A STUDY ON ESTIMATION METHOD OF OD TABLE BY TRAFFIC COUNT DATA AND ON-STREET QUESTIONNAIRE SURVEY

Tetsuro HYODO
Associate Professor
Information Engineering & Logistics
Tokyo University of Mercantile Marine
Etchujima, Koto-ku, Tokyo
135-8533 Japan
Fax: +81-3-5620-6492
E-mail: hyodo@ipc.tosho-u.ac.jp

Yoji TAKAHASHI
Professor
Information Engineering & Logistics
Tokyo University of Mercantile Marine
Etchujima, Koto-ku, Tokyo
135-8533 Japan
Fax: +81-3-5620-6492
E-mail: takaha@ipc.tosho-u.ac.jp

Takehiko TSUBOI
Kanto Branch Office
Toshiba Logistics Corporation
242-2 Toyofuta-minami-okinahara,
Kashiwa-city, Chiba Japan
Fax: + 81-471-45-2131
E-mail: Takehiko.Tsuboi@tl.dx.toshiba.co.jp

Abstract: An efficient and low-cost transportation survey is required for future transportation planning in developing counties. However, almost all the developed OD estimation methods need some prior OD-table information such as parameters of the Gravity model or OD-table estimated by small samples. In this study, we propose an OD estimation method which only uses on-street survey. It combines traffic count data and questionnaire data about sampled drivers' OD information. Solutions are derived by simple matrix algebra. A case study using a transportation survey conducted in Vietnam is examined. The estimation results clarify some features of the method and its limitation. Finally, further topics to be studied are discussed.

Key Words: Estimation of OD table, Traffic survey, Design of transportation survey

1. BACKGROUND AND OUTLINE

In developing countries, lack of accurate traffic OD data is one of the serious constraints for future transportation planning. Thus, efficient and low-cost survey and OD estimation method are required. A lot of researches that estimate OD by prior information have been developed (i.e. Maher(1983), Cascetta(1984), Lo *et al.*(1996) or Lo *et al.*(1999)). However, they need parameters of Gravity model or past OD survey. On the other hand, estimation methodologies without prior information have not been well discussed. In this paper, we examine OD estimation method using on-street traffic count data and questionnaire data on each car's OD. The basic idea is to enlarge the OD patterns asked to sampled drivers by traffic volume at the survey points. But this simple method has several problems as follows:

1) Same drivers may answer his/her OD at several survey points ("duplicate problem").
2) Enlarging method considering the above duplicate data has not been developed.
3) The differences of sampling rate between survey points should be considered.

We propose an equation to estimate OD table from on-street survey. The equation is based on the least squares, and it is described as simple matrix algebra. But the method has several problems. First, the survey does not ask the route of each sample so that the route should be identified by "minimum path" condition. Second, the answer does not satisfy the "non-negative condition". Therefore we should introduce a constraint condition for the estimation.

To test the applicability and limitation of the proposed method, we use actual survey data. The examined data is traffic volume and on-street survey in Vietnam in 1998. The survey was conducted for the nationwide transportation network master plan in Vietnam by JICA. We calculated the OD table by using the proposed method and the results showed the features of the method. For example, it was recognized that the "non-negative condition" does not affect the estimation result. We also compared the estimated OD table with the authorized OD table calculated by comprehensive data, and we clarified the advatages of the proposed method. Finally, we summarize the applicability of the method, and recommend some topics to be studied.

2. ESTIMATTION METHOD OF OD-TABLE BY ON-STREET QUESTIONNAIRE SURVEY

2.1 Issues of OD-table estimation methods

The principal method to estimate accurate OD-table is to calculate it by home-based travel surveys. However, the home-based survey costs too much and accurate population data is also necessary for the estimation. One of the ways to estimate OD-table at low cost is to combine an on-street questionnaire survey and observed traffic volume at each "survey point", hereafter we call it "station". The on-street questionnaire asks the origin & destination of sampled drivers and the observed traffic volume at each station can give the enlarge coefficient. The well-examined distribution pattern of the station's location might give reasonable OD-table. But this simple method has several problems as follows:

1) Same drivers may answer his/her OD at several survey points ("duplicate problem").
2) Enlarging method considering the above duplicate data has not been developed.
3) The differences of sampling rate between survey points should be considered.

If the questionnaire involves a question to identify whether he/she had answered the survey at other stations, we could solve the problem by omitting the sample. However, a nation-wide OD survey can not be conducted on a same day, because the trip length is very long and sometimes overnight trips exist. This means the above additional question is not useful and some methods to avoid "duplicate problem" should be developed.

2.2 Basic idea of proposed estimation method

We examine a simple method to estimate OD-table by on-street questionnaire and traffic count data. The basic concept is based on minimizing the least squares of the differences between the results of questionnaire and traffic volume at each station.

Here, we introduce the equations to estimate an OD-pair, therefore the subscript of the OD-pair, "$i\,j$", is omitted. We assume that the OD-pair has R routes. The total volume of the OD-pair (t) is calculated by the summation of R route's volume.

$$t = \sum_{r=1}^{R} t_r \qquad (1)$$

Then we define a dummy variable δ_{lr} as follows.

$$\delta_{lr} = 1, \text{ if link } l \text{ is part of route } r \\ = 0, \text{ otherwise} \tag{2}$$

And q_l is the observed traffic volume at a station in link l. The objective function is to minimize the differences between the summation of trip volume of routes and the observed traffic volume at each station (or each link). We apply the least squares equation as the objective function.

$$S = \sum_l \left(\sum_r \delta_{lr} t_r - q_l\right)^2 \rightarrow \min_{t_r} \tag{3}$$

The unknown variable is t_r, so the minimum point of the equation (3) is given as follows.

$$\partial S / \partial t_r = 0 \quad \forall r \tag{4}$$

The solution of the equation (4) is written as the following matrix formulation.

$$\begin{bmatrix} t_1 \\ \vdots \\ t_r \\ \vdots \\ t_R \end{bmatrix} = \begin{bmatrix} \sum_l \delta_{l1}\delta_{l1} & & \cdots & & \sum_l \delta_{l1}\delta_{lR} \\ & \ddots & & & \\ \vdots & & \sum_l \delta_{lr}\delta_{lr'} & & \vdots \\ & & & \ddots & \\ \sum_l \delta_{lR}\delta_{l1} & & \cdots & & \sum_l \delta_{lR}\delta_{lR} \end{bmatrix}^{-1} \begin{bmatrix} \sum_l \delta_{l1} q_l \\ \vdots \\ \sum_l \delta_{lr} q_l \\ \vdots \\ \sum_l \delta_{lR} q_l \end{bmatrix} \tag{5}$$

Where the link non-observed the OD-pair ($q_l = 0$) is excluded. This simple method can derive the solution ($\hat{t}_r$) only by calculation of inverse matrix. But there are two major problems.

1) $\hat{t}_r$ does not always satisfy "non-negative condition" ($\hat{t}_r \geq 0$).
2) Some algorithms to identify each route should be explored.

Here we assume that the on-street questionnaire does not include the information about drivers' route, because it is very difficult to record all drivers' route. The survey is conducted on the way to their destination so that all drivers may not have accurate information from the station to their destination.

If the appropriate route identification algorithm could be developed, the former problem would be avoided. However, we do not have enough information about the route choice

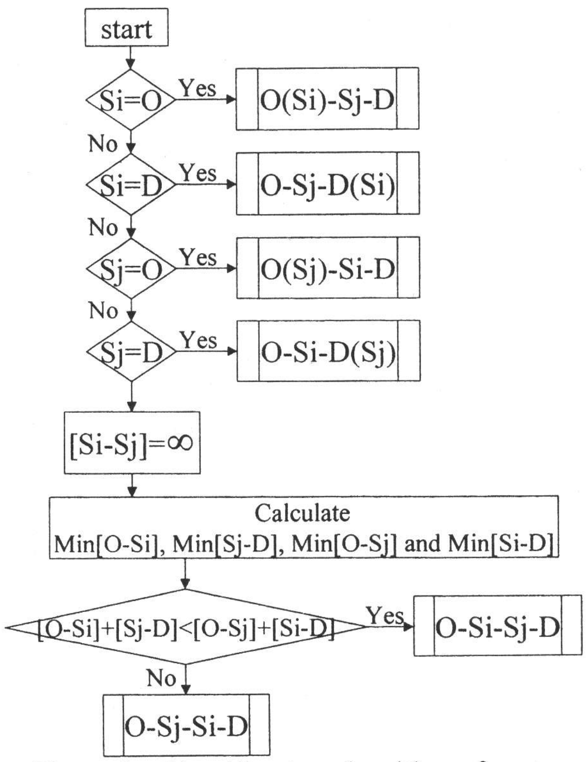
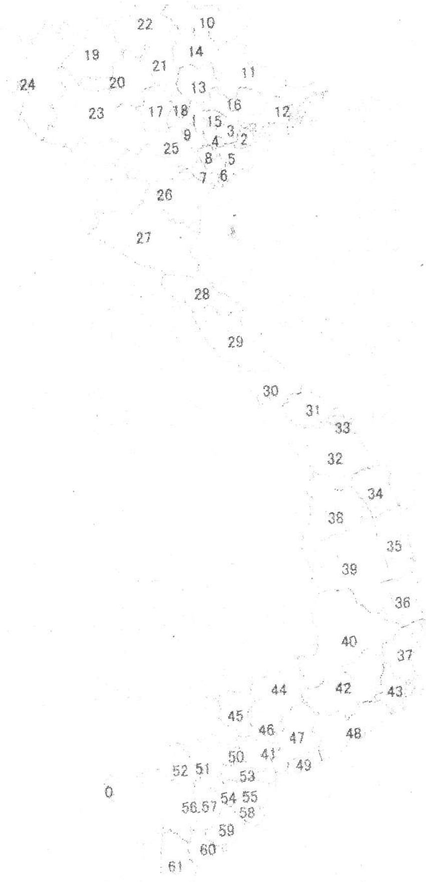


Figure 1 Identification algorithm of route
[O:origin, D:destination,
Si, Sj: both nodes of a station's link]

behavior, especially in developing countries. One of the methods to identify each route is to assume that all drivers may choose minimum travel time route or minimum distance route. In this paper, we apply this "minimum path condition". The algorithm is shown in Figure 1. Our surveyed OD-data at each station does not include direction information. For example, if an OD pair, "O to D", was surveyed at a station which locates between node Si and Sj, we do not identify the actual path was "O -> Si -> Sj -> D" or "O -> Sj -> Si -> D". Therefore, some algorithm to discriminate the reasonable path should be required. One of the ways is to compare the distance of two paths and identify the shorter one as actual path. This algorithm is shown in Figure 1.

2.3 Outline of analyzed data

In this paper, travel surveys in Vietnam is examined as an example of the proposed method. The nationwide survey was conducted in 1998-1999 for "The Study on The National Transport Development Strategy in The Socialist Republic of Vietnam (VITRANSS)". VITRANSS was a study by the Japan International Cooperation Agency (JICA) and Ministry of Transport of Vietnam to suggest log-term transportation development strategies. In the study, the country was divided into 61 zones (Figure 2) and 39 stations were located. Traffic count survey and questionnaire survey, and sampled data were conducted to estimate the OD-table between 61 zones. The survey covered all modes; passenger car, truck, ship. But



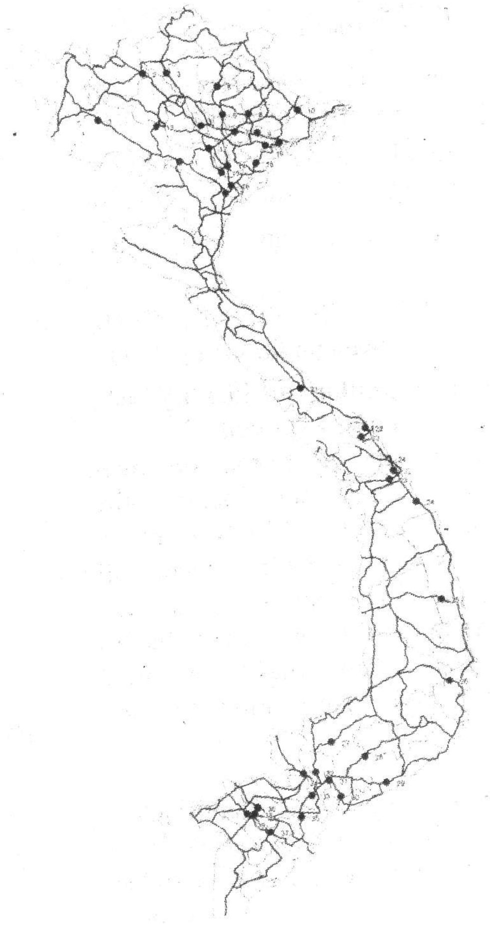


Figure 2 Zoning in VITRANSS study Figure 3 Road network and Location of stations

in this study, we only analyze "car trip" consisting of passenger cars and trucks.

We treat the VITRANSS survey result to examine the proposed OD estimation method, however, the actual VITRANSS study did not use the method. Careful heuristic process was applied to the estimation, and our study was independent of VITRANSS results.

3. RESULTS OF IDENTIFICATION ALGORITHM OF ROUTE

The number of OD-pairs in VITRANSS data is 3,261. This number includes the samples which have duplicated routes. Therefore, the identification algorithm of route (Figure 1) should be applied.

Figure 4 shows one of the results of the route identification. "Detour rate" means the distance of estimated route divided by the minimum distance between each observed sample. A detour rate of about 6% of sampled OD-pairs is over 3, which may be an unreasonable number. Example of typical unreasonable and reasonable results are shown in Figure 5 and 6. In Figure 5, the station of the sampled data did not locate between the answered origin and destination. The reason for this may be imagined as follows;

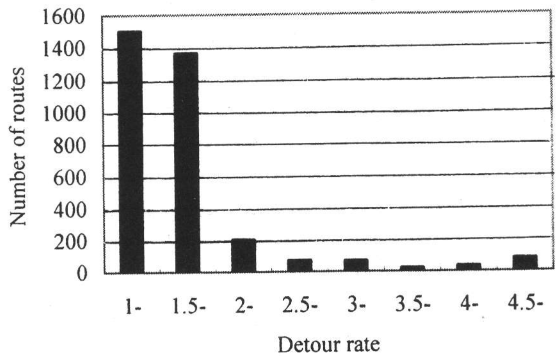


Figure 4 Frequency of detour rate

1) The driver did not answer his/her true origin or destination.
2) The answered OD was the trip chain OD of the day, and the trip observed at the station was one of the chained trips.
3) The hypothesized zone center is far from the true origin/destination point.
4) Some mistakes in the process of recording data.

To avoid reason 2) above, a careful questionnaire survey should be designed. Some process

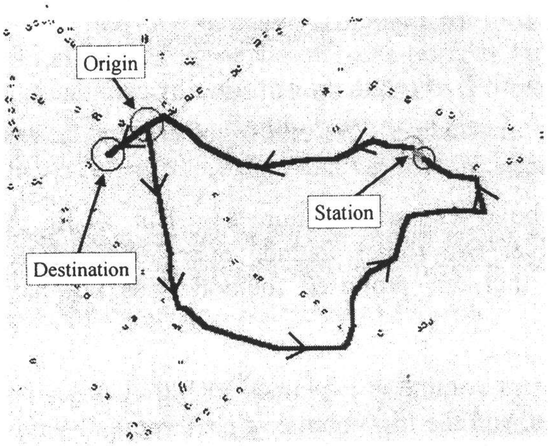


Figure 5 Example of unreasonable result

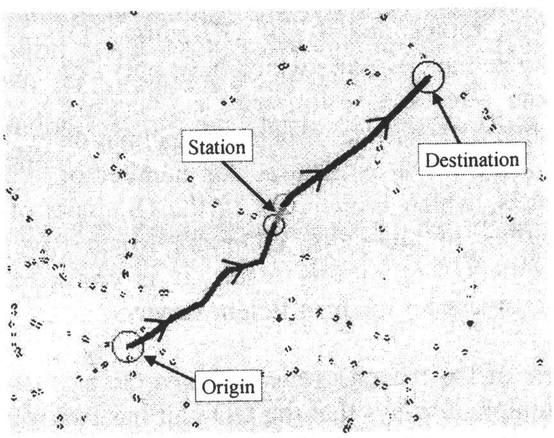


Figure 6 Example of reasonable result

which can omit the unreasonable answers should be introduced in the questionnaire. Reason 3) above is based on the accuracy of demand analysis system. More precise zoning could reduce it, but the burden of handling spatial data increases. This problem may relatively occur at short distance trips so that some heuristic process should be required especially for the short trips.

The identified routes still include duplicated samples. The "duplicated samples" mean that the samples observed at different stations have same origin and destination. By checking each sample's route with other samples' route, the duplicated routes are merged. Consequently, the 3,261 OD-pairs were reduced to 1,720 OD-pairs (Table 1).

Table 1 The number of OD-pairs by merging duplicated samples

Number of routes	Number of OD-pairs
1	980
2	452
3	187
4	62
5	27
6	12
Total	1,720

4. ESTIMATION OF OD-TABLE BY MINIMIZING LEAST SQUARES

The trip volume of each OD-pair is estimated by the least squares method, given in Equation (5). As mentioned in subsection 3.2, this method does not have non-negative condition, so that the results may involve some negative number. Figure 7 shows the frequency of estimated route's volume. Almost all the volume of each route is positive. However, there are small negative results.

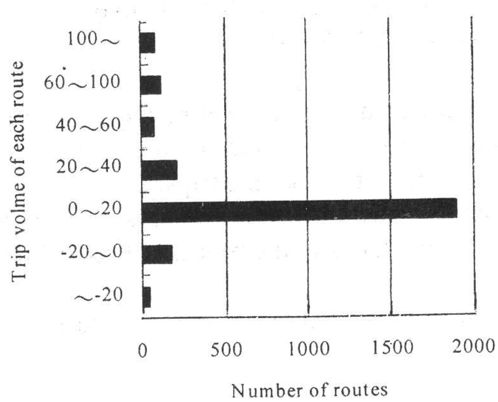


Figure 7 Frequency of routes by estimated trip volume

The data for the estimation consists of 2,900 routes and 1,720 OD-pairs. Of course, the OD-pairs which have only one route are not estimated as negative number. The rate of negative estimation is about 18% to 22% as for number of routes, which is included in the OD-pairs of above one route (Figure 8). But, as for the number of OD-pairs, OD-pairs which have over two routes include over 50% negative numbers. From this result, it is considered that our proposed method does not have robustness for the insufficient surveys.

One of the reasons which derives the negative trip volume is explained in Figure 10. The example displays that the OD-pair has two routes, and the trip volume of each route observed is 60 (route 1) and 10 (route 2). But at the station which contains the two routes, the observed volume is very small compared to the summation of each route's observation

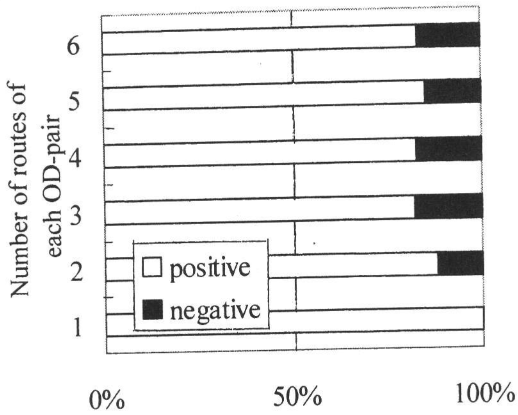


Figure 8 Proportion of negative estimations [proportion as for number of routes]

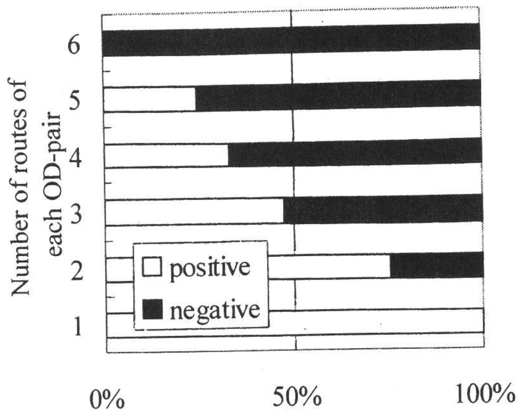


Figure 9 Proportion of negative estimations [proportion as for number of OD-pairs]

($q_3 = 10 \leq q_1 + q_2 \fallingdotseq 70$). If we apply the least squares method (Equation (5)) to this pattern, one of the route's volume is estimated as a negative number as follows.

$$\begin{bmatrix} t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 60+10 \\ 10+10 \end{bmatrix} = \begin{bmatrix} 40 \\ -10 \end{bmatrix} \tag{6}$$

The complicated routes and stations structure tends to generate this situation, therefore, the rate of negative results is relative to the number of routes of each OD-pair as shown in Figure-9.

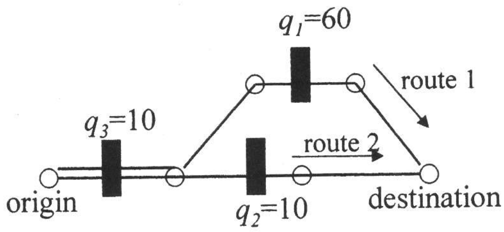


Figure 10 An example generating negative result

Discussion for further analyses

One of the ways to avoid non-negative numbers is to conduct an accurate survey, and design a careful questionnaire as mentioned. Here, we would discuss the other technical methods.

To satisfy the non-negative condition by force, non-linear optimization with non-negative condition may be applied. Recent popular optimizer could easily solve this problem. For example, if we solve the former example (Figure 10), the solution is given as $t_1 = 36.7,\ t_2 = 0$. However, this method offers only a plausible solution. It can not clarify the fundamental problem such as the errors of survey and questionnaire.

Another technical method is to introduce an accuracy of sampling at each station. If we assume that the observed traffic volume of the r-th route at the l-th station (q_{lr}) is distributed normal, the dispersion is defined as follows.

$$\sigma_{lr}^2 = \frac{Q_l^2 q_{lr}}{q_l}\left(1 - \frac{q_{lr}}{q_l}\right) \tag{7}$$

Where, Q_l is the total observed traffic volume at the l-th station, q_l is the number of drivers who answered the questionnaire ($\sum_r q_{lr} = q_l$). From the logarithmic likelihood function, the following equation is derived as well as equation (3).

$$L = \sum_l \left(\frac{\sum_r \delta_{lr} t_r - q_l}{2\sigma_{lr}^2} \right)^2 \to \min_{t_r} \qquad (8)$$

The first order derivation as for equation (8) gives

$$\begin{bmatrix} t_1 \\ \vdots \\ t_r \\ \vdots \\ t_R \end{bmatrix} = \begin{bmatrix} \sum_l \delta_{l1}\delta_{l1}/\sigma_{l1}^2 & \cdots & \sum_l \delta_{l1}\delta_{lR}/\sigma_{l1}^2 \\ \vdots & \sum_l \delta_{lr}\delta_{lr'}/\sigma_{lr}^2 & \vdots \\ \sum_l \delta_{lR}\delta_{l1}/\sigma_{lR}^2 & \cdots & \sum_l \delta_{lR}\delta_{lR}/\sigma_{lR}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_l \delta_{l1} q_l/\sigma_{l1}^2 \\ \vdots \\ \sum_l \delta_{lr} q_l/\sigma_{lr}^2 \\ \vdots \\ \sum_l \delta_{lR} q_l/\sigma_{lR}^2 \end{bmatrix}. \qquad (9)$$

This method can consider that the difference of accuracy, and sampling rate between stations. If the "negative number problem" caused some inaccurate observations, it might contribute to the generation of negative numbers. Actually, we applied this method to the VITRANSS data. However, the numbers of negative ones do not decrease. As for our data, more fundamental process should be considered.

Our proposed method needs an appropriate route identification algorithm before applying the least square estimation. In this paper, we assume drivers choose the minimum distance path. If there are many traffic congestion points in contrast to Vietnam, the minimum travel time condition which considers congestion or grade of road should be adopted. Basically, both minimum distance and minimum travel time are special case of route choice model. Conventional route choice model such as the Logit model ought to be examined. In any case, the best way to get accurate route information is to ask drivers his/her actual route directly as mentioned in Section 3.

5. CONCLUDING REMARKS

A new method for OD-table estimation by using on-street survey is examined in this paper. It was clarified that the method requires sufficient accuracy; especially, validity of route specification is sensitive to the estimation result. If the survey does not gather actual route information, appropriate route choice model (including minimum distance or travel time) should be developed to improve accuracy of route identification algorithm.

The case study on the nationwide travel survey in Vietnam could not arrive at reasonable results. The number of estimated negative trip volume is not negligible, and we do not have enough information to improve it. However, the proposed method would give important insights into the estimation of OD-table. If there are experts who have sufficient experience and knowledge for drivers' behavior, the results may be improved easily and the method can save their effort. This method may not give accurate answers, otherwise, it can assist to get OD-table easily under the condition of few survey data.

Because of the limitation of examined data, we can not study the method in detail. More simple OD patterns should be tested to clarify the feature of it. And the simulation study may discuss location of survey points and appropriate sample size; the distribution and the number of stations for efficient survey.

Acknowledgements

The authors wish to thank JICA for our usage of VITRANSS data.

References

Cascetta, E. (1984) **Estimation of Trip Matrices from Traffic Counts and Survey Data: A Generalized Least Squares Estimator**, Transportation Research B, Vol.18B, No.4/5, pp.289-299

Lo, H.P., Zhang, N. and Lam, W.H.K. (1996) **Estimation of An Origin-Destination Matrix with Random Link Choice Proportions: A Statistical Approach**, Transportation Research B, Vol.30B, No.4, pp.309-324

Lo, H.P., Zhang, N. and Lam, W.H.K. (1999) **Decomposition Algorithm for Statistical Estimation of OD Matrix with Random Link Choice Proportions from Traffic Counts**, Transportation Research B, Vol.33B, pp.369-385

Maher, M.J. (1983) **Inferences on Trip Matrices from Observations on Link Volumes: A Bayesian Statistical Approach**, Transportation Research B, Vol.17B, No.6, pp.435-447