INTER-URBAN ORIGIN AND DESTINATION MATRIX ESTIMATION USING AUTOMATICALLY OBTAINED AADT TRAFFIC COUNT DATA

Seungjae LEE
Assistant Professor
Dept of Transportation Engineering
The University of Seoul
Dongdaemoon-Ku, Seoul, Korea.
Tel:+82-2-2210-2172
Fax:+82-2-2215-5097
E-mail: sjlee@uoscc.uos.ac.kr

Hunju LEE
Researcher
Dept of Transportation Engineering
The University of Seoul
Dongdaemoon-Ku, Seoul, Korea.
Tel:+82-2-2210-2172
Fax:+82-2-2215-5097
E-mail: hjlee@uoscc.uos.ac.kr

Hyunho CHANG
Graduate Student
Dept of Transportation Engineering
The University of Seoul
Dongdaemoon-Ku, Seoul, Korea.
Tel:+82-2-2210-2990
Fax:+83-2-2215-5097
E-mail: chhho@sidae.uos.ac.kr

Injoon CHOI
Graduate Student
Dept of Transportation Engineering
The University of Seoul
Dongdaemoon-Ku, Seoul, Korea.
Tel:+82-2-2210-2990
Fax:+83-2-2215-5097
E-mail: injoon74@sidae.uos.ac.kr

Abstract: The accuracy of an adjusted O-D matrix depends very much on the reliability of the input data, i.e. link traffic counts, and the number and locations of traffic counting points in the road network. The former has been studied widely in small example networks, while the latter has gained very limited attention. But in the field, the number and locations of traffic counting points is one of the most annoying problems, because networks become larger, the number of traffic counting points is required more. Therefore, this paper investigates these issues as an experiment using a nation-wide network in Korea. We have compared and contrasted the set of link trips assigned by the old and the adjusted O-D matrices with the set of observed link trips. It has been analyzed by increasing the number of the traffic counting points. Finally we can see an optimal set of the number of counting links through statistical analysis.

Key Words: An Old Origin and Destination Matrix, The number and location of traffic counting points, Gradient method, AADT (Annually Average Daily Traffic), TLFD (Trip Length Frequency Distribution)

1. INTRODUCTION

An origin and destination (O-D) matrix is crucial for transportation planning process. The accuracy of the O-D matrix plays key roles to make and evaluate various transportation policies. However, it is very hard and costly to estimate the O-D matrix. Though high amounts of resources are required for the O-D survey, yet the accuracy is relatively low. Even more, the transportation situation and land use have changed very steeply and quickly, and thus the transportation environments have been unstable. Therefore the transportation

planning should be frequently rectified according to the newly planned environments in order to capture the changed situations in the limited budget and time.
Fortunately, there are methods to adjust an old O-D matrix. Because traffic counts are readily available and relatively inexpensive to collect, the estimation method has great economic advantage and is directly applicable for tactical and operational transportation planning (Yang et al., 1992, 1994; Yang, 1995). Generally, the quality of the estimated O-D matrix is greatly dependent on the accuracy of the input data (especially, traffic counts and prior matrix) and the number and locations of traffic counting points (Lam and Lo, 1990; Yang et al., 1991). However, there has been very limited attention that has been devoted to the key problem of identifying a set of links for which flow information should be collected and used. Lam and Lo (1990) proposed some heuristic procedures of identifying the order in which the links should be selected for estimating O-D matrices. Based on the maximum possible relative error (MPRE), Yang et al. (1991) examined the reliability of the estimated O-D matrix with regard to the number and locations of counting points in the network. Yang et al. (1991) proposed a heuristic greedy algorithm to determine the desirable number and locations of counting points. But that algorithm are subject to a critical assumption that path flows associated with a prior O-D matrix must be known. This is a restrictive assumption because path enumeration is required. H. Yang and J. Zhou (1998) proposed a heuristic greedy algorithm. Berman et al. (1995) proposed a new approach for the location of facilities without requirement of the knowledge of path flows. As long as the turning probabilities and initial O-D trip distribution are known in advance, the problem can be formulated as an average-reward Markov decision process. H. Yang and J. Zhou (1998) reformulated Markov decision process. But above methods simulated in a small example network. So we addressed the problem of the number and locations of traffic counting points in the Korean road network, and analyzed and selected the set of the number of traffic counting points through the analysis of both difference between surveyed and assigned link volumes and TLFD(Trip Length Frequency Distribution).

2. EXSTING STUDIES

2.1 MPRE and the location of traffic counting points (Yang et al. 1991)

Since a link traffic volume is the aggregation of trips using selected paths between origin and destination trips, an old O-D matrix can be adjusted and renewed using link traffic volume. Models such as Entropy Maximizing, Maximum Likelihood, Generalized Least Square, Bi-Level, and Gradient model have been developed to update and renew an old O-D matrix from link traffic volume. However, these have been applied on the very limited small road network. It is obvious that there are various errors occurring in the process of the O-D trip matrix estimation using link volume. In the side of the scale of road network, the more errors the larger network, because some link traffic volumes cannot make any contribution to the trip matrix estimation, or even retain garbage information. To decrease errors and obtain the acceptable number and location of counting points, Yang et al. (1991) proposed a new concept of MPRE (Maximal Possible Relative Error), and defined an index of reliability of and estimated O-D matrix. MPRE represents the maximum possible relative deviation between the estimated O-D matrix and the true one.
Now, we first review this concept which is adopted from Yang et al (1991). Let $G(N,A)$ is a road network, where N is the set of nodes, and A is the set of links. Suppose there is a subset

of links $\hat{A}(\hat{A} \subset A)$ on which traffic flows are observed. Let $n,l,\hat{l}$ denotes the number of elements in N, A, and $\hat{A}$, respectively. Let W be the set of O-D pairs and R be the set of routes in the network. The true and estimated trips between O-D pairs $w \in W$ are denoted by t^*_w and t_w, respectively. For each O-D pair w, the proportion of trips using link $a \in \hat{A}$ is denoted by p_{aw} $(0 \le p_{aw} \le 1)$ and the observed flow on link a by v_a. Assume that the proportion p_{aw} has been determined exogenously and that the link traffic counts are error-free. Thus the true and estimated O-D trips must satisfy the following.

$$\sum_{w \in W} p_{aw} t_w = v_a, a \in \hat{A} \tag{1}$$

$$\sum_{w \in W} p_{aw} t^*_w = v_a, a \in \hat{A} \tag{2}$$

Subtracting eq.2 from eq.1, we have eq.3

$$\sum_{w \in W} p_{aw}(t^*_w - t_w) = v_a, a \in \hat{A} \tag{3}$$

The proportion p_{aw} can be obtained by using from simple all-or-nothing trip assignment to more complex equivalent trip assignment. If all p_{aw} and v_a are given, t_w calculated by L simultaneous equations come to N^2 or at least $N^2 - N$, not considering intrazonal trips where, L denotes the number of observed traffic, and N is the number of trip matrix cells. Let $\lambda_w = (t^*_w - t_w)/t_w$ denote the relative deviation of the estimated trips from the true one for O-D pair $w \in W$. Because $t^*_w \ge 0$, $t_w \ge 0$, thus $\lambda_w \ge -1, w \in W$. From eq.3, we have

$$\sum_{w \in W} p_{aw} t_w \lambda_w = 0, a \in \hat{A} \tag{4}$$

Define eq.5

$$G(\lambda) = \sqrt{\sum_{w \in W} \lambda_w^2 / m} \tag{5}$$

as a measure of the estimation error of the O-D matrix where m is the number of elements in W. Evidently, the smaller $G(\lambda)$, the higher the accuracy of estimation. So the maximal possible relative error can be defined as the maximum $G(\lambda)$ subject to constraint eq.4:

$$MPRE(\lambda) = \max G(\lambda) \text{ subject to } \sum_{w \in W} p_{aw} t_w \lambda_w = 0, a \in \hat{A} \tag{6}$$

2.2 Location rules (H. Yang and J. Zhou, 1998)

Yang et al. (1998) derived four counting location rules from the theory of MPRE. Obviously,

link traffic counts should contain information as much as possible to increase the certainty or reduce the feasible space of the O-D matrices defined by the system of eq.1. This is equivalent to selecting traffic counting points so that the resultant $MPRE(\lambda)$ is minimized.
H. Yang and J. Zhou proposed the following four location rules.
Rule 1 (O-D covering rule): the traffic counting points on a road network should be located so that a certain portion of trips between any O-D pare will be observed.
Rule 2 (maximal flow fraction rule): for a particular O-D pair, the traffic counting points on a road network should be located at the links so that the flow fraction between this O-D pair out of flows on these links is as larger as possible.
Rule 3 (maximal flow-intercepting rule): under a certain number of links to be observed, the chosen links should intercept as many flows as possible.
Rule 4 (link independence rule): the traffic counting points should be located on the network so that the resultant traffic counts on all chosen links are not linearly dependent.

2.3 Shortcomings to the adjustment O-D matrix from link counts

In the estimation of O-D matrix using observed link traffic volume, there exist some problems such underspecification, dependency and inconsistency, congestion effects, and the number and location of counting points. We can briefly state these issues mentioned by Yang et al (1991) as follows.

1) Underspecification
If observed link volume is less than assigned volume, underspecification problem occurs. Mathematically, this problem is that the number of unknown quantity is more than that of equations, so the one solution value doesn't exist. To prevent solution from roller-coaster condition, the way to cut down on an unknown quantity is required. Techniques for unraveling this problem, taken for the following two examples, can be considered. The one is that the method to diminish the number of possible solutions on the assumption of trip behaviors gives consistency through such as the minimization of information, entropy maximizing, and so on. The other is that past information can be used to update an old O-D table. Especially the difference of a present traffic volume from a past traffic volume is acceptable.

2) Dependency and inconsistency
Dependency problem is that some link traffic counts among observed traffic counts in the whole road network bear a linear relation with a surrounding link traffic counts. That is to say, if inflow is equal to outflow, a linearly dependent relationship arises between links, and we can't gain any additional information from these links. Figure 1 shows this relationship.

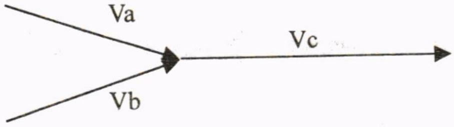

Figure 1. Y-shaped network

If inflow is equal to outflow, Vc = Va+Vb. Vc is linearly dependent, and both Va and Vb are independent, so link c hasn't any additional information. In reality, because an observed traffic counts contain themselves errors, there is an inconsistency problem (e.g., Vc = Va+Vb+α) which leads to failure to estimate acceptable O-D trips. It is hard to eliminate this inconsistency problem. But we can use the way of giving error bounds to models. The cause of inconsistency problem can be divided into two. First, both in and out boundary volume on a link can't be equivalent in the both side of time and space. Second, an assumed trip assignment model can't explain clearly observed link volume. For example, it maybe occurs that an assignment model loads no trips on the link where trips are observed. Under those conditions any O-D trip, which can describe the link volume through the assignment model never exist.

3) Congestion effect

The congestion-effect problem is that the link-choice probability can either reflect the degree of traffic congestion on the road network or not. An O-D table can be estimated by the probability of link choice under the conditions subject to surveyed link data. Either trip assignment model or the probability model of link choice can be used to obtain a link-choice probability. In the case of the former, the selection of a suitable trip assignment method plays a key role in estimating O-D table. And trip assignment method can be divided into two; the one is proportional trip assignment not considering congestion, the other is none proportional trip assignment considering congestion. Due to not considering congestion, proportional trip assignment irrelative to the estimation process of O-D decides independently the proportion of link choice. Therefore, proportional trip assignment is effective in the road network, which does not comprise congested links, but ineffective in the road network do congested. But in the case of capacity-restricted equivalent trip assignment, the proportion of link choice is continually changed according to the fluctuation of trip demands. More grave underspecification problem may be occur because the selected proportion of links to the whole network is changed depending on the rise and fall of trip demands, and the proportion of link choice is distorted.

2.4 Gradient method (Spiess, 1990)

To adjust the O-D matrix using link traffic counts, we used the gradient method, also called the method of steepest descent, proposed by Spiess. It is formulated as an optimization problem in order to minimize a difference of surveyed and assigned volumes. The simplest function of this type is the square sum of the differences, which leads to the convex minimization problem. Since the matrix estimation problem as formulated in the Spiess is highly underdetermined, it usually admits an infinite number of optimal solutions, i.e. possible demand matrices which all reflect the observed volumes equally well. In the actual planning process, we expect the resulting matrix to resemble as closely as possible the base matrix, because it contains important structural information on the inter O-D movement, taken for example TLFD. Therefore, just finding one solution to the problem in Spiess in evidently not enough.

If we would have a solution algorithm that inherently finds a solution close to the starting point, we could leave the objective function as is. Fortunately, the gradient method has exactly this property that we search for. It follows always the direction of the largest yield regarding minimizing the object function and, thus, it also assures us not to deviate from the starting solution more than necessary. To implement the gradient method, we also need to provide values for the step lengths. Choosing very small values for the step length has the advantage of following more precisely the gradient path, but has the disadvantage of requiring more steps. On the other hand, when choosing too large values for the step length, the objective function can actually increase and the convergence of the algorithm would be lost. Thus, the optimal step length at a given demand can be found by solving the one-dimensional subproblem.

3. STRUCTURE OF NETWORK AND SELECTED LOCATION

The nationwide road network for this study, including from expressway, arterial, to collector and local, comprises 3,055 nodes, 1,051 links, and 237 zones. And a supplementary Cntpost.mac program, operated in a suite of traffic network analysis programs, EMME/2, was used to select the optimal number and location of traffic counting points. Table 1 shows the number of selected traffic counting points according to the level of road.

Table 1. The number of selected traffic counting points

Road types	Num of links	Num of selected location	Pct. (%)
Expressway	528	338	64.02
Principal Arterial	4,303	288	6.69
Minor Arterial	1,282	596	46.49
Collector and local	643	236	36.70
Others	3,295	-	-
Total	10,051	1,458	14.51

Table 2 shows PCE, Passenger Car Equivalent, required to unify various vehicles into PCU, Passenger Car Unit, according to road and vehicle types.

Table 2. PCE according to road and vehicle types

Road/vehicle type	Car	Bus		Truck
		Small	Standard	
Expressway	1.0	1.0	1.3	1.5
Others	2.0	1.3	1.3	1.5

4. ANALYSIS

4.1 Measure of analysis

It is clear that the better O-D trips the better explanations of inner traffic phenomena, such as the fluctuation of the link volume and TLFD (trip length frequency distribution). Therefore, we will compare and contrast the set of link trips assigned by the old and the adjusted O-D matrices with the set of observed link trips depending on a step-up increase in the number of the traffic counting points [see table 3]. We have some problems about the analysis of estimated O-D tables, because it is almost impossible to know the true O-D table especially on the large size road network. Therefore, there exists nothing that can be used for the fixed points. Having said that, TLFD fused in an O-D table is indicative of trip behaviors. So we will make a comparison and draw a parallel between the base TLFD and estimated TLFDs. The statistical measures of error analysis for stated above are the following.

$$\text{MAE}=\frac{1}{N}\sum_i |x(t)-\hat{x}(t)| \tag{7}$$

$$\text{EC}=1-\frac{\sqrt{\sum_i (x(t)-\hat{x}(t))^2}}{\sqrt{\sum_i x(t)^2}+\sqrt{\sum_i \hat{x}(t)^2}} \tag{8}$$

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_i (x(t)-\hat{x}(t))^2} \tag{9}$$

Where, $x(t)$ denotes observed link traffic counts (or base TLFD), $\hat{x}(t)$ denotes assigned link traffic counts (or estimated TLFDs), N denotes the number of traffic counting points (or the number of time slices).

4.2 Analysis of link volume

In the analysis of link volume, Link-Scattergram and R-square value are used together with the above three measures. The nearer R-square values go to 1.0, the nearer gradients 0.5, the higher the accuracy of estimation. Figures 2, 3, 4, and 5 show the link-scattergrams according to the increase in the number of traffic counting points.

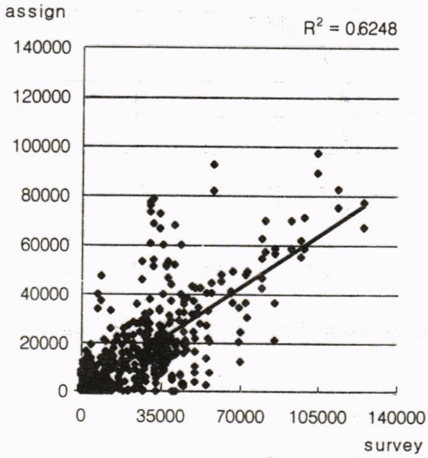

Figure 2. Link-Scattergram (before adjusted)

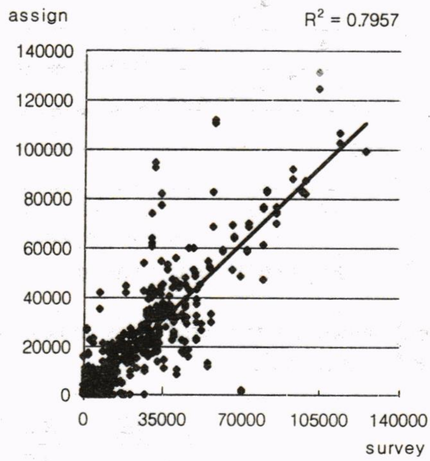

Figure 3. Link-Scattergram (642)

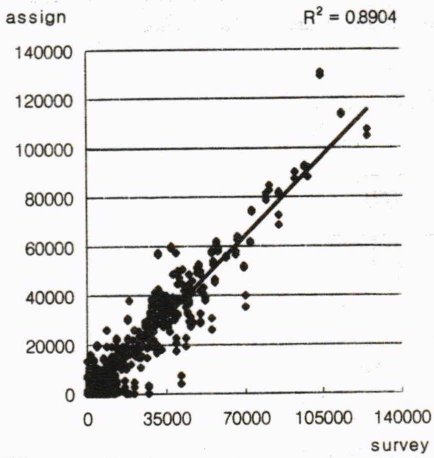

Figure 4. Link-Scattergram (1,164)

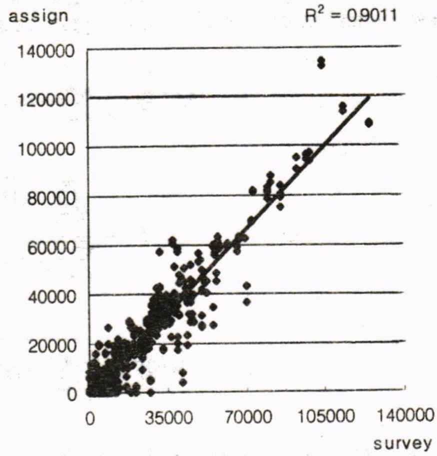

Figure 5. Link-Scattergram (1,458)

In figures 6 and 7, the error values converge steeply from 0 to 1060 and smoothly from 1164 to 1458 depending on the number of traffic counting points.

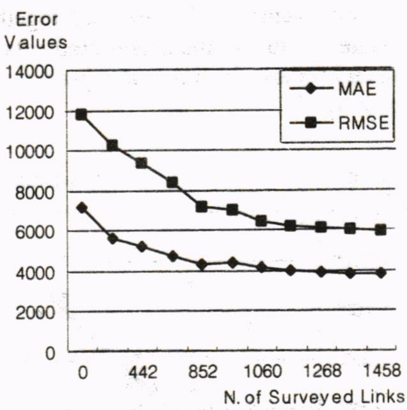

Figure 6. Error estimation (1)

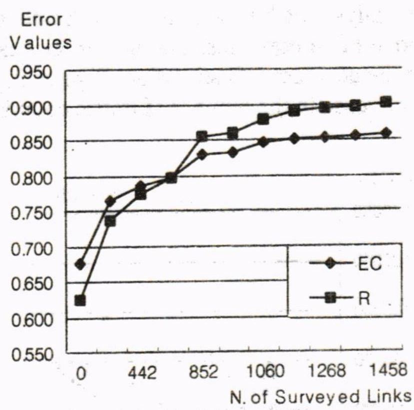

Figure 7. Error estimation (2)

Table 3 shows the estimated error values. And we used the following function to estimate above four measures.

$$\left(\frac{MAE_i}{MAE_{\max}} + \frac{RMSE_i}{RMSE_{\max}} + \frac{EC_{\max}}{EC_i} + \frac{R_{\min}}{R_i}\right) / N \qquad (10)$$

Where, N is the number of the measures

Table 3. Link volume analysis according to the number of traffic counting points

N. of Surveyed Links	MAE	RMSE	EC	R-Value	Estimated Error
0	7131	11826	0.675	0.625	1.000
242	5649	10249	0.763	0.735	0.848
442	5181	9329	0.784	0.774	0.796
642	4761	8418	0.796	0.796	0.753
852	4350	7189	0.829	0.854	0.691
954	4392	7002	0.831	0.859	0.687
1060	4155	6456	0.845	0.878	0.660
1164	3995	6173	0.850	0.890	0.645
1268	3930	6079	0.853	0.894	0.639
1356	3842	5987	0.855	0.898	0.633
1458	3862	5961	0.857	0.901	0.632
Mean	4659	7697	0.813	0.828	0.726

It is turned out that the common idea, the more the number of traffic counting points the higher the accuracy of an adjusted O-D table, is not a myth.

4.3 Analysis of TLFD

With regret, we don't know the true O-D trips. But adjusted O-D trips can be explained by TLFD (Trip Length Frequency Distribution), because TLFD explains trip behaviors fused in the base O-D matrix that can be initially estimated on the base of the expansion techniques using the person-trip-surveyed data. Like the analysis of the comparison of observed and assigned link volume, TLFD is analyzed according to the increase of the number and locations of traffic counting points. Figures 8, and 9 show the TLFD according to the increase in the number of traffic counting points. In the two Figure The TLFD of adjusted O-D trips becomes almost similar to that of base O-D trips depending on the number of counting points.

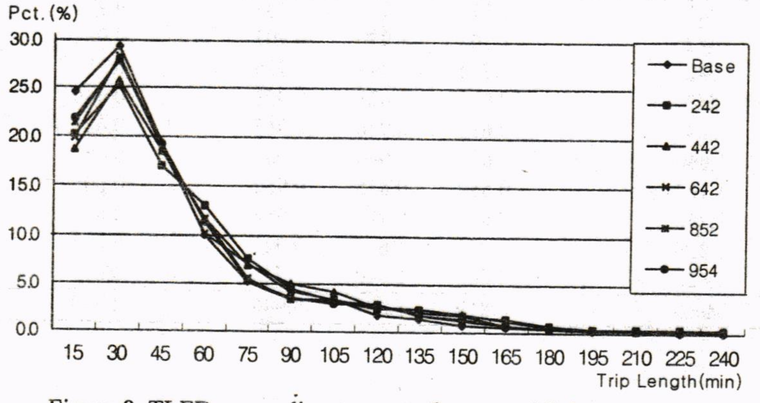

Figure 8. TLFDs according to a set of surveyed link counts (1)

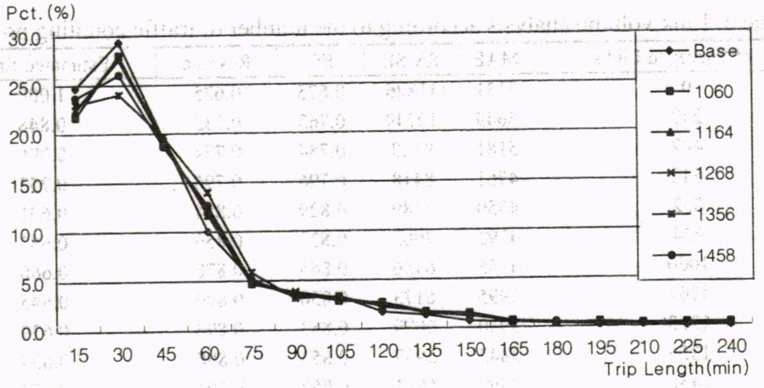

Figure 9. TLFDs according to a set of surveyed link counts (2)

Figure 10 and Table 4 show the estimated error values. And we used the following function to estimate changes of TLDF.

$$\left(\frac{MAE_i}{MAE_{\max}} + \frac{RMSE_i}{RMSE_{\max}} + \frac{EC_{\max}}{EC_i}\right) / N \tag{11}$$

Where, N is the number of the measures

In Table 4 and Figure 10, when a set of traffic counting points is from 1060 to 1268, Error values are acceptable. Especially, Figure 10 says that TLFDs of adjusting O-D tables is not continually similar to the base TLFD according to the increase in a set of traffic counting points.

Table 4. TLFD analysis according to the number of traffic counting points

N. of Surveyed Links	MAE	RMSE	EC	Estimated Errors
242	1.379	1.951	0.908	0.996
442	1.305	1.970	0.907	0.982
642	0.794	1.393	0.935	0.751
852	0.641	1.028	0.953	0.646
954	0.611	0.930	0.957	0.621
1,060	0.697	1.040	0.952	0.662
1,164	0.589	0.875	0.960	0.605
1,268	0.630	0.942	0.957	0.628
1,356	0.874	1.615	0.921	0.813
1,458	0.666	1.140	0.948	0.673
Mean	0.819	1.289	0.940	0.738

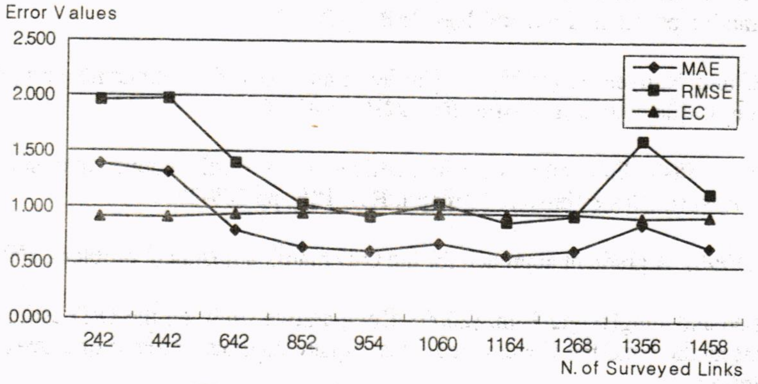

Figure 10. Error Analysis of TLFD

5. SUMMARY

In this study, the estimation of O-D matrix has been tested on the nationwide network with the Gradient algorithm provided in EMME/2, depending on the increase in the number of traffic counting points. And the inner structure, i.e. TLFD, in the base O-D matrix is more or less maintained after the adjustment process has been carried out using the Gradient method. Obviously, the difference of the link volume assigned by an adjusted O-D matrix from that assigned by the base O-D matrix decreases depending on the increase in the number of traffic counting points, but it can occurs that Trip Length Frequency Distribution is changed. Though adjusted O-D matrices diminished the discrepancy of the surveyed counts and assigned counts, yet the underspecificaion problem is occurred. Therefore, if the difference between assigned link volume with an adjusted O-D matrix and surveyed link volume is acceptable, we should reconsider to select a set of traffic counting points, because it is not desirable that the base TLFD is seriously distorted. Above statement needs seriously to be considered, especially when the transportation situation and land use change very steeply and quickly.

REFERENCES

a) Books and Books chapters

INRO, Consultants Inc. (1999), **EMME/2** User's Manual.

b) Journal papers

H. Yang, Y. Iida, T. Sasaki, Y. Asakura. (1992). Estimation of origin destination matrices from link traffic counts on congested networks. **Transpn. Res. 26B**, pp 417-434

H. Yang, (1998). Otimal traffic counting locations for O-D estimation, **Tranpn. Res. 25B**, pp 351-363

H. Yang and Y. Iida and Tsuna Sasaki, (1994). The Equilibrium based origin destination matrix estimation problem. **Transpn. Res. 28B**, pp 23-33

Cascetta E. and Nguyen S. (1998). A Unified framework for estimating or updating OD matrices from traffic counts, **Transpn. Res. 22B**, pp 437-445

Cascetta E., (1984). Estimation of trip matrices from traffic counts and survey data : generalized least squares estimator, **Transpn. Res. 18B**, pp 289-299

Spiess, H. (1990). A gradient approach for the O-D matrix adjustment problem, **INRO**.

Seungjae Lee and JongHyung Kim, (2000). Comparative study on the O-D estimation using the Gradient method and the Generalized Least Square method, **Journal of Korea Society of Transportation**.

Seungjae Lee, EuiYoung Son, JongHyung Kim, (2000). Evaluation on the traffic count based O-D matrix using Trip Length Frequency Distribution, **Journal of Korea Society of Transportation**.

Seungjae Lee and JongHyung Kim, (2001). Development of a demand estimation method by using multiclass traffic assignment based on traffic counts, **Journal of Korea Society of Transportation**.

JongHyung Kim (2000). Travel Demand Estimation from traffic counts using the Gradient Method – In the Focus of Multiclass traffic assignment on the Large scale network, PhD Thesis, Univ. of Seoul.

Gur Y.J. et al. (1980). Estimation of an origin-destination trip table based on observed link volumes and turning movements-volume 1: **Technical report. Report No. RD-80/034**, FHWA, U.S. Department of Transportation Washington, D.C.

Fisk C. S. and Boyce D. E. (1983). A note on trip matrix estimation from link traffic counts data, **Transpn. Res. 17B**, pp 245-250

Fisk C. S. (1989). Trip matrix estimation from link traffic counts: The congested network case, **Transpn. Res. 23B**, pp 331-356

Berman, O., Larson, R. and Fouska, N. (1992). Optimal location of discretionary service facilities. **Transportation Science 26**, pp 201-211