A STATISTICAL ESTIMATION OF THE AADT USING MONTHLY, WEEKLY AND WEATHER VARIATION FACTORS

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

Namchul BAEK
Researcher
Advanced Highway System Research Group
Korea Institute of Construction Technology
Tel:+82-31-910-0194
E-mail: nc100@kict.re.kr

Heejeong KWEON
Graduate Student
The University of Seoul
Tel:+82-2-2210-2990
E-mail:khj_korea@sidae.uos.ac.kr

Daesoon CHOI
Research Fellow
Advanced Highway System Research Group
Korea Institute of Construction Technology
Tel:+82-31-910-0194
E-mail: nc100@kict.re.kr

Abstract: In inter-urban regional road sections, only one to three daily traffic data are counted manually according to the schedule and financial budgets. Using these manually obtained few daily traffic data, we have to estimate the AADT. In order to do so, we have to obtain information and characteristics from continuous monitoring sites' data. And then, we have classified these continuous monitoring locations into similar subgroups using the cluster analysis. After these processes, we have obtained weekly and monthly adjustment factors of each pattern cluster so as to estimate the AADT. The most important thing in this paper is that we have used the time series analysis to analyze traffic patterns in conjunction with weather data as an irregular variation. We have compared the AADT estimations using time series analysis with those of the conventional estimation models used in FHWA(1995). Accuracy for the result of the AADT estimation was highest when we used monthly and daily variables.

Key Words: Annually Average Daily Traffic, Weekly Monthly Weather Variation Factors. Statistical Estimation Techniques.

1. INTRODUCTION

Intelligent Transport Systems have been introduced and deployed many high-tech and reliable surveillance systems. These enables to obtain and use the accurate data for the estimation of the AADT (Annually Average Daily Traffic) data. In Korea, the traffic counted data are automatically obtained using the loop detectors on the approximately 430 selected locations of the Korea inter-urban regional road networks. On the other hand, in about 1200 locations on inter-urban regional road sections, only one to three daily traffic data are counted manually according to the schedule and financial budgets.

However, for these manually obtained few daily traffic data, we have to estimate the AADT. In order to do so, we have to obtain information and characteristics from automatically counted data. In the first process, we have classified 430 locations into similar subgroups using the cluster analysis. In the second process, manually obtained daily traffic on 1200 locations are also classified into the subgroup made in the first stage according to its own characteristics using statistical techniques such as Discriminant and Factor Analysis. And then we estimated the ADT (Average Daily Traffic) using these obtained data. After these processes, we have obtained weekly and monthly adjustment factors in order to estimate the AADT.

Until recently, we use only weekly and monthly adjustment factors in order to estimate the AADT. However, in order to reduce random error components, we have looked for the weather conditions as an influential factor. Weather conditions have impacts on traffic activity. Therefore, in this paper, we have studied how to estimate AADT by considering weather irregular factors. In the first stage of the estimation, we had to know various factors showing characteristics of traffic pattern. And then we extracted variation factors from traffic data. Factors are based on directional, monthly, weekly and daily variations. In addition to these factors, we have to consider impacts of an irregular variation including weather conditions such as rainfalls, fog, temperatures, and sunshine hours among others. It is because traffic is influenced by weather conditions. In order to factor it out, we gathered hourly traffic volume data of 200 permanent count spots and 1200 temporary count spots in Korea regional roads. We also collected meteorological data according to those spot areas.

Meteorological data was used to determine whether weather conditions have irregular variations. The differences in the observed values in traffic and weather have been used to determine if unseasonable weather was linked to changes in traffic activity. According to some results, the changes in the average weekday traffic activity were less than 3% for unseasonable sunshine hours, maximum temperature, minimum temperature and rainfall. However on weekends there were reduction of more than 4% in the average traffic activity on both the days of the highest rainfalls and the days of lower than expected minimum temperatures. In the case of weather conditions being linked to changes in traffic activity, we have to include irregular weather variation factors to estimate AADT in addition to monthly and weekly factors.

In this study, the AADT estimation method has been developed and applied using three different models. First, we have applied the traffic pattern clustering analysis which is the traditional AADT estimation model. Second, we estimate the AADT using time-series analysis without considering irregular variation. Third, we estimate the AADT using time series analysis with considering irregular weather variation, Then, the developed models of the AADT estimation have been compared each other using real data obtained in Korea regional roads.

2. THE AADT ESTIMATION METHOD IN FHWA(1995)

The AADT estimation method by traffic pattern clustering analysis has been introduced in the traffic volume inspection guide of FHWA (1965, 1985, 1992, 1995) and has been studied by David (1981), Joe(1996), and Alan(1997). In Korea, AADT estimation of the nationwide inspection sites has gone through a dramatic change having 1995 as the starting point. First of all, before 1995 we had used the method of assuming and inspecting the third Thursday of October as the nearest day to AADT. However, currently this method has been applied in the national highway and regional roads, and had been applied in the general national roads until 1994. Secondly, after 1995 for the general national roads, the daily traffic had been inspected five times randomly between March and December. We have used the method of simply averaging this value. However, in the FHWA(1995), the variables from the CM (Continuous Monitoring) sites have been applied in the nationwide inspection of traffic volume to estimate the AADT by the following equation.

$$AADThi = 1/2\sum (Volhi * Mh * Dh * Ah * Gh) \quad \text{.........................}(1)$$

Where,
AADThi: AADT of functional classification h at I site
Volhi: 24-hour corridor traffic volume of functional classification h at I
Mh: monthly variable, which is appropriate for functional classification h
Dh: daily variable, which is appropriate for functional classification h (if needed) or (necessary)
Ah: adjustment factor, which is appropriate for functional classification h (if necessary)
Gh: growth factor, which is appropriate for functional classification h

Pattern Clustering is calculating the parameters, which show the average traffic characteristic of each cluster, as clustering the CM sites with similar traffic characteristics. The most widely used parameter, which represents the traffic characteristics, is a parameter, It shows the monthly change aspects and the daily change aspects. Therefore, clustering in this study shows the similar traffic characteristics of a continuous road section. And this phenomenon is assumed to continue for a long period of time.

3. ESTIMATION PROCESS OF THE AADT

The accuracy of the AADT estimation depends on the reasonableness of the TPG(Traffic Pattern Grouping) clustering and TPG assignment. It goes through the following process step by step. First of all, we have to make variables for the TPG clustering so that we discriminate the traffic changing characteristics of traffic. Secondly, we cluster traffic in each case using monthly variables and daily variables. Thirdly, we apply the monthly variables and the daily variables of the each TPG to the ADT, and then estimate the AADT. In this step, we compare the AADT estimation method by time series analysis with the others by applying daily variables and monthly variables.

3.1 Pattern analysis of CM sites' traffic

① Variation factor

$$\text{Monthly variables} = \frac{MADT}{AADT}, \quad \text{Daily variables} = \frac{WADT}{AADT}, \quad \text{Hourly variables} = \frac{AHT}{AADT}$$

Where, the MADT is the Monthly Average Daily Traffic, the WADT is the Weekly Average Daily Traffic and the AHT is the Average Hourly Traffic.

We calculated variation factors for each site. Monthly variables and Daily variables are used as factors of clustering analysis. And we used Hourly variables to discriminate clusters for each site.

② Variation factor by time series analysis

We calculated these factors to make a new model. Because traffic is one of data affected by time, we can consider it as time series and use time series analysis. And then, after estimating the AADT, we compared this result with the result of an exiting method using monthly variables and daily variables.

Variations of time series are Trend(T), Seasonal variation(S), Cyclic variation(C) and Irregular variation(I). Model of AADT estimation is $AADT = ADT \times T \times S \times C \times I$.

We found Trend by using linear regression and Seasonal variation by 7 days moving average since traffic pattern similarly changes every week. Irregular variation is the traffic pattern remaining after removing T, S and C variation from each site's traffic pattern.

We thought about what factors affecting Irregular variation, analyzed relation between those factors and Irregular variation, and then intended to make a method of explaining Irregular variation. There are weather effects (from snow, rain, temperature among others)and oil price among others, but we only used the amounts of rainfall and the degree of lowest temperature in a day.

Next figures show changes of traffic data in each stage applying time series analysis. The stages consist of processes removing T, C, S and I from law data.

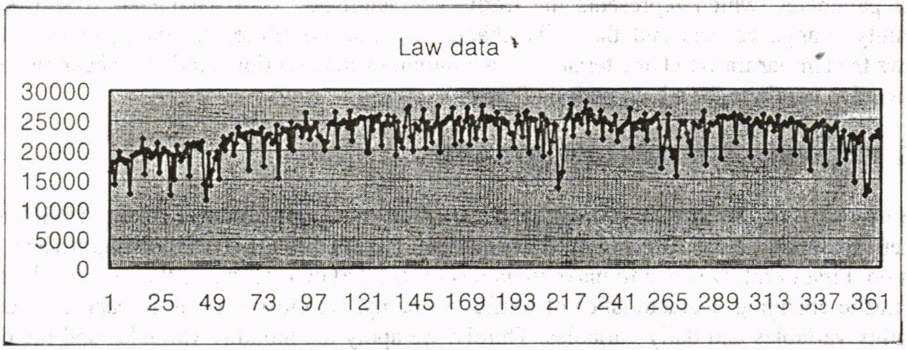

Figure 1. Law data of traffic during 365 days

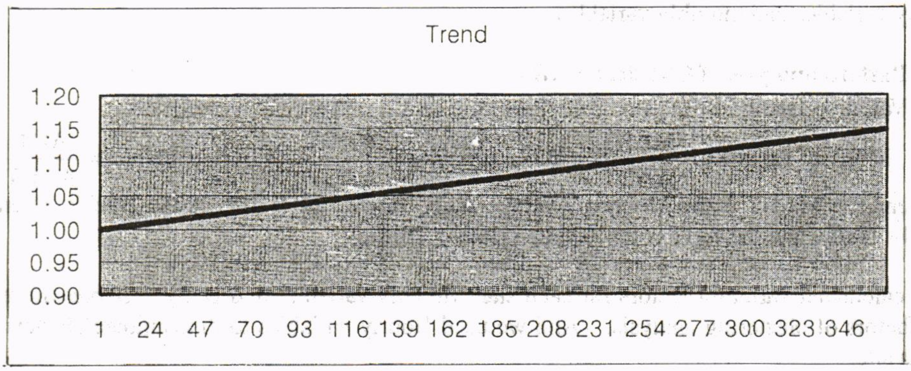

Figure 2. Trend

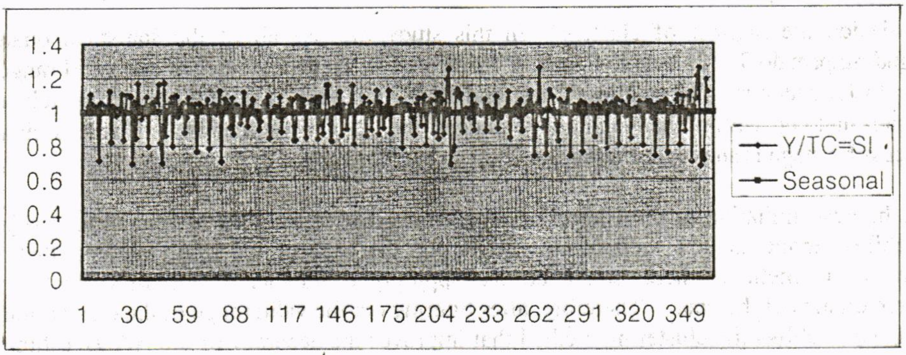

Figure 3. Seasonal variation (S) and SI

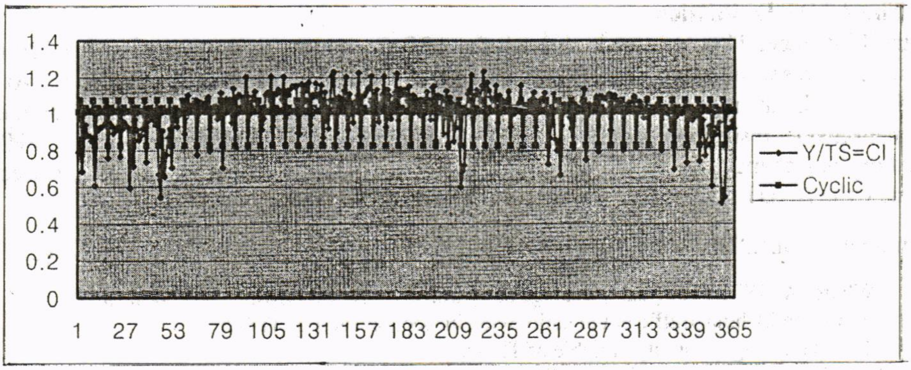

Figure 4. Cyclic variation (C) and CI

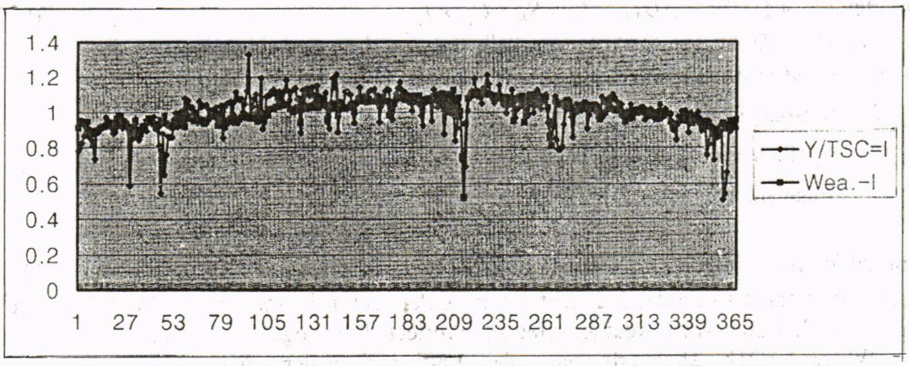

Figure 5. Irregular variation (I) and I joined with weather effect (Wea. –I)

3.2 TPG Clustering by the Clustering Analysis

Clustering analysis is a statistical method, which clusters in the base of the similarity among the cluster members. Three problems have to be positioned. First of all, what should be the standard for measuring the similarity? In this study, squared Euclidean Distance that is accepted theoretically and practically for measuring the similarity. Secondly, what clustering methods are to be used? In this study, we have examined and compared the Ward method and the K-means Clustering method. Thirdly, what kind of statistical quantity should we use when we decide the number of clusters? In this study, we use all of the statistical quantity, including pseudo T**2, pseudo F, and R**2 offered by the SAS to decide the valid range for the cluster number. We applied U-test and RMSE amount to decide how accurately each cluster number within the valid range judges the AADT estimation. This examination process by stage is shown in Figure 1.

The biggest limitation of the Clustering Analysis is that there is no equipment to verify the reliability about the analyzed result. Also, for the case of the Ward method that is a hierarchical method, there could be an appropriate number of clusters due to the characteristics of the data. Therefore, after figuring out the valid range of the cluster number first, we establish the clustering method that improves the accuracy of the AADT estimation and decide the number of clusters at last.

3.3 The AADT Estimation

At the first stage, TPG was made through the TPG Clustering by using data of the 220 CM sites. At the second stage, the TPG was assigned at the nationwide inspection site. At the third stage, the average monthly variables and daily variables for 24 hour-traffic volume of the nationwide inspection site were multiplied together to calculate the AADT. FHWA equation was changed to fit the condition of Korea. The following equations are used to estimate the AADT

① model 1 : $AADThi = VOLhi \times \frac{1}{Mh} \times \frac{1}{Dh}$ ....(2)

Where, AADThi: The yearly average daily traffic of the nationwide inspection site
VOlhi: the 24 hour-traffic at the nationwide inspection site I of TPG h
Mh: The average monthly variable of TPG h
Dh: The average daily variable of TPG h

② model 2 : $AADThi = ADT_{hi} \times T_{hi} \times S_{hi} \times C_{hi} \times I_{hi}$ ....(3)

Where, AADThi: The yearly average daily traffic of the nationwide inspection site
T_{hi} : Trend of at site I of TPG h
S_{hi} : Seasonal factor at site I of TPG h
C_{hi} : Cyclic factor at site I of TPG h
I_{hi} : Irregular factor at site I of TPG h

③ model 3: $AADThi = ADT_{hi} \times T_{hi} \times S_{hi} \times C_{hi} \times IW_{hi}$ ....(4)

This model is same to model 2 except for Irregular factor. In this model, the I is joined to weather data. In other word, weather effects are reflected in irregular variation.

Where, AADThi: The yearly average daily traffic of the nationwide inspection site
T_{hi} : Trend of at site I of TPG h
S_{hi} : Seasonal factor at site I of TPG h
C_{hi} : Cyclic factor at site I of TPG h
IW_{hi} : Irregular factor at site I of TPG h

4. RESULT ANALYSIS

4.1 TPG Clustering

We have performed the clustering analysis for the monthly fluctuations and the daily fluctuations. Next figures show the comparison of statistics from the time when the number of cluster is one to when it is 20 by using the ward method about monthly variables and daily variables. The statistical quantity is a generated result of SAS/STAT.

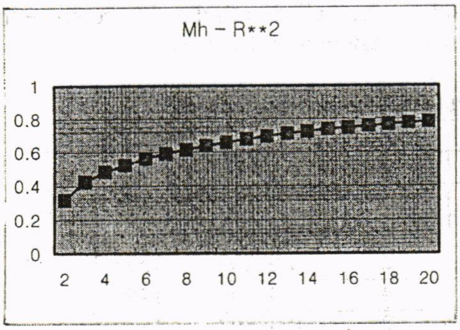

Figure 6. R**2 in case of Mh.

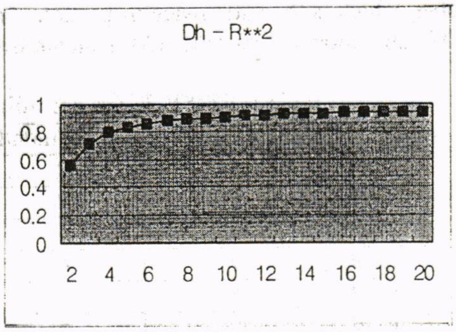

Figure 7. R**2 in case of Dh

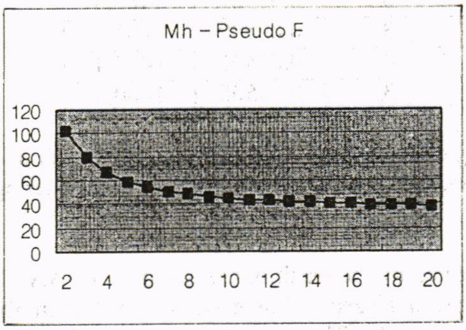

Figure 8. Pseudo F in case of Mh.

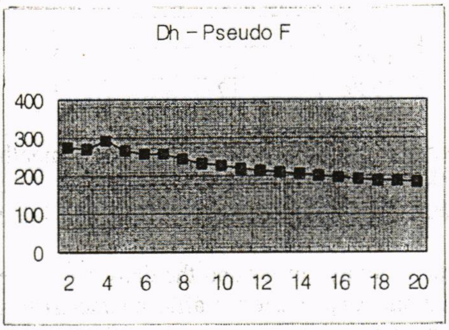

Figure 9. Pseudo F in case of Dh

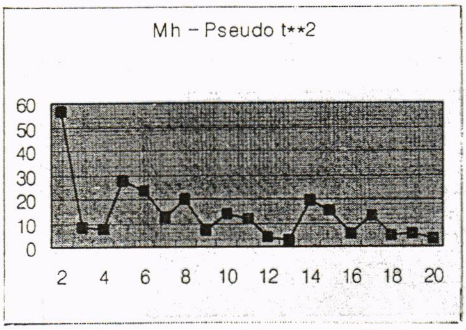

Figure 10. Pseudo t**2 in case of Mh.

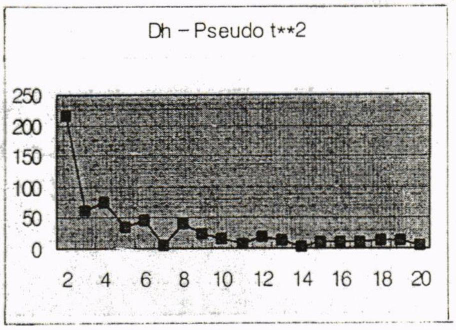

Figure 11. Pseudo t**2 in case of Dh

In Figure 7, 9, 11 about daily variables, the R**2 increases rapidly when there are 4 clusters, but from the fifth cluster, the fluctuation is not great. Pseudo t**2 shows the local extreme point when there are 4 and 6 clusters. Pseudo F value shows the local extreme point when there are 4 clusters. If we unite the statistical quantity of these four values, the appropriate number of cluster is appeared to be 4 or 6. After we analyzed the various statistical quantities applying the ward method for the monthly variable, 3 to 8 clusters were judged to be included in the valid range. This process is applied identically to the k-average

Clustering method. As a result, the valid range of the number of clusters is shown to be identical as that of the ward method. By the accuracy evaluation of AADT estimation, we have finally decided the method of clustering and the number of clusters. The number of clusters, clusters average monthly variable and the average daily traffic of third week of October, which is the nationwide inspection period, are multiplied to find and to compare the AADT estimation and the true AADT drawn from the clustering analysis result.

Table 1. RMSE, U-test according to cluster number changing

Year	N. of cluster	Cl. By month	Cl. By day	U-test	RMSE
1999	4	2	2	0.03999	1678.4
1999	6	2	3	0.03618	1507.5
1999	6	3	2	0.03801	1592.5
1999	7	4	2	0.03803	1593.2
1999	8	2	4	0.03600	1499.8
1999	9	5	2	0.03802	1592.6
1999	9	2	5	0.03574	1488.8
1999	9	3	3	0.03507	1460.3
1999	10	4	3	0.03509	1461.1
1999	11	3	4	0.03491	1453.4
1999	12	5	3	0.03508	1460.4

Looking at the Table 1 when there are 3 monthly clusters and 4 daily clusters, the U-test and the RMSE are the smallest. There fore, we choose the ward method among the clustering methods, 3 monthly clusters for TPG, and 4 daily clusters.

For the monthly cluster and the daily cluster selected by the result of the accuracy of AADT estimation, each clusters traffic characteristics are analyzed. For each of the 3 monthly clusters, cluster average monthly variable was calculated. The monthly cluster variables are shown in figure 12.

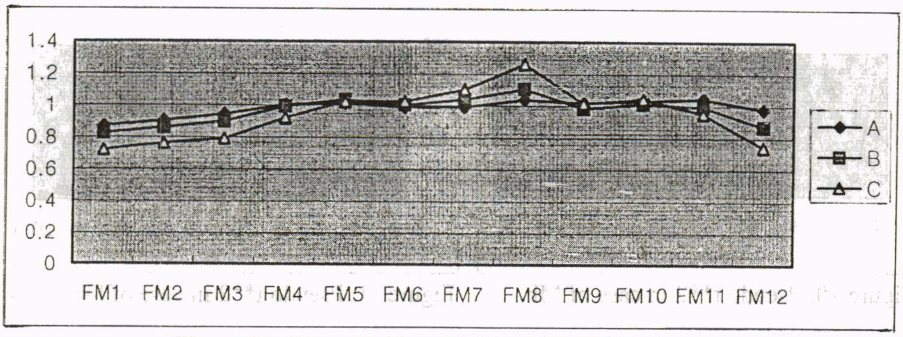

Figure 12. The monthly variables of monthly clusters

Moving from the monthly cluster A to C, July and August go up, but the rest goes down. Also, June is the safest month closest the AADT among all of the monthly clusters. We can see that having June, July and August at the center, from Jan. to Apr. and from Nov. to Dec., the difference among variables by cluster gets bigger.

For the daily clusters, the changing characteristic is shown in figure 13. Tuesday, Wednesday, and Thursday show commonly a stable shape, and in the order of daily cluster a, c,

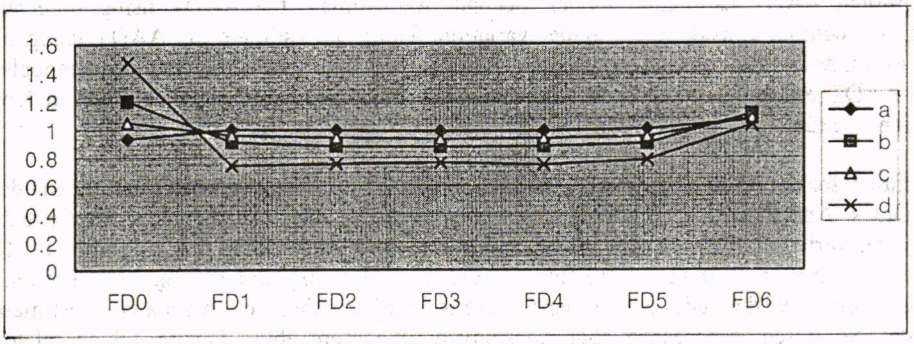

Figure 13. The Daily variables of daily clusters

b and d, the variable for Sunday increases. Especially, the daily cluster, a, has stable weekday traffic, but decreased weekend traffic. The cluster c is quite similar to cluster b, but the Sunday traffic reduces to the level of weekday traffic. The daily cluster, d, has a characteristic of having concentrated weekend traffic.

4.2 AADT estimation and evaluation

① AADT estimation by clustering analysis
ADT, Mh and Dh are used as coefficients of the AADT estimation model 1.

② AADT estimation by time series analysis
ADT, T, S, C, and I are used as coefficients of the AADT estimation model 2 and 3.

RMSE and U-test are a method of evaluation, comparing estimated AADTs and true AADTs.

Table 2. RMSE, U-test in case of model 1,2,3

N. of cluster	model 1		model 2		Model 3	
	RMSE	U-test	RMSE	U-test	RMSE	U-test
1	2443	0.0537	6526	0.1363	5886	0.1237
2	1759	0.0875	6675	0.2913	5158	0.2343
3	1530	0.0623	4319	0.1681	2504	0.0999
4	347	0.0705	781	0.1547	547	0.1104
5	4479	0.0715	15992	0.2334	13938	0.2069
6	448	0.0714	1200	0.1817	855	0.1325
7	3265	0.1725	10979	0.4658	7507	0.3620
8	272	0.0639	1255	0.2568	1015	0.2132
9	647	0.1067	2525	0.3473	1899	0.2772
10	410	0.0759	1842	0.2849	1598	0.2531
11	943	0.1506	4639	0.5198	3227	0.4163

5. CONCLUSIONS AND RECOMMENDATIONS

In this study, the AADT estimation method has been developed and applied using three different models. First, we have applied the traffic pattern clustering analysis which is the traditional AADT estimation model. Second, we estimate the AADT using time-series analysis without considering irregular variation. Third, we estimate the AADT using time series analysis with considering irregular weather variation, Then, the developed models of the AADT estimation have been compared each other using real data obtained in Korea regional roads.

Accuracy for the result of the AADT estimation was highest when we used monthly and daily variables. And in case of applying time series analysis, when weather data was joined to irregular variation, the results of RMSE and U-test were smaller than that not joined to. This result tells how important it is to find out the method of explaining irregular variation. In other word, if we deal with that properly, there is possibility that we can get a good estimation model. So it seems necessary that we study a little more about factors, which relate to irregular variations.

REFERENCES

a) Books and Book chapters

Hyeongjin Noh (1999) **Multivariate statistics**, In Seok-jeong, Korea

Jiseong Yu   Yeohwan Yun (1997) **Modern statistics**, In Pak-young publishing co., Korea

Byeongseon Choi (1995) **Multivariate time series analysis**, In Se-kyeong, Korea

Seongcheol Lee   Yanggyu Pak (1999) **Expeditionary force to SAS statistics**, In 21th century co., Korea

Chunglyeon Kim (2000) **Statistical box named of SAS**, In Dataplus, Korea

Seongmo Yu (1999) **SAS Multivariate statistics**, In Academy of freedom, Korea

Yeohwan (2000) **Grouping study of Traffic monitoring sites for AADT estimation**, In Korea Institute of Construction Technology, Korea

Gary A. Davis (1997), **Estimation theory approach to monitoring and updating average daily traffic**, Minnesota Department of Transportation

b) Journal papers

A. Faghri   J. Hua (1995) Roadway seasonal classification using neural networks, **Journal of computing in civil engineering**

Satish C. Sharma   Pawan Lingras   Fei Xu   Guo X Liv (1999) Neural networks as alternative to traditional factor approach of annual average daily traffic estimation from traffic counts, **TRR**

Satish C. Sharma   Brij M. Gulati   Samantha N. Rizak (1996) Statewide traffic volume studies and precision of AADT estimates, **Journal of Transportation engineering**