable. If yes, stop.

7) If not, identify the link priority and the sample size by a link selection process on the basis of the estimated t^n_{ij} and the estimated link flows V^k (see section 2.3); and then conduct

supplementary counts on the selected links. Go to Step 5 with the additional traffic flow data.

Step 1 can be replaced by conducting traffic counts at the external cordon links and using the outputs of trip generation model if planning data of internal traffic zones are available. In Step 2, an initial prior O-D matrix and the average number of trips per O-D pair can be obtained. It is expected that the prior O-D matrix will be getting closer to the "true" values during the above procedure. The link choice proportions could be determined externally in Step 4 with the use of the proportional assignment algorithm, if the congestion effects are negligible. Although EM and IM models can be employed in Step 5, GLS is proposed to be used for the estimation of O-D matrix because GLS is computationally efficient and is capable to take into account the error variances of O-D estimates and traffic count data. Step 6 in this method requires measures for the estimation reliability of the O-D matrix and link flows, while Step 7 would require a selection index and a cut-off point in order to determine the link priority and the sample size required for improving the estimation reliability.

2.3 Estimation Reliability and Selection Index

There exists a great need to find better criteria for evaluating final O-D estimates. Standard criteria, such as the Coefficient of Determination (R^2), based on comparing cell by cell "distance" are inadequate because of equal weight given to each error term particularly the small cell values. Further, inherent in many of these measures is the assumption that a "true" O-D matrix exists for comparison with the estimate. Rarely is this the case in practice, so it is impossible to evaluate exactly the estimation accuracy. In view of this, Yang, Iida and Sasaki (1991) proposed an index "Maximum Possible Relative Error" (MPRE) for the analysis of the scope for un-reliability of the estimated O-D matrix. The formulation and disadvantages of MPRE are discussed in the following.

Let T_{ij} and U^k be the real O-D trip matrix and observed flow on link k. Their relationship would be linear as shown in eqn (1). If the link choice proportions are known with certainty and the traffic counts are error free, the estimated link flows should be equal to the observed link volumes. Thus, we obtain

$$\sum_i \sum_j P^k_{ij} T_{ij} = U^k, \qquad \text{for } k = 1, 2, \ldots, K \qquad (3)$$

From (1) and (3), we have

$$\sum_i \sum_j P^k_{ij} (T_{ij} - t_{ij}) = 0 \qquad \text{for observed link } k \qquad (4)$$

The relative deviation (d_{ij}) of the estimated O-D trip from the true volume for the O-D pair from i to j is defined as

$$d_{ij} = (T_{ij} - t_{ij}) / t_{ij} \geq -1 \qquad (5)$$

(from T_{ij} and $t_{ij} > 0$)

Thus, eqn (4) can be rewritten as

$$\sum_i \sum_j P^k_{ij} \, d_{ij} \, t_{ij} = 0 \qquad \text{for observed link } k \qquad (6)$$

As a result, for a particular fitted O-D matrix t_{ij} using a selected estimation method (such as GLS, EM and IM), the relative deviation sum of square (SD) between the estimated and true O-D trips can be formulated as the following quadratic programming problem P1 :

$$\text{P1: Maximize } SD = \sum_i \sum_j (d_{ij})^2 \qquad \text{for N number of i-j pair (7)}$$

subject to eqn (5) and (6).

Therefore, the MPRE in the estimated O-D matrix is

$$MPRE = \sqrt{SD/N}\,. \qquad (8)$$

The MPRE can show the upper bound of the real error for a particular fitted O-D matrix. However, MPRE would always lead to under-estimate the reliability of the O-D estimation. In addition, it can be proved that the solution of problem P1 which may not be unique will be attained at the vertexes of its feasible region. In fact, the quadratic objective function in eqn (7) is strictly convex with respect to the variables, and the feasible region set by the linear constraints (5) and (6) is also convex. Hence, the minimization problem to the same objective function of eqn (7) and the relevant constraints would certainly have an unique solution; this solution would be located at the tangent point between the isoquant lines of the objective function and the feasible region. However, it is not the case for the MPRE. For the maximization problem P1, the isoquant lines surrounding the global minimium would be extended outward for increasing the objective function value, until the maximum points which are attained at some vertexes. As a result, only some of the O-D pairs would have values while the others are zero. The above mathematical difficulty has not been mentioned in Yang, Iida and Sasaki's paper (1991).

It appears that the problem P1 can be solved by checking the objective function values at all vertexes. However, it is impossible to checking all vertexes for large scale problems because of the computational burden.

In view of the computational inefficiency of MPRE and its over-estimation effect on the real relative error, a new selection index is proposed as a criterion for selecting traffic count locations on a road network; namely, "Estimated Relative Deviation from the Mean" (ERDM); where ERDM is in fact the chi-square value and may be defined as below:

$$ERDM = \sum_i \sum_j (t_{ij} - \bar{T})^2 / \bar{T}. \qquad (9)$$

As most of the O-D estimation models tend to equalize the cell values to their prior values subject to the traffic count constraints, the links which are likely to deviate the O-D estimates as far as possible from their prior values, or the mean ($\bar{T}$) if the prior values are not available, should be chosen for traffic counts particulary when the estimation reliabilty of the solution is not yet satisfactory. The ERDM function is also convex but can be evaluated easily since t_{ij} and $\bar{T}$ can be calculated. As $\bar{T}$ and N (the number of cells to be estimated in the O-D matrix) are constants, ERDM can be easily converted to the variance of the estimated O-D from the true mean VAR($\bar{T}$). Thus,

$$VAR(\bar{T}) = [\sum_i \sum_j (t_{ij} - \bar{T})^2] / (N - 1)$$

$$= ERDM * [\bar{T} / (N - 1)]. \qquad (10)$$

Both of the ERDM and VAR($\bar{T}$) can be used as the selection index; the resulting link priority would be similiar. On the basis of the estimated t_{ij} from Step 5 and subseqently the estimated V^k, the link priority and the required sample size can be determined by the following link selection process.

a) Firstly, the initial surveyed links (used in Step 5) are excluded from the selection process. However, they are used to calculate the estimated t_{ij} and the corresponding ERDM, which will be used as the cut-off point for determining the sample size required.

b) A systematic "forward" selection procedure is then used for sorting the priority order of the uncounted links, where the first link selected for inclusion is the one which produces the estimated O-D matrix with maximum ERDM. The forward selection continues to determine which succeeding link (combining with the previous selected links) to be entered into the O-D estimation model in such a way as to maximize the ERDM.

c) If the maximum ERDM found in Step (b) is less than the cut-off point obtained in Step (a), the forward selection can be stopped and those selected links will be the additional count locations required for improving the O-D estimation.

The convergence of the selection process can be easily seen, as the same estimated t_{ij} will be obtained if the link flows of the uncounted links are entered into the O-D estimation model. On the other hand, the less the number of uncounted links in the network, the higher accuracy of the estimated t_{ij}. In this case, the maximum ERDM obtained in Step (b) would not be greater than the cut-off point. Hence, no additional link is required for observation.

The GLS model should not be used in the link selection process as the prior matrix affects the ERDM results. Thus, the effect of a particular link on the ERDM cannot be seen explicitly. In view of this, no prior matrix should be used in the link selection process. However, the GLS model doesn't work without a prior matrix.

In order to determine the estimation reliability of the solution, independent data should be used to check the estimation accuracy. Based on the proposed sampling procedure, the turning movements at the selected links can be available for this purpose. Sometimes, a partial "true" O-D matrix (T^*_{ij}) which is usually estimated by conducting a roadside interview survey at a major road may become available. This partial information should be used to guarantee the quality of the estimated O-D matrix, especially for a large study area.

As a result, a more realistic measure (χ^2) for the estimation reliability of the solution which includes two "chi-square statistics" (χ_1^2 and χ_2^2) can be defined as follows:

$$\chi_1^2 = \sum_{(i,j)\in S} \sum \{ [T^*_{ij} - t_{ij}]^2 / t_{ij} \} \quad (11)$$

and $$\chi_2^2 = \sum_{(k,l)\in L} \sum \{ [U^{kl} - V^{kl}]^2 / V^{kl} \} \quad (12)$$

where t_{ij} and V^{kl} are obtained from the O-D estimation model, S is the set of all (i,j) pairs having observed interzonal trips T^*_{ij} and L is the set of all links having observed turning movements U^{kl} from link k to link l. Since the χ_1^2 and χ_2^2 are calculated on two different classes of information, a weight factor γ can be added to the measure in which the original estimation reliability problem becomes

$$\text{Min } \chi^2 = \gamma_1\chi_1^2 + \gamma_2\chi_2^2, \quad (13)$$

where $\gamma_1 + \gamma_2 = 1$; and γ_1, γ_2 is in the range of 0 and 1 and would be dependent on the availability and accuracy of the two data sets.

It should be noted that χ^2 is always greater than zero. If χ^2 is equal to zero, it implies that the estimated O-D trips and the turning movements would be consistent with the 'true' values. With the use of χ^2 and $\gamma_1 = \gamma_2 = 1$, hypothesis tests can be used to examine whether the "Estimation Reliability" of the model solution is acceptable in terms of the confidence level of the estimation.

A cost analysis should also be made to demonstrate the trade-off between the model accuracy and amount of link flow information. Suppose that the cost for traffic counts is proportional to the number of links for which flow information is required. On this basis, a measure of effectiveness of the additional traffic count data can be defined as below:

$$\text{\% Effectiveness} = \frac{\text{Percentage reduction in } \chi^2}{\text{No. of extra link(s) with flow data}} \quad (14)$$

It is expected that the measure of effectiveness would decrease with the increasing number of observed links. Moreover, the better selection of traffic count locations, the further the reduction in the % effectiveness should be.

3. A NUMERICAL EXAMPLE

Evaluation tests were carried out to study the performance of the three models with respect to the proposed sampling method for selecting traffic count locations.

The test network is shown in Fig.1. It consists of 10 links and 6 nodes. Nodes A, B and C are both origin and destination zones,

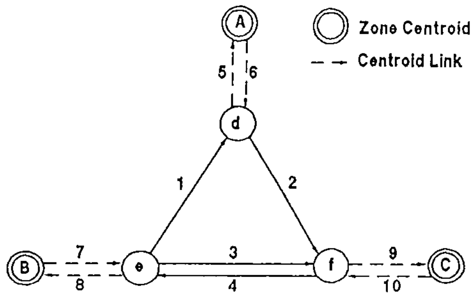

Fig. 1. Test Network

while nodes d, e and f are intersections only. Table 1 shows the "true" O-D matrix and the actual link flows. The link choice proportions by O-D pairs are given in Table 2. This example is designed for three purposes, namely (a) to illustrate the discussions and findings in section 3, (b) to test the sampling procedure based on the three O-D estimation techniques, and (c) to demonstrate the effectiveness of the proposed link selection process.

Three models for O-D estimation from traffic counts were tested : namely, generalised least square (GLS), entropy maximisation (EM), and information minimisation (IM). Those links sorted by the selection process were used as input to the O-D estimation models. It was assumed that there was no prior information on the O-D pattern and the link flow. Traffic counts were initially conducted on all the in-bound centroid links; namely, links 6,7 and 10.

The results of the ERDM are shown in Table 3. It was found that the priority order of links for observation is the same for both EM and IM models. Also, only Link 3 is identified for additional count location as its ERDM is greater than the corresponding cut-off point value. There is no link with ERDM greater than the cut-off point in the GLS results. It showes that the GLS model is not suitable for determining the additional count locations as the prior matrix would affect the ERDM due to a particular link. For the EM and IM results, the ERDM approaches to the cut-off point value as increasing the number of links with estimated flow data. It demonstrates the convergence of the proposed selection process with the use of ERDM.

The estimation reliability of each model, with the set of observed links sorted by the selection process, are displayed in Table 4. The χ_1^2 statistic for the O-D estimation was calculcated by eqn (12) using the assumed "true" O-D matrix, while the χ_2^2 statistic

TABLE 1 TRUE O-D MATRIX AND LINK FLOWS FOR TEST NETWORK

O-D Pair	O-D Matrix	Node Pair	Link Number	Link Flows
A-B	2.53	e-d	1	11.58
A-C	3.04	d-f	2	5.57
B-A	6.32	e-f	3	21.40
B-C	21.40	f-e	4	14.00
C-A	5.26	d-A	5	11.58
C-B	6.21	A-d	6	5.50
		B-e	7	27.72
		e-B	8	8.74
		f-C	9	24.44
		C-f	10	11.47

TABLE 2 LINK CHOICE PROPORTIONS BY O-D PAIRS

Link Number	O-D Pairs A-B	A-C	B-A	B-C	C-A	C-B
1	0	0	1	0	1	0
2	1	1	0	0	0	0
3	0	0	0	1	0	0
4	1	0	0	0	1	1
5	0	0	1	0	1	0
6	1	1	0	0	0	0
7	0	0	1	1	0	0
8	1	0	0	0	0	1
9	0	1	0	1	0	0
10	0	0	0	0	1	1

TABLE 3 TEST RESULTS OF THE ESTIMATED RELATIVE DEVIATION FROM THE MEAN (ERDM)

Links with Estimated Flows	Estimated Relative Deviation from Mean (ERDM) GLS model	EM model	IM model
3	5.29	32.94	32.94
3,1	6.72	12.42	12.79
3,1,4	12.39	13.37	13.35
3,1,4,9	14.35	16.31	16.72
3,1,4,9,8	15.92	17.06	17.32
3,1,4,9,8,2	16.67	17.70	17.68
3,1,4,9,8,2,5	16.74	17.70	17.70
Cut-off point	17.40	17.70	17.70

TABLE 4 CHI-SQUARE STATISTICS OF ESTIMATED O-D MATRIX (χ_1^2) AND TURNING MOVEMENTS (χ_2^2)

Observed Links	Model Type	Chi-square Statistics χ_1^2	P_1	χ_2^2	P_2
6,7,10	GLS	8.35	17.25	8.23	4.32
	EM	8.33	17.45	8.20	4.36
	IM	8.67	17.58	8.20	4.36
6,7,10	GLS	0.13	>99.5	0.01	>99.5
3	EM	0.13	>99.5	0.00	>99.5
	IM	0.68	98.30	0.00	>99.5
6,7,10	GLS	0.06	>99.5	0.01	>99.5
3,1	EM	0.05	>99.5	0.00	>99.5
	IM	0.37	>99.5	0.00	>99.5
6,7,10	GLS	0.02	>99.5	0.01	>99.5
3,1,4	EM	0.00	>99.5	0.00	>99.5
	IM	0.00	>99.5	0.00	>99.5

Note : P1 and P2 are the confidence levels of the estimated O-D and turning movements respectively.

TABLE 5 CHI-SQUARE STATISTICS OF ALTERNATIVE INITIAL SURVEY LINKS

Initial Surveyed Links	Supplementary Observed Links	Chi-square statistic of both O-D & turning movement estimation (χ^2) GLS	EM	IM
6,7,10	--	16.58	16.53	16.88
Case (I)	3	0.15	0.13	0.68
	3,1	0.07	0.05	0.37
	3,1,4	0.03	0.00	0.00
5,8,9	--	26.02	25.95	26.67
Case (O)	3	3.18	3.68	4.84
	3,7	3.12	3.66	4.86
	3,7,4	3.53	3.59	5.03

TABLE 6 EFFECTIVENESS OF SELECTED SETS OF OBSERVED LINKS

Initial Surveyed Links	Supplementary Observed Links	Percentage of Effectiveness of the Additional Link Flow Data (%) GLS	EM	IM
6,7,10	3	99.10	99.21	95.97
Case (I)	3,1	0.48	0.48	1.84
	3,1,4	0.24	0.30	2.19
5,8,9	3	87.78	85.82	81.85
Case (O)	3,7	0.23	0.08	-0.07
	3,7,4	-1.58	0.27	-0.64

was calculated on the turning movements at the observed links only. P_1 and P_2 are the confidence levels of the estimated O-D and turning movements respectively.

In general, both χ_1^2 and χ_2^2 decrease as the number of observed links increases. It was found that the confidence levels of the estimated O-D and of turning movements reach to 99.5 percent or above after Link 3 is included. It indicates that a better O-D estimate can be obtained if the best locations for traffic counts are identifed.

There are two sets of alternative initial surveyed links; namely, (I) in-bound and (O) out-bound centroid links. The chi-square statistics and the effectiveness of these alternative starting-points are illustrated in Tables 5 and 6, respectively. The χ^2 is calculated by assuming that complete information on the 'true' O-D is available and the 'observed' turning movements on the selected links are provided. The highest percentage of effectiveness was found in the 4th link selected by the proposed selection process for the case starting with out-bound trip ends, because the intial O-D estimate of Case (I) is better than that of Case (O). With 50 percent coverage (i.e 5 links selected out of 10), the largest χ^2 was only 0.37 for the case (I) and was 4.86 for the case (O).

In addition, the results shown in Table 6 indicated that although the accuracy of the O-D estimation is improving with the increasing percentage of link coverage, the rates of effectiveness are not strictly decreasing especially for case (O) where the link priority was sorted on the basis of poor initial O-D estimates. In other words, the poor the initial O-D estimates, the less accuracy of the resulting link priority particularly for those links with ERDM less than the cut-off point.

4. CONCLUSIONS

The performance of the O-D estimation models is affected by three sources of link flow information, namely, the locations of the initial surveyed links, the priority order of the selected links and the percentage of links counted.

In this paper, we have introduced a selection index (ERDM) and a systematic sampling procedure for choosing traffic count locations, which involves two iterative parts : (1) estimation of O-D matrix with supplementary count data; and (2) determination of additionaal count locations by a link selection process. The GLS model is suggested to be used in the part (1) as it is computationally efficient and is capable to take into account the error variances of supplementary traffic count data, while only EM or IM model should be used in the link selection process when they can examine explicitly the effect of a particular link on the ERDM.

Moreover, an estimation reliability index (χ^2) from both O-D and turning movement information is defined. With the use of this

Chi-square statistic χ^2, standard hypothesis test can be used to examine whether the "Estimation Reliability" of the model solution is acceptable with respect to a particular confidence level.

The proposed systematic sampling procedure for selecting links with flow information is proven to be efficient even when there is no prior O-D and link flow information available. For the purpose of the verification, the proposed sampling method will be applied to the real life situation such as in Shenzhen, a special economic zone in China. Research is currently under way attempting to investigate the effects of the variability of link choice proportions on the selection of an optimal set of links for observation, in order to improve the O-D estimation from traffic counts.

BIBLIOGRAPHY

Van Zuylen, H.J. and Willumsen, L.G. (1980) : "The most likely trip matrix estimated from traffic counts". Transportation Research, Vol. 14B, No. 13, pp. 218 - 293.

Maher, M.J. (1983) : "Inferences on trip matrices from observations on link volumes : A Bayesian statistical approach". Transportation Research, Vol. 17B, No. 6, pp. 435 - 447.

Han, A.F. and Sullivan, E.C. (1983) : "Trip table synthesis for CBD networks : evaluation of the LINKOD model". Transportation Research Record 944, TRB, National Research Council, Washington D.C., pp. 106 - 112.

Cascetta, E. (1984) : "Estimation of trip matrices from traffic counts and survey data : A generalized least squares estimator". Transportation Research, Vol. 18B, No. 4/5, pp. 289 - 299.

Cascetta, E. and Nguyen, S (1988) : "A unified framework for estimating or updating origin-destination matrices from traffic counts". Transportation Research, Vol.22B, No.6, pp. 437 - 455.

Logie, M. and Hynd, A. (1990) : "MVESTM matrix estimation". Traffic Engineering and Control, Vol. 31, No. 8/9, pp. 454 - 459; and Vol. 31, No. 10, pp. 534 - 549.

Lam, W. H. K. and Lo, H. P. (1990) : "Accuracy of O-D estimates from traffic counts". Traffic Engineering and Control, Vol. 31, No. 6, pp. 358 - 367.

Bell, M.G.H. (1991) : "The estimation of origin-destination matrices by constrained generalised least squares". Transportation Research, Vol. 25B, No. 1, pp. 13 - 22.

Inaudi, D.; Marquis, G.; Ploss, G.; Sykora, R; Cremer, M.; Lange, J.; Maher, M. and Bell, M. (1991) : "OD information vs traffic control". W.P. 421 and W.P. 431. Report for DRIVE Project V1047, Commission of the European Communities Directorate General XIII.

Yang, H.; Iida, Y. and Sasaki, T. (1991) : " An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts". Transportation Research, Vol. 25B, No. 5, pp. 351- 363.