# A TRAFFIC DEMAND ESTIMATION MODEL BASED ON LINK TRAFFIC COUNTS CONSIDERING THE INTERACTION OF THE TRAFFIC MODES


| Shogo KAWAKAMI | Huapu LU | Yasuhiro HIROBATA |
|---|---|---|
| Professor | Graduate student | Associate professor |
| Nagoya University | Nagoya University | Toyohashi Univ. of Technology |
| Nagoya-Japan | Nagoya-Japan | Toyohashi-Japan |


## INTRODUCTION

It is necessary to estimate the O-D matrices when a transportation network analysis is carried out. The method commonly used to get O-D matrices is the traffic survey, i.e., home or roadside interviews. However, conducting a large-scale traffic survey is very expensive, time consuming and labor intensive. Moreover, sampling errors of survey data also occur. On the other hand, link traffic count data is easier to obtain and less expensive. Furthermore, in most cases, it is collected regularly for other purposes. That is especially true in developing countries where rapid economic growth and frequent land use changes require conducting frequent O-D matrices estimations. In this case, of course, it is necessary to carry out large scale traffic surveys, but it is impossible to do so frequently. This makes the estimation of O-D matrices from link traffic counts, as a new method, very attractive.

For these reasons, much research work has been done and many models have been developed. However, further work is required to apply the O-D estimation model to practical forecasting and planning. The purpose of this research is to develop a new model in which multiple traffic modes and their interaction can be considered, and the separate outputs for different traffic modes can be obtained. Moreover, the path choice rate, which is the proportion of O-D traffic volume using the given route, can be obtained internally. This path choice rate is often required as input by most of the existing methods, thus cutting down their practicality and accuracy.

A brief review of the existing research is presented in Section 1. In Section 2, a combined distribution, modal split and traffic assignment model in the form of an optimization program is proposed. From the Kuhn-Tucker condition of the proposed optimization problem, a distribution demand model is derived. Then, the uniqueness and the existence of a solution are demonstrated by checking Hessian matrices of the objective function and of the constraint conditions with respect to the O-D variables.

In Section 3, the proposed model is applied to a Nagoya road network as a case study. In order to consider the interaction of the traffic modes, a revised version of BPR function as a generalized cost function are introduced and sensitivities of parameters are also analyzed. The O-D matrices for different modes are estimated by iterative calculation of given traffic volume generated in each zone as well as the observed link traffic counts.

A comparison between the observed and the estimated O-D matrices is carried out and the accuracy of the estimation is discussed in Section 4. Moreover, in order to study the property of the proposed model, a comparison analysis of the estimation results obtained by using using the proposed model and the existing methods are performed.

## 1. REVIEW OF EXISTING RESEARCH WORKS

It is well known that estimation of O-D matrices from link traffic counts is a underdefined problem. We can not estimate the O-D matrices using only link traffic counts because information contained in the link traffic counts is insufficient, thus some other information or constraints must be added in order to find unique O-D matrices. According to the method of adding additional information, the existing research can be divided into two main groups. One of them uses prior information given about distribution patterns of O-D traffic volume, and an O-D matrix which is the nearest one to the given O-D matrix can be found. For example, giving a target O-D matrix or expressing the O-D distribution pattern by a gravity model belongs to this group. The other group is that in which a distribution pattern is not assumed and the O-D matrix is estimated directly based on entropy maximization or information minimization principles.

The representative models can be summarized as follows: first is maximum likelihood model proposed by J. Holm et al.(1976), Henk. J van Zuylen etc(1982), Heinz Spiess(1987), and so on. Holm et al. assume that link flows are random and independent variables of each other, and they estimate the O-D matrix by maximizing the likelihood.

Second is least square method and regression model proposed by P. Hogberg(1976), Sue McNeil and Chris. Hendrickson(1985), Takayama and Iida (1987), M. G. H. Bell (1991) and so on. Takayama and Iida take generation traffic volume as a unknown variables and find O-D matrix from observed link counts by minimizing the square errors between the estimated and observed link flows.

Third is the programming method proposed by Malachy Carey et al.(1981), Maud Brenninger-Gothe etc(1989). Malachy Carey et al. formulates a quadratic programming method to estimate matrix entry as an equivalent constrained generalized least square estimation problem.

Fourth is the entropy maximization and information minimization method proposed by Van Zuylen, Henk, J. and Willumsen L. G.(1980) and so on. Hens J. Van Zuylen proposed an information minimizing approach. The idea behind the model is that since the information available in the link traffic counts is insufficient to determine a complete trip matrix, it seems reasonable to choose a trip matrix that adds as little information as possible to the knowledge contained in conservation condition of links. He defines the information contained in a set of N observations using Brillouin's information measure. Using the Lagrangian multiplier method for objective function and constraint conditions, a multi-proportional model is finally obtained. The O-D matrix can be found by solving this model.

The last one is the network equilibrium or combined model proposed by S. Erlander etc.(1979), LeBlanc, L. J., Farhangian, K.(1982), Fisk, C.S. Boyce, D.E.(1983), Kawakami S., Hirobata Y. and Lu H.(1989) and so on. Nguyen et al. introduced the network equilibrium concept to estimate O-D matrices from link traffic counts data and proposed an excellent method. He proves that the optimal solution obtained by solving the model proposed by him does reproduce observed flows. However, since the objective function is not strictly convex with respect to the variables $T_{ij}$, there can be more than one set of optimal trip variables from the model. In order to overcome this shortcoming, Gur et al. suggested a method using an externally generated trip table, or target trip table, to provide information on the structure of the actual (i.e. unknown) trip table. A unique solution can be obtained this way. Furthermore, Fisk et al. (1983) proposed a combined distribution and assignment model to estimate the O-D matrix from link counts data. These models are superior in giving the path choice rate internally.

## 2. A COMBINED DISTRIBUTION, MODAL SPLIT AND TRAFFIC ASSIGNMENT MODEL

### 2.1. Optimization Program

When the choice behavior of the trip maker is analyzed, Wardrop's principle is most commonly used. However, it can not describe all choice behaviors by travellers over the four stages (Boyce et al. 1983). Wardrop's cost minimizing behavior would appear to be appropriate to route choice, to a lesser extent to mode choice, and hardly at all to destination and location choice. These deviations from cost minimizing behavior occurs principally in location and destination choices and to a lesser extent in mode and route choices. On the other hand, entropy can be used to measure these deviations from Wardrop's cost minimizing principle. Based on these analysis, a combined distribution, modal split and traffic assignment model in the form of a optimization model is proposed, which extends Fisk's model (Fisk et al. 1983)by introducing the entropy constraint with respect to the mode choice behavior as follows:

$$Max \quad -\sum_i \sum_j (\sum_m \sum_r P_{ijmr}) \ln(\sum_m \sum_r P_{ijmr}) \tag{1}$$

$$s.t \quad \frac{1}{T} \sum_m \sum_a \int_0^{V_a^m} S_a^m(x) dx = \hat{C} \tag{2}$$

$$V_a^m = \sum_i \sum_j \sum_r P_{ijmr} \delta_{ijmr}^a T \quad \forall \quad a, m \tag{3}$$

$$\sum_j \sum_m \sum_r P_{ijmr} = \bar{P}_i \quad \forall \quad i \tag{4}$$

$$-\sum_i \sum_j \sum_m (\sum_r P_{ijmr}) \ln(\sum_r P_{ijmr}) \leq E_M \tag{5}$$

$$P_{ijmr} \geq 0 \quad \forall \quad i, j, m, r \tag{6}$$

where $P_{ijmr}$ refers to the proportion of the traffic flows traveling from origin $i$ to destination $j$ by mode $m$ and route $r$. $V_a^m$ is traffic flow on link $a$ by mode $m$. $S_a^m(V_a^m)$ is the generalized cost function of mode $m$ on link $a$ at traffic flow $V_a^m$. $\bar{P}_i$ is the proportion of the known traffic volume generated from each zone. $T$ is total traffic flow of the study region. $E_M$ is the given entropy constraint with respect to traffic modes. $\hat{C}$ is a constant which will be calculated from link traffic counts data and $\delta_{ijmr}^a$ is an element of the incidence matrix.

The objective function of this optimization problem (1) represents the entropy for the traffic distribution. Constraint condition (2) specifies an Beckmann type user equilibrium. Constraint condition (5) is the entropy constraint with respect to traffic modes.

This maximization problem can be rewritten into a minimization problem and the Lagrange function can be obtained. Finally, from optimal conditions, a distribution model can be obtained as follows:

$$\sum_r P_{ijmr} = \bar{P}_i \frac{\exp[-\beta(\eta,\mu)\tilde{C}_{ij}]\exp[-\mu C_{ijm}]}{\sum_j \exp[-\beta(\eta,\mu)\tilde{C}_{ij}]\sum_m \exp[-\mu C_{ijm}]} \tag{7}$$

where $\mu = \eta \,/\, \gamma$, $\beta(\eta,\mu) = \frac{\mu^2}{\eta+\mu}$, $\tilde{C}_{ij} = \frac{1}{\mu}\ln\sum_m \exp\{-\mu C_{ijm}\}$

Parameters $\gamma$ and $\eta$ are Lagrange multipliers. $C_{ijm}$ is the minimum path cost from zone $i$ to zone $j$ by mode $m$. This equation should be satisfied when the optimization program is solved.

The method of solving this optimization program is as follows: first, the constant $\hat{C}$ is found from traffic link counts data. Then the derived distribution model (equation(7)) and other optimal conditions are used to find the O-D matrices for different traffic modes and parameters $\gamma$ and $\eta$ simultaneously by iterative calculation.

### 2.2. Existence and Uniqueness of the Solution

It is well known that if the objective function of a non-linear minimization program is a concave function and if all constraint conditions of optimization program are also concave functions, then the program is called a concave program and a unique solution exists. The condition of being strictly concave for a function is equivalant to the Hessian matrix of that function being positive definite. When the maximization program is changed into the minimization program, the second-order derivative of the objective function with respect to the variable $P_{ijmr}$, which is between zero and 1, is as follows:

$$\frac{\partial^2 A}{\partial P_{ijmr}^2} = \frac{1}{\sum_m \sum_r P_{ijmr}} > 0 \tag{8}$$

where, 'A' represents the equation (1) with opposite sign. The value of the other non-diagonal cells of the Hessian matrix are zeros, $\partial^2 A / \partial P_{ijmr} \partial P_{i'j'm'r'} = 0$, so that the Hessian matrix of the objective function is positive definite. Similarly, the Hessian matrix of constraint (5) is also positive definite. If we take the second order derivative of the constraint condition (2) with respect to $P_{ijmr}$, we get

$$\frac{\partial^2 B}{\partial P_{ijmr}^2} = \frac{\partial S_a^m(x)}{\partial P_{ijmr}} \tag{9}$$

where,'B' represents the equation (2). Equation (9) is the first-order derivative of generalized cost function with respect to $P_{ijmr}$. All the non-diagonal cells of the Hessian matrix become zeros. From equation (9) , it is clear that for the diagonal cells $\partial S_a^m(x)/\partial P_{ijmr}$, whether or not they are positive depends on the form of the generalized cost functions. A suitable form can be taken to make these cells positive and a unique solution of the proposed model exists.

## 3. APPLICATION TO THE CITY OF NAGOYA, JAPAN

### 3.1. Input Data

In order to study the efficiency and applicability of the proposed model, it is applied to a Nagoya road network. Input data used by the proposed model to estimate the O-D matrices are link traffic counts data by modes and traffic volume generated in each zone. Output data consists of the estimated O-D matrices for different traffic modes and parameters of the distribution model.

The O-D matrices and link traffic counts are obtained from the Japan's O-D automobile traffic census carried out in 1985. Traffic volumes are divided into two categories. One category contains large-sized trucks and buses, another consists of small trucks and cars. For simplicity, the former is called large-sized truck and the latter is called cars. Without loss of the generality, large-sized trucks and cars are regarded as two different modes and estimation is carried out.

### 3.2. Simplified Nagoya road network

A simplified Nagoya road network is shown in Fig 1. It consists of 154 nodes and 240 links. All links are two-way and the total link number is 480. In order to use the survey data, the study region is divided into zones just as did in the survey. Nagoya city is divided into 16 zones with a centroid in each zone. Moreover, the simplified network consists of general national routes and main prefectural roads, and some

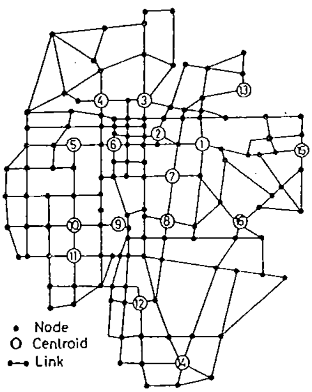

FIGURE 1. Simplified Nagoya Road Network

community roads are eliminated. Because the inner-zone traffic volume is ignored the variables to be estimated are $2 \times n \times (n-1) = 480$.

### 4.3. Generalized cost function and sensitivity analysis

In this case study, the traffic modes are taken to be two. namely, large-sized trucks and cars. For taking the interaction of modes into account, the generalized cost functions for two modes are assumed to have the following form:

$$t_{aT} = t^0_{aT}[1.0 + \alpha_T(\frac{x_{aT} + \xi x_{aC}}{C_{aT}})^{\beta_T}] \qquad \forall a \tag{10}$$

$$t_{aC} = t^0_{aC}[1.0 + \alpha_C(\frac{x_{aC} + \eta x_{aT}}{C_{aC}})^{\beta_C}] \qquad \forall a \tag{11}$$

where equation (10) is the generalized cost function for large-sized trucks, and equation (11) is for cars. $\xi$ and $\eta$ express the effect to the large-sized truck given by the cars and the effect to the cars given by the large-sized truck respectively. $t_{aT}$ and $t_{aC}$ represent the link traffic time of the large-sized truck and cars respectively. $t^0_{aT}$ and $t^0_{aC}$ represent zero-flow link traffic time of the large-sized trucks and cars respectively. $X_{aT}$ and $X_{aC}$ are link traffic volumes of the large-sized truck and car, $C_{aT}$ and $C_{aC}$ are the capacities of the link for large-sized trucks and cars respectively. $\alpha_T, \beta_T, \alpha_C$, $\beta_C, \xi$ and $\eta$ are the parameters.

In this way, the interaction of the different modes can be considered. The link travel time of large-sized trucks is not only affected by large-sized trucks but also affected by cars, and vice versa. For the time being, there is not enough the observed data to calibrate the parameters of the generalized cost function, thus they have to be determined according to the experience. In order to study the effects caused by the errors of parameters, a sensitivity analysis is carried out. The changes of the generalized cost with parameters $\alpha_C, \beta_C, \xi, \alpha_T, \beta_T$ and $\eta$ are shown in the Fig. 2,3,4,5,6 and 7. The horizontal coordinates of these figures are the values of parameters and the vertical coordinates are the values of the generalized cost. It can be seen that changes with parameters $\alpha_C, \beta_C, \xi, and\ \eta$ are relatively sensitive but the changes of the generalized cost with parameters $\alpha_T$ and $\beta_T$ are not very obvious as shown in the figures. In summary, the sensitivity of the generalized cost to parameters are not very strong. Therefore, in the case of lacking available data to calibrate the parameters, we can determine them by the experience. Referring to the analysis of Kawakami et al.(1989), the parameters of the generalized cost functions are given as follows. $\alpha_C = 0.15$, $\beta_C = 4.0$ and $\xi$ =1.5 for cars and $\alpha_T = 0.06$, $\beta_T = 4.0, \eta = 0.45$ for large-sized trucks.

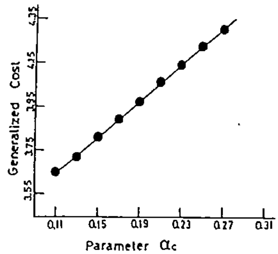


FIGURE 2. Change of Generalized Cost with Parameter $\alpha_C$

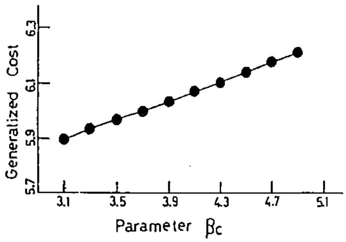


FIGURE 3. Change of Generalized Cost with Parameter $\beta_C$

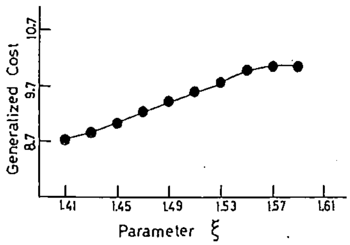


FIGURE 4. Change of Generalized Cost with Parameter $\xi$

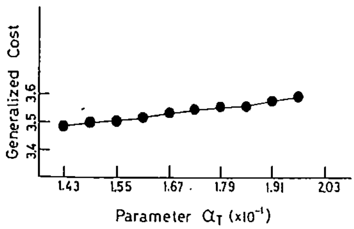


FIGURE 5. Change of Generalized Cost with Parameter $\alpha_T$

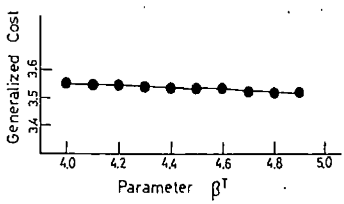


FIGURE 6. Change of Generalized Cost with Parameter $\beta_T$

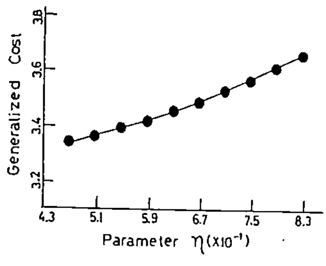


FIGURE 7. Change of Generalized Cost with Parameter $\eta$

## 4.4. Statistical measures

The statistical measures, Mean Absolute Error, Root Mean Square Error and Percent Root Mean Square Error, have been introduced to facilitate the error analysis and comparison analysis.

$$MAE\% = \frac{\sum_i \sum_j \left|\hat{t}_{ij} - t_{ij}\right|}{\sum_i \sum_j t_{ij}} \times 100\%, RMSE = \sqrt{\left(\frac{\sum_i \sum_j (\hat{t}_{ij} - t_{ij})^2}{Z}\right)}$$

$$\%RMSE = \frac{RMSE}{(\sum_i \sum_j t_{ij})/Z} \times 100\%$$

where Z is a number of non-zero cells of the observed O-D matrix. RMSE and %RMSE show the 'error dispersion' of the estimated O-D matrices.

## 4. ESTIMATED RESULTS AND COMPARISON ANALYSIS

In this study, the O-D traffic volumes for different traffic modes are estimated based on link traffic counts data as well as the traffic volume generated in each zone. A comparison between observed O-D matrices and estimated O-D matrices are carried out. These results are shown by Fig 8 and Fig 9 respectively.

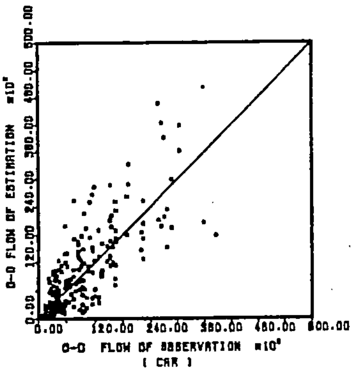


FIGURE 8. Comparison of observed and estimated O-D matrix (car)

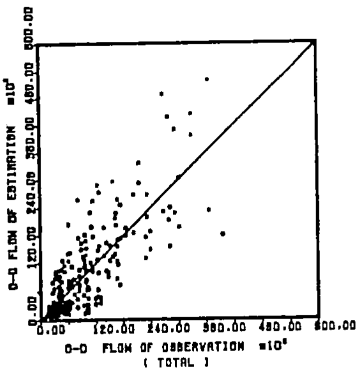


FIGURE 9. Comparison of observed and estimated O-D matrix (Total)

The horizontal and vertical coordinates of these figures represent the traffic volume, and the ratios of the estimated and observed traffic volume are represented by dots in the figures corresponding to traffic volume. Because the proportion of the large-sized trucks in total traffic volumes is very small(only 5 %), for saving space, only the figures of cars and total are given here. The correlation coefficients of the observed and estimated O-D traffic volume for cars, large-sized trucks and total are respectively $r = 0.78$, 0.59,0.79 showing that a good estimation result has been obtained. Furthermore, the parameters of the distribution model which are calibrated simultaneously are respectively $\gamma = 0.052$, $\eta = 0.158$, and $E_M$ is given as 4.845.

In order to study the property of the estimation model and accuracy of estimation, the estimated results by using the proposed model and other two existing models (Maximum likelihood model and least square method) are compared. For least square model, it is necessary to give a prior information on O-D matrices when estimation is performed. For discussing the effects caused by a prior information on O-D matrices, two extreme cases are taken into account in comparison analysis. One case(A) is assumed that planner knows prior information on O-D matrices completely and anothor case(B) is that prior information is really poor, thus we can think prior O-D information is almost nothing.

The statistical measures, correlation coëfficents, average absolute error, % RMSE and RMSE of the estimated values and observed ones for cars, large-sized trucks and total are shown in table 2 and 3.

Table 2. Correlation Coefficients of the estimated valus and observed ones

| Model / Corelation coefficients | Proposed Model | Likelihood Model | Least square Model (A) | Least square Model (B) |
|---|---|---|---|---|
| Cars | 0.79 | 0.73 | 0.82 | 0.66 |
| Large-sized truck | 0.59 | 0.47 | 0.79 | 0.56 |
| Total | 0.79 | 0.75 | 0.83 | 0.68 |

From the correlation coefficients and other statistical measures, except least square method case A, the results estimated by proposed model are the most satisfactory. For the least square model, it is very clear that the estimation accuracy is different corresponding to different prior O-D information. The difference is caused by the property of the least square model. The estimation accuracy is affected by prior information very strongly. In our calculation, destination choice rate is calculated by using the observed O-D matrices for case A and are calculated by multiplying the random coefficient to the observed O-D traffic volumes for case B. So that the estimation accuracy is rather different between these two cases. It is very difficult to know precise prior information on the O-D matrices, so that case A is only an ideal case.

The proposed model takes the interaction of traffic modes into accounts, therefore, a good estimation results are obtained. These results show that the interaction among the traffic modes exists and if we consider this in modelling, then estimation results can be improved.

Table 3. Comparison of Estimation Accuracy

| Statistical measures / Model | MAE% | %RMSE | RMSE |
|---|---|---|---|
| The Proposed Model | 45 | 66 | 6094 |
| | 65 | 95 | 522 |
| | 43 | 64 | 6226 |
| Maximum Likelihood Model | 76 | 180 | 7367 |
| | 116 | 80 | 987 |
| | 70 | 76 | 7391 |
| Least Square Model | 52 | 57 | 5325 |
| | 83 | 205 | 1127 |
| | 49 | 58 | 5687 |

## 6. SUMMARY

A new traffic demand estimation model, a combined distribution, modal split and traffic assignment model in the form of optimization program, is proposed based on existing research works. From this optimization program, a traffic distribution demand model is derived and the existence and uniqueness of solution is also discussed. Additionally,the interaction of traffic modes are considered by introduing the revised BPR performance function. A sensitivity analysis for parameters of the generalized cost function is carried and results shows that the estimation accuracy is affected by parameters. However, the effects are not very significant if the value of parameters veries within certain range. Furthurmore, the proposed model is applied to a Nagoya road network to estimate the O-D matrices for different traffic modes and comparison of estimation accuracy among the proposed modeel and other existing model is carried out. The results of comparison show that the proposed model is superior to the existing models in terms of estimation accuracy, conviendence and applicability. The proposed model can be used to metropolitan without difficulty because route choice rate is internalized and traffic modes have been taken into account.

## 7. References

Bell M. G. H.. The Estimation of Origion-Destination Matrices by Constrained Generalised Least Squares, Transpn. Res.-B, Vol. 25B, No.1, 1991, pp. 13-22.

Boyce, D.E. et al.. Implementation and Computational Issues for Combined Models of Location, Destination, Mode and Route Choice, Environmentand Planning A, Volume 15, 1983, pp. 1219-1230.

Erlander S. etc.. On the Calibration of the Combined Distribution-Assignment Model, Transpn. Res.-B, Vol. 13B, 1979, pp. 259-267.

Fisk, C.S. and Boyce, D.E. A Note on Trip Matrix Estimation from Link Traffic Count Data, Trans. Res 17B, 1983,pp. 245-250.

Heinz Spiess. A Maximum-Likelihood Model for Estimating Origion-Destination Matrices, Transpn. Res.-B, Vol. 21B, No. 5, 1987, pp. 395-412.

Henk J. van Zuylen et al.. Consistent Link Flow Estimation from Counts, Transpn. Res.-B, Vol. 16B, No.6, 1982, pp. 473-476.

Henk J. Van Zuylen and Willumsen L.G.. The Most Likely Trip Matrix Estimated from Traffic Counts, Transpn. Res.-B, Vol. 14B, 1980, pp. 281-293.

Hogberg P.. Estimation of Parameters in Models for Traffic Prediction: A Non-linear Regression Approach, Transpn. Res. 10, 1986, pp. 263-265.

Holm, J., Jensen, T. and Nielson, S.K. et al.. Calibrating Traffic Models on Traffic Census Results only, Traffic Engineering and Control, Vol.17, No.4, 1976, pp.137-140.

Iida, Y. and Takayama, J.. The estimation model of O-D traffic volume based on the link flow, text published by JSCE, 1987.

Kawakami, S., Hirobata, Y. and Xu, Z.. A study on the Multi-class Traffic Assignment Method with Separation of Car and Truck Link Flows, Infrastructure planning review, No.7, 1989, pp. 243-250.

Kawakami S., Hirobata Y. and Lu H.. Estimation of Origin-Destination Matrices from Link Traffic Counts Considering the Traffic Modes, Infrustructure Planning Review, No. 8, 1990, pp. 57-64.

LeBlanc, L.J. and Farhangian, K.. Selection of a Trip Table Which Reproduces Observed Link Flows, Trans.Res-B, Vol. 16B, No.2, 1982, pp. 83-89.

Maud Brenninger-Gothe etc.. Estimating Origin-Destination Matrices from Traffic Counts Using Multiobjective Programming Formulations, Transpn. Res.-B, Vol. 23B, No.4, 1989, pp. 257-269.

Malachy Carey, Chris Hendrickson et al.. A Method for Direct Estimation of Origin/Destination Trip Matrices, Transportation Science, Vol.15,No.1, 1981, pp. 32-49.

Sue McNeil and Chris Hendrickson. A note on Alternative Matrix Entry Estimation Techniques, Trans.Res.-B Vol.19B,No.6, 1985, pp. 509-519.