A Cluster Analysis of the variations of Macroscopic Fundamental Diagram: A case study in Tokyo metropolitan areas

Thuong Thi LE[1], Kouki SATSUKAWA[2], Kentaro WADA[3], Takashi OGUCHI[4] and Miho ASANO-IRYO[5]

[1]Graduate student, Department of Civil Engineering, The University of Tokyo
(4-6-1, Komaba, Meguro-ku, Tokyo, Japan)
E-mail: lethuong@iis.u-tokyo.ac.jp
[2]Graduate student, Department of Civil Engineering, The University of Tokyo
(4-6-1, Komaba, Meguro-ku, Tokyo, Japan)
E-mail: kouki@iis.u-tokyo.ac.jp
[3]Assistant Professor, Institute of Industrial Science, The University of Tokyo
(4-6-1, Komaba, Meguro-ku, Tokyo, Japan)
E-mail: wadaken@iis.u-tokyo.ac.jp
[4]Professor, Institute of Industrial Science, The University of Tokyo
(4-6-1, Komaba, Meguro-ku, Tokyo, Japan)
E-mail: takog@iis.u-tokyo.ac.jp
[5]Associate Professor, Institute of Industrial Science, The University of Tokyo
(4-6-1, Komaba, Meguro-ku, Tokyo, Japan)
E-mail: m-iryo@iis.u-tokyo.ac.jp

The purpose of this paper is to investigate the variations of Macroscopic Fundamental Diagrams (MFD) for Tokyo Metropolitan region. To achieve this purpose, we use the observed traffic data of one year period (1/1/2012 – 12/31/2012) in 51 areas. Based on initial analysis, it was found that some tentative factors such as time periods, seasons and days of week could impact on MFD's variance. In order to analyze the impact of factors in depth, a cluster analysis was adopted so that sub-MFDs with lower variance are obtained for each area. Then, the reasons that caused reduction in variance will be interpreted by comparing the composition of these factors between clusters. Furthermore, the discussion about the variations of MFD in different areas is conducted.

Key Words : *Macroscopic Fundamental Diagram, empirical analysis, cluster analysis, detector data*

1. INTRODUCTION

In theory, most recent computer simulation models can predict almost anything on a multi-modal transportation network but not in practice[1)]. It is because of the model requires many detail inputs. In addition, driver behavior is unpredictable in different situation. And, when network experiences congestion, it behaves chaotically[1)].

To "shift the modeling emphasis from microscopic prediction to macroscopic monitoring and control"[1)], a macroscopic fundamental diagram has been introduced. Macroscopic fundamental diagram (MFD) represents a relationship between network density and network flow which can be obtained by aggregating density and flow in a whole network. MFD is an extension of fundamental diagram (FD) which had used by traffic engineers traditionally for a purpose of traffic analysis and control. The link fundamental diagram is plotted for one link and it shows the relationship of link density and link flow. **Fig.1** illustrates FD and MFD. As same as FD, based on MFD, network performance can be divided into two regimes: free flow and congestion. Through MFD, network performance can be easily managed. Hence, optimal performance can be achieved once network density is controlled to be equal or smaller than critical value.

Recent studies have shown existence of MFD in difference types of network. A well-defined MFD in

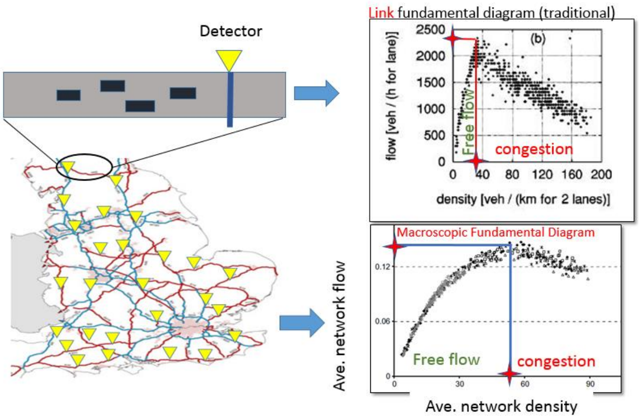


Fig.1 Illustration of difference between FD and MFD[4) 2)]

Yokohama downtown area has been discovered by Geroliminis and Daganzo[2]. Buisson and Ladier[3] have also acknowledged the existence of MFD in Toulouse's surface road and freeway. Additional studies about MFD can be listed such as Geroliminis and Sun (2011)[5], Saberi and Mahmassani (2012)[6], and Saeedmanesh and Geroliminis (2015)[7]. Well-defined MFD with less scatter has been only noticed in Yokohama downtown network[2]. However, majority of MFD studies experiences scatter shapes. This observation would raise a question about reasons why such kind of MFD variation happens. In addition, although characteristics of MFD have been shown in many researches, those studies only focused on several days of traffic data (e.g. 2 days in Geroliminis & Daganzo study, and 3 days in Buisson & Ladier study). To understand comprehensively MFD features, several days of traffic data do not capture everything. Hence, there is a need of a study with long-term data, wider network, and different type of network structure to validate the robustness of MFD's properties and further reliable analysis.

To fulfill the gap of previous studies and to answer above question, the paper investigates on MFD's variation in Tokyo Metropolitan region. This region is divided into fifty-one areas of different wards, cities, and towns with various network structures as well as characteristics. Firstly, initial study was carried in 50 areas according to tentative factors such as monthly, days of week, and time periods. This analysis has shown that MDFs of center Tokyo region and suburban Tokyo region present some dissimilar characteristics between them. The abundance of characteristics of each area would provide a better understanding about MFD's variation and how is its scatter level affected by tentative factors. Then, to understand deeply about initial analysis observations, cluster analysis was adopted. Its results showed that different area presents significantly various features of MFD. Final composition analysis of tentative factors also

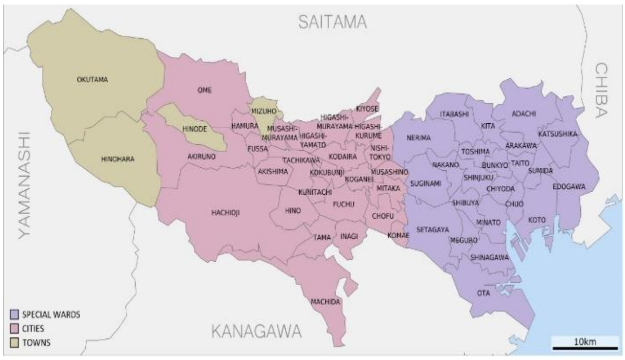


Fig.2 Map of Tokyo Metropolitan region

helped to draw a conclusion that even located in a same region, MFD's variance of individual area is significantly different. The finding of this study would add more information to MFD study before it could be used for purpose of real time traffic control.

The remaining of this paper is organized as follows: Chapter 2 describes Tokyo Metropolitan region and traffic data that was used for the analysis. Chapter 3 presents initial analysis of MFD for 51 areas. The first half of Chapter 4 includes the details of cluster analysis methodology and its results, and composition analysis is presented in the last part of Chapter4. Finally, Chapter 5 gives some discussion regarding to the results of this paper.

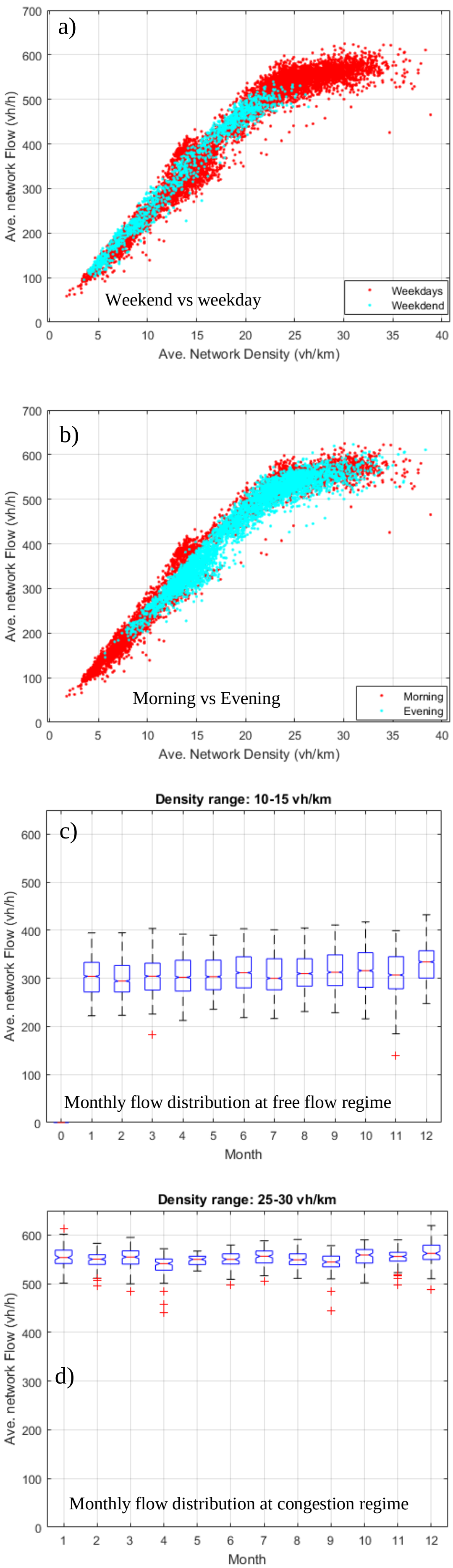
2. DATA

2.1 Tokyo Metropolis

Tokyo Metropolis is a large region with the combination of 53 areas including 23 special wards, 26 cities 3 towns and 1 village. According to Tokyo Metropolitan Government's report, the population of Tokyo was estimated to be 13.216 million in October 2012. And it was about 10% of Japan's total population. **Fig.2** shows the map of Tokyo Metropolis region.

2.2 Data

Hourly aggregated vehicle-kilometers and existing number of vehicles aggregated on the major streets/avenues for each of 51 areas (except 2 local villages) are provided by Tokyo Metropolitan Police Department for one year (366 days from January 1, 2012 to December 31, 2012). One day traffic data for each area consists of 24 data points consequently. Vehicle-kilometers and number of vehicles are able to be normalized and be converted to traffic volume (in vehicles per hour) and traffic density (in vehicles per kilometer) with information of total length of the major streets/avenues for each area. The major streets/avenues are defined as the road network equipped with detectors utilized by the central traffic



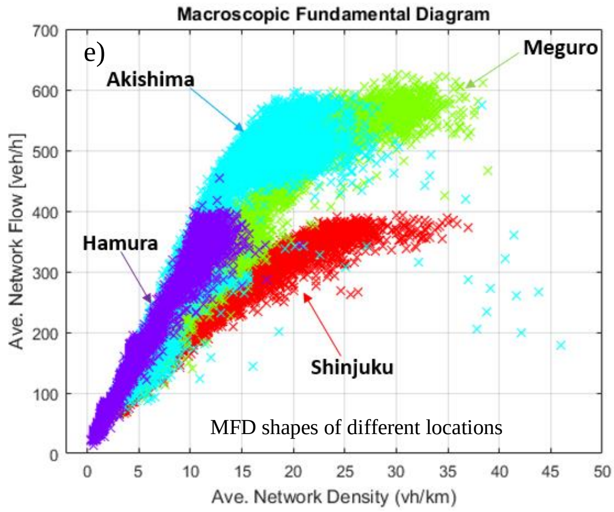


Fig.3 Example of initial analysis results (a) Weekend vs Weekday, b) Morning vs Evening, c) Monthly flow distribution at free flow regime, d) Monthly flow distribution at congestion regime, e) MFD shapes in different locations)

control center of Tokyo Metropolitan Police Department. Overall total length of streets/avenues for 51 areas is 515 km.

3. INITIAL ANALYSIS

First of all, initial analysis was done for understanding the features of the data. In **Fig.3**, different areas show dissimilarities regarding to average network flow distribution in term of tentative factors and MFD shape.

Based on **Fig.3a**, it is clear to see a significant difference in the distribution of network flow between weekend and weekday. At free flow regime of MFD, traffic flow of both weekend and weekday are presented. However, at congestion regime, weekday flow is completely dominant. MFD also shows its variation in term of different time periods (**Fig.3b**). Morning traffic flow has higher variation compared to evening traffic flow. In addition, more outliers are expected to see in morning time. In **Fig.3c** and **Fig.3d**, regarding to monthly factor, as boxplots show a monthly distribution in 2 small density bins (one in free flow regime, one in congestion regime), it is clear to see that distribution levels of flow in different months are lightly different. However, we believe that MFD's variance in different areas would behaves variously according to monthly factor. For example, in months of December and January, intensive traffic activities are expected to observe in sub-urban areas. It is obvious to say that MFD shapes of areas in different locations are dramatically dissimilar (**Fig.3e**). High traffic flow and density value are observed in Meguro ward and Akishima city. However, Hamura's MFD shape seems not to have congestion regime. Shijuku ward has a very neat MFD form.

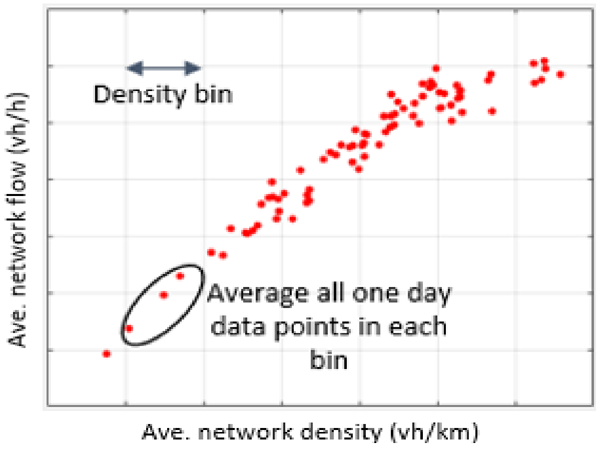


Fig. 4 Data point illustration

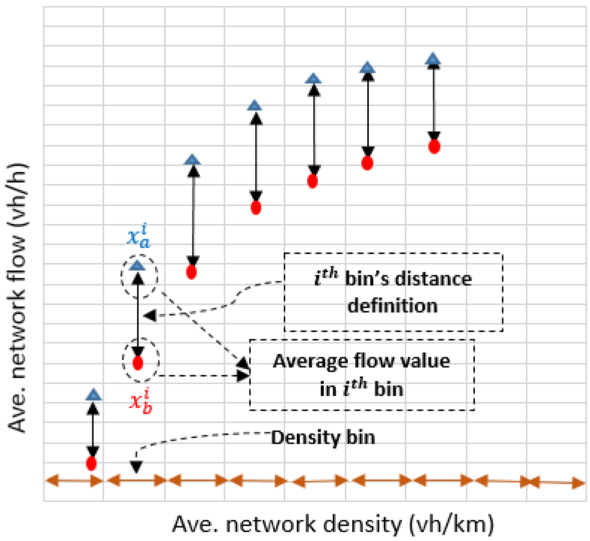


Fig. 5 Illustration of distance definition

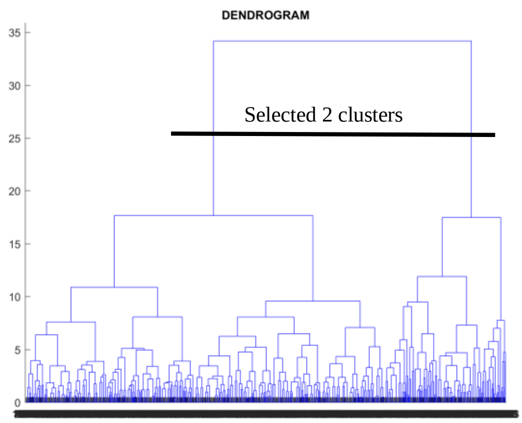


Fig. 6 Dendrogram of cluster analysis

The observations of MFD variance regarding to tentative factors inspired us to explore more deeply about factors' impact on MFD variation.

4. CLUSTERING RESULTS

4.1 Clustering Analysis

a) Clustering Methodology

The study was carried out in 51 areas in total with time period is from 6:00 to 21:00. By applying clustering analysis, MFDs' scatter level in each sub-MFD is expected to be smaller than original one. This method was adopted to understand how variance of macroscopic fundamental diagram in individual area is affected by different factors. In addition, similar characteristics of tentative factors would be found in one sub-MFD but would be very distinguished with the factors not belonging to the other. Totally, there were 51 areas used for cluster analysis in this paper. However, there was an error found with data in Akiruno city, so we could not do any analysis for this area.

As already mentioned in Chapter 2, one day traffic data consisted only 24 points. This limitation in data would affect cluster analysis results. Hence, an idea to increase the number of data was considered to make sure that at least there is one data point of one day in each bin. Consequently, linear interpolation methodology was adopted. And, the number of data was increased by two with 30-minute period. Traffic data shows that there are two peaks in its time series. Hence, we separated the data into two-peak groups before applying cluster analysis.

Cluster classification was done based on average flow variance in density bin for each data x_a, x_b which is calculated by taking average value of average network flows in each density bin (**Fig.4**). The maximum of density bin was calculated by dividing the maximum density value to density range unit (Unit of density range is 2vh/km for this paper).

Fig.5 illustrates the definition of distance between data. To define the distance between x_a, x_b, let:
$I_{a,b}^i$ is the indicator function of x_a, x_b in i^{th} bin.

$$I_{a,b}^i = \begin{cases} 1 \; if \; x_a^i > 0 \; and \; x_b^i > 0 \\ 0 \qquad\qquad otherwise \end{cases}$$

x_a^i is averaged flow value of x_a in i^{th} bin
x_b^i is averaged flow value of x_b in i^{th} bin
$D_{a,b}$ *is* distance between data x_a and data x_b

$$D_{a,b} = \sum_i I_{a,b}^i * \left| x_a^i - x_b^i \right|$$

The distance between x_a, x_b of i^{th} bin which is fail to have $x_a^i > 0 \; and \; x_b^i > 0$ was ignored.

We used Ward[8]) method to classify clusters based on defined distance. Results of cluster analysis can be described as dendrogram **Fig. 6**. The number of clusters can be chosen according to each area's dendrogram. In this paper, 2 clusters were selected for the purpose of composition analysis.

b) Statistical Analysis

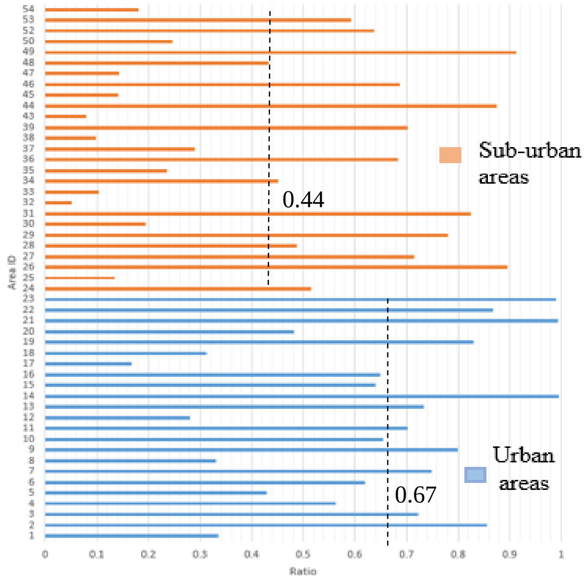


Fig.7 Ratio of sample size between 2 clusters

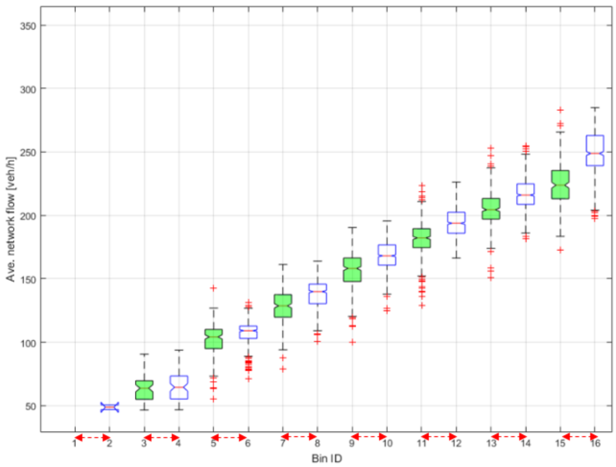


Fig.8 Boxplots for parallel bins between two clusters (Chiyoda Ward)

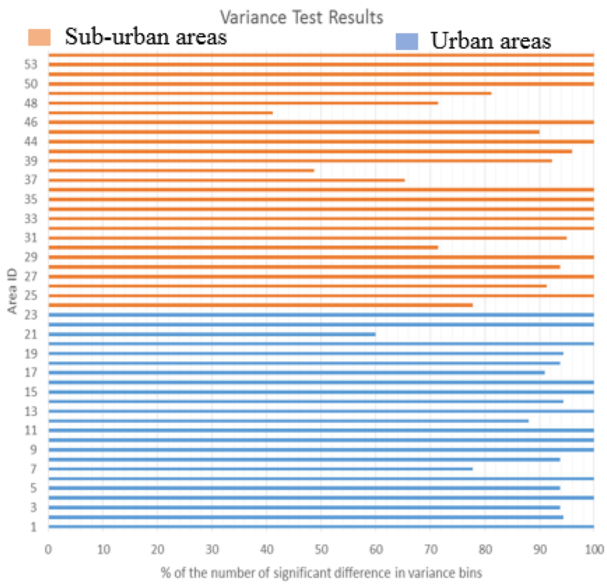


Fig.9 Variance test results

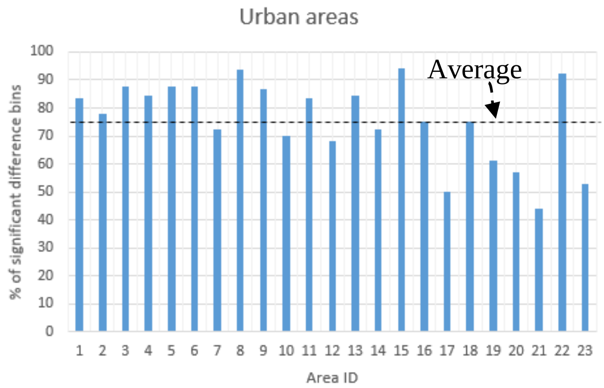


Fig.10 t-test results for urban Tokyo areas

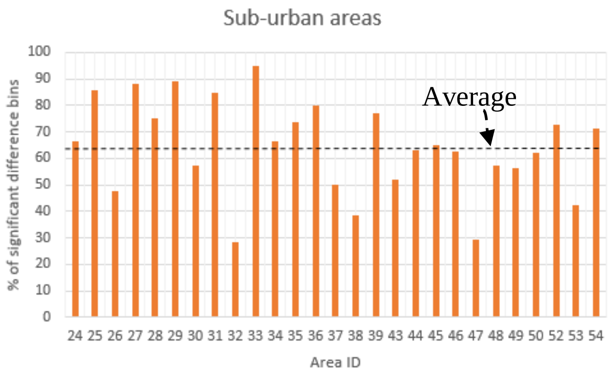


Fig.11 t-test results for sub-urban Tokyo areas

Statistical tools are needed for our investigation on the decision of whether or not sub-MFDs are significantly difference. If sub-MFDs in one area are statistically significant difference, that area will be chosen for composition analysis.

First ratio between clusters' sample size was checked to compare data sample size between sub-MFDs. In **Fig.7**, the results shows a quite-equal sample size with average ratio value 0.67 in urban region. The same outcome does not clearly present in areas of sub-urban region. The ratio of sample size between sub-MFDs in this is area is only 0.44.

In **Fig.8,** red arrows represent for average network flow at the same density bin. White and green colors present for different clusters. In Chiyoda ward, boxplots of average network flow at same density bin show very different in variance as well as mean between sub-MFDs. Hence, to statistically validate if this observation is exist for the rest areas, F-test and t-test are conducted.

Based on **Fig.9**, F-test results show that sub-MFDs' variances significantly different in most of areas. The variance test results showed that, 100% parallel bins in some areas are statistically significant difference in variance. This result is understandable since in working time periods, expectation of congestion or typical traffic activities still happen throughout the

Fig.12 Areas with worked well cluster analysis (t-test results > 80%)

Fig.13 Areas with failed cluster analysis (t-test results < 50%)

year.

In **Fig.10**, two-sample t-test results for paired bins in urban Tokyo areas are at very high percentage which means two clusters are significantly different and cluster methodology works well in those areas. Results show that 17 out of 23 wards in former Tokyo area have more than 70% of significant different bins. Typically, some area such as Koto, Suginami, and Katsushika have highest t-test results with more than 90%. For sub-urban areas (**Fig.11**), the distribution of t-test results dramatically vary compared to center Tokyo areas. The number of areas have more than 70% of significant different bins are eleven. There are some areas with t-test results lower than 50% like Komae, Akishima, Kokubunji, and Mizuho. Areas with t-test results smaller than 50% mean cluster analysis was failed in those areas. The reason why there is such difference in t-test results between center Tokyo region and sub-urban region could be an extremely variation in sample size between clusters. Therefore, statistical test accept a hypothesis which is means of parallel bin are equal.

4.2 Clustering Results

Observation from clustering results can be summarized as following:

- According to t-test outcome, areas in center Tokyo region present better cluster results compared to the cities and towns in sub-urban region. According to **Fig.10** and **Fig.11,** averages of t-test results are 75.6 % and 64.3 % in center Tokyo region and sub-urban Tokyo region respectively.

- In sub-urban region, the distribution of t-test results is dramatically dissimilar between areas compared to center Tokyo region. There are some areas with t-test results lower than 50% like Komae, Akishima, Kokubunji, and Mizuho which means cluster analysis was failed in those areas. This observation can be understandable due to distinguished characteristics between locations in sub-urban Tokyo region such as small area (Komae), small and mountain area (Mizuho).

- Observation from **Fig.12** and **Fig.13** shows that cluster analysis works quite well in center areas. However, cities and towns in sub-urban areas seem to have lower performance of cluster analysis.

4.3 Composition Analysis

Composition analysis was used for interpreting tendencies of tentative factors like months, days of week, and time periods which could impact on MFDs' variance. For monthly factor, there are 12 components which are month from January to December. Days of week factor consists of weekend and weekday component. Morning (6:00 – 13:00) and evening (13:00-21:00) are elements of time period factor. The method to choose an area for composition analysis was decided based of the statistical test results. Hence, those areas which t-test results were larger than or equal 80% will be chosen. An exception was made for Okutama area since it was the only town that t-test result is larger than 70% and Meguro ward.

Let:

X_1^i is sample size of factor's i^{th} component in cluster 1

X_2^i is sample size of factor's i^{th} component in cluster 2

X_k is total sample size in cluster k (k= 1,2)

$$Ratio_1^i = \frac{X_1^i}{X_1}$$

$$Ratio_2^i = \frac{X_2^i}{X_2}$$

Percentage of difference of factor's components:

$$\% = 100 * \left(\left|Ratio_1^i - Ratio_2^i\right| / P\right)$$

Where P :

$$P = \frac{1}{12} \quad \text{for monthly factor}$$

$$P = \frac{1}{7} \quad \text{for days of week factor}$$

$$P = \frac{1}{2} \quad \text{for time period factor}$$

If the percentage of difference of factor is equal or larger than 80%, it could be confident to conclude that factor has significant impact on MFD variance.

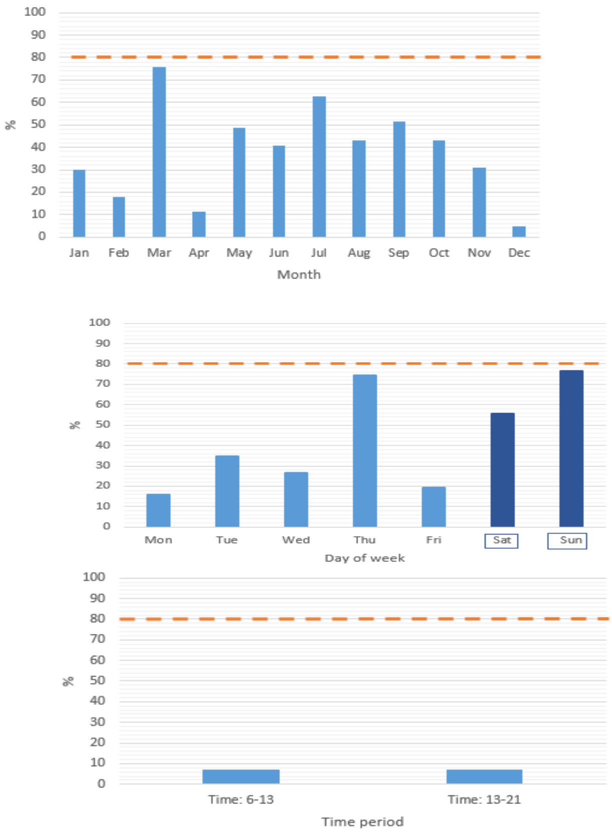


Fig.14 Chiyoda’s composition analysis results

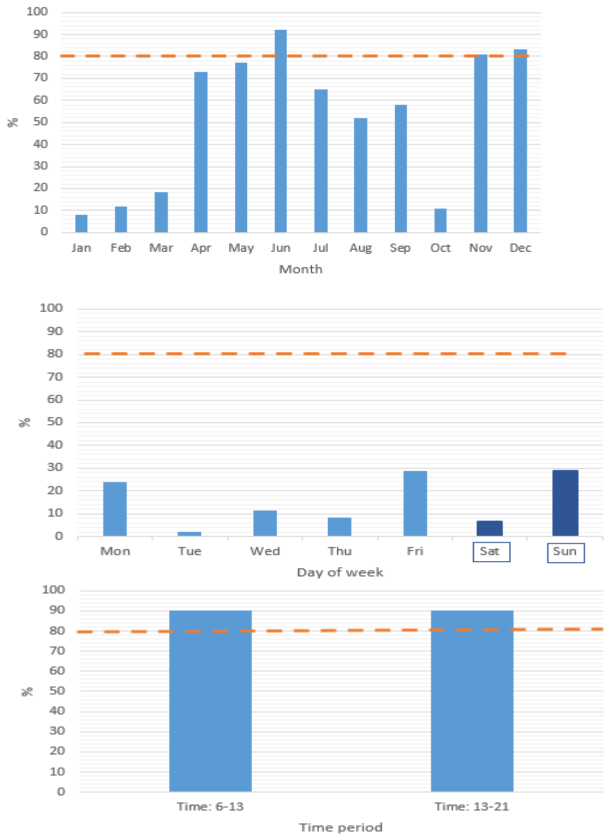


Fig.16 Suginami’s composition analysis results

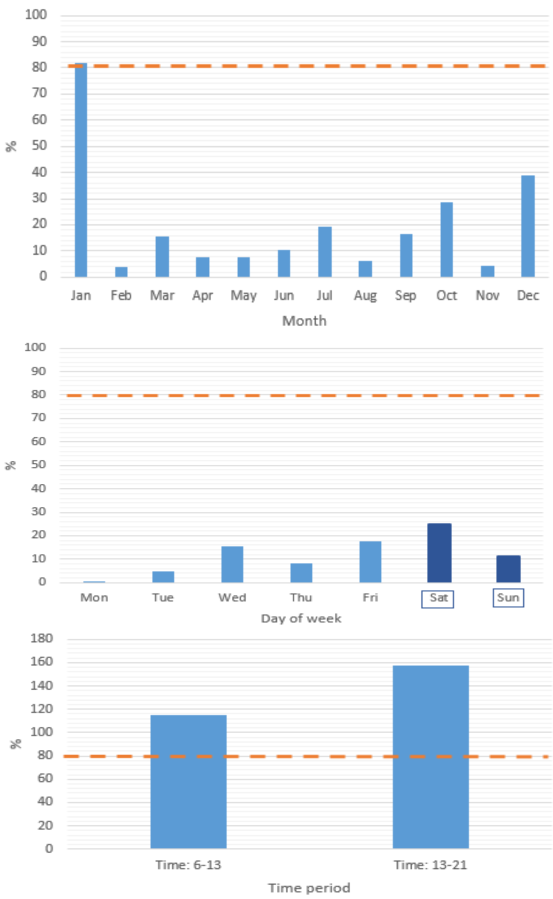


Fig.15 Meguro’s composition analysis results

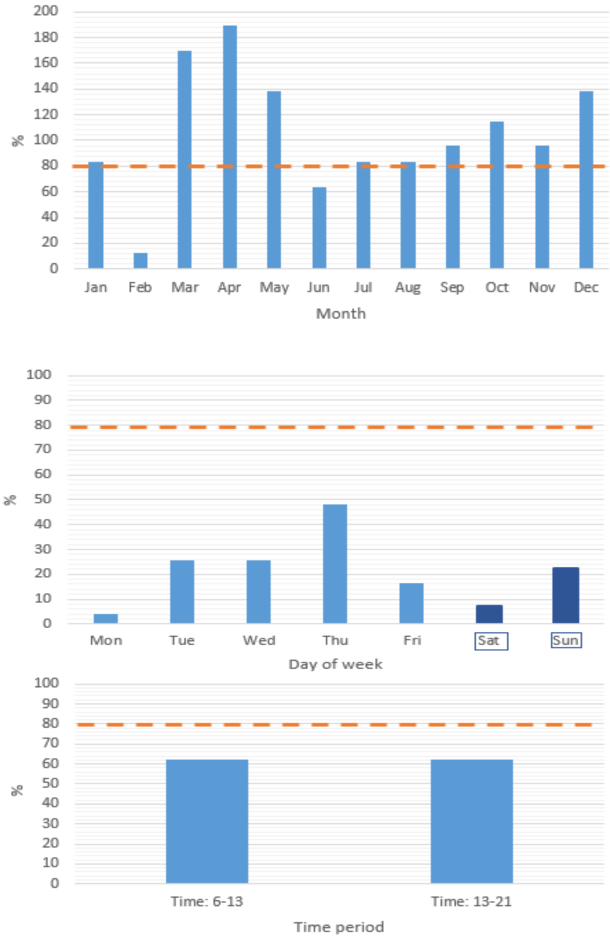


Fig.17 Tachikawa’s composition analysis results

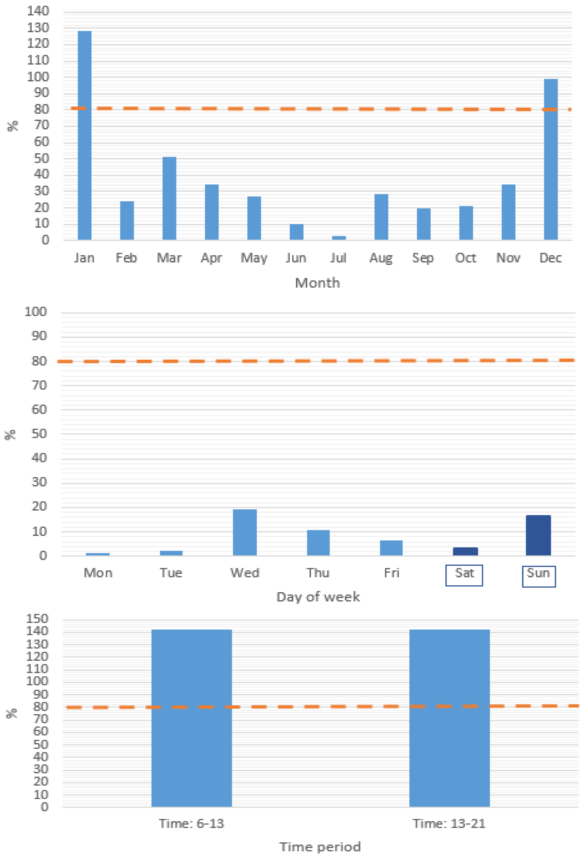


Fig.18 Oume's composition analysis results

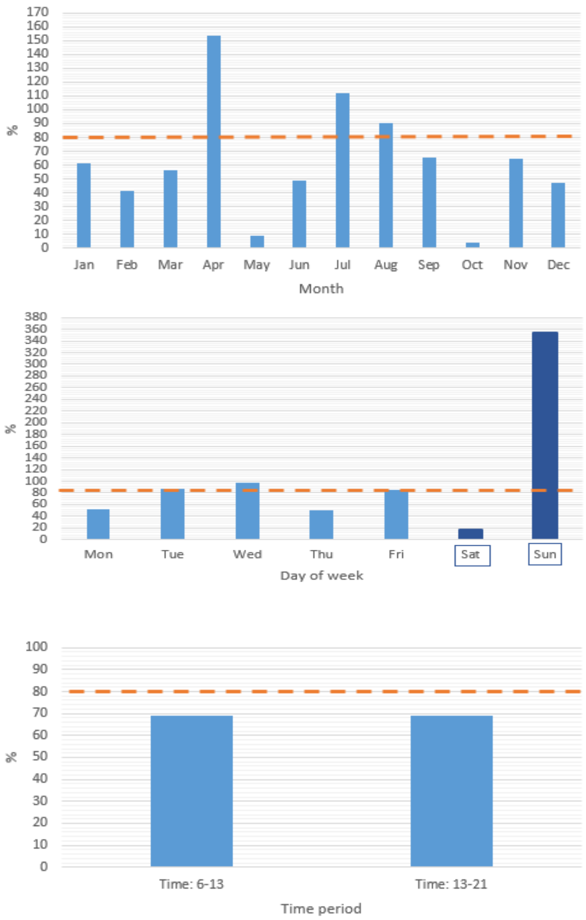


Fig.19 Machida's composition analysis results

Fig.20 Okutama's composition analysis results

Composition analysis results of presentative areas are presented for several areas. Note that for day of week factor, dark blue bars represent for weekend (Saturday and Sunday).

Chiyoda ward **(Fig.14)**

Located at center location of former Tokyo region, Chiyoda is famous with many historical places and attracts many tourists. With average of the percentage of weekend difference is 66%, weekend factor has the strongest impact on cluster results compared to others. Monthly factor also shows its effect according to composition analysis results in Mar or July. However, checking for overall 12 months shows that monthly factor does not have strong meaningful influence in network performance variance. There is not significant impact of time period factor on the result of cluster analysis.

Meguro Ward **(Fig.15)**

Compared to Chiyoda ward, time period factor with more than 100% of difference percentage is the most influent factor that impact MFD variance in Meguro ward. Interestingly, monthly factor only shows its strong impact in January with more than 80% of difference percentage. There should be expected outlier traffic activities in this month. Weekend does not show its effect in cluster results.

Suginami ward **(Fig.16)**

Suginami is located near to the boundary between former Tokyo region and sub-urban Tokyo region. There are interesting points of composition analysis results in this area. Observation of season effect is clearly found in monthly factor with the average of 73% in difference for months except Jan, Feb, Mar and Oct. In addition, time period factor is also a factor with high influence in cluster results. Time period of percentage difference is about 90%. Weekend does not show any significant impact on cluster results in Suginami area.

Tachikawa city **(Fig.17)**

With 106% in average of percentage difference, it is clearly to see that monthly factor has the strongest effect in cluster results. Spring season seems to show the most outstanding of percentage difference compared to other seasons. One of a reason for this strong tendency would be because there is a famous park named Showakinen in Tachikawa area. This park is not only an attractive place for foreigners but also for Japanese people. Time period also shows a significant impact on cluster results with more than 60% of percentage in difference. However, day of week factor does not have any meaningful influence in Tachikawa area.

Oume city **(Fig.18)**

It is clear to see that time period is the strongest factor with 140% of difference percentage. January and December are two outstanding months that impact strongly in cluster results. The average of percentage difference for these two month is about 114%. Significant effect on MFD variance is not found in day of week factor.

Machida city **(Fig.19)**

Located near to Kanagawa area, Machida is also a city with very high result in t-test (>90%). Interestingly, it is shown that monthly, day of week and time period factor all have significant impact on cluster results. With 350% in percentage of difference, weekend is a factor which has the strongest impact on cluster outcome. Composition analysis result also indicates that individual weekday has significant impact on cluster results such as Tuesday, Wednesday and Friday. This is an extremely interesting point that only noted in Machida area. In addition, monthly factor shows its impact very randomly. Several months such as April, July and August have outstanding effect with percentage of difference is more than 80%. Furthermore, with nearly to 70% in percentage difference, time period also presents its significant

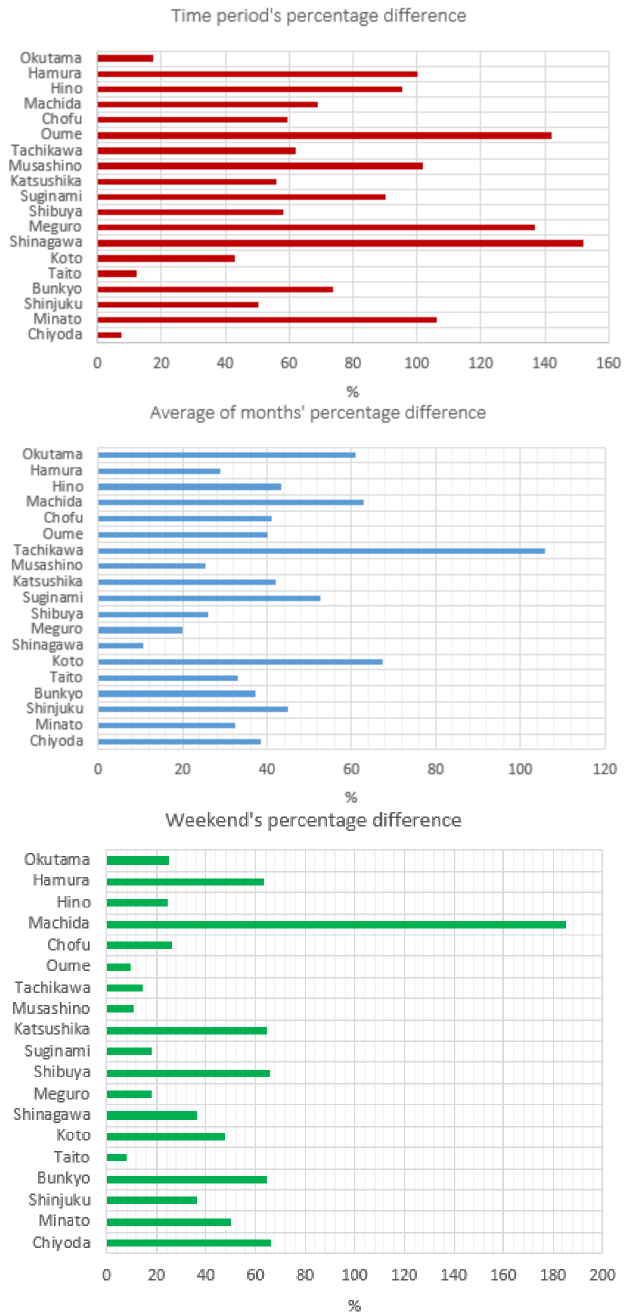
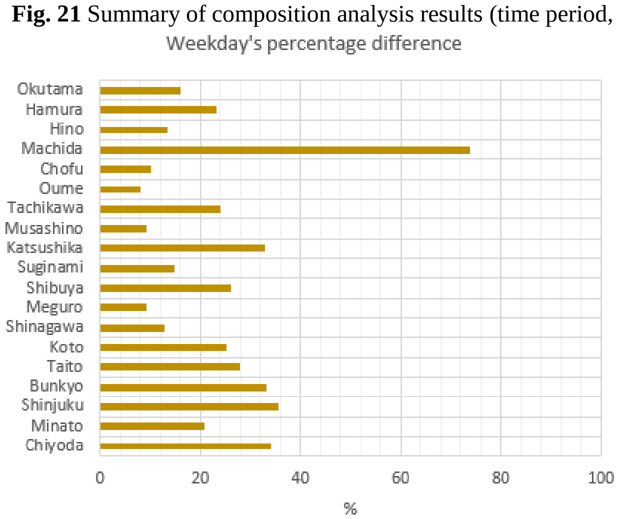


Fig. 21 Summary of composition analysis results (time period,



months, weekend, and weekday)

influence on cluster results. Composition analysis would give a hint that there should be combination effect between the three factors: months, day of week and time periods.

Okutama town (**Fig.20**)

Obviously, time period and day of week factors do not have significant impact on cluster results. However, for monthly effect, July, September and October are three months have highest percentage of difference which ranges from 76% to 172%. These months have the strongest influence in MFD variance.

Summary of composition analysis

Accoring to **Fig.21**, time period is a factor which has the strongest impact on cluster results. Its impact is randomly found distribution over Tokyo metropolitan area with percentage of difference range is distributed from lower than 10% to more than 140%. Compared to time period factor, month factor also shows its influence in MFD variance. Considering for individual month, it would be seen clearer impact in different area. Overall, day of week factor does not have significant effect except typical Machida city. The summary includes 17 areas which have t-test results larger than or equal 80%, Meguro ward and Okutama town.

5. DISCUSSION

Results of this study have shown how different factors like monthly, days of week and time period factor can affect the variance of macroscopic fundamental diagram in overall Tokyo Metropolitan region. Clearly, the level impact of different factor is randomly distributed from area to area. This observation provides additional evidence for understanding MFD variance. Depending on the characteristics of individual area like population distribution, network characteristic, or the combination information of network and demand, the variance of MFD would be affected.

To understand deeply about MFD's variance, additional study is needed to discover possible factors that would impact on the scatter level of MFD shape. According to initial analysis, it has been noted that there are some differences in MFD's properties between free flow regime and congestion regime regarding to flow distribution and MFD shape. In free flow stage, demand pattern seems to be the most important factor that impacts on these properties. However, in congestion regime, a combination mechanism between demand and supply patterns could affects the distribution of network flow. The knowledge about such differences between two regimes in macroscopic fundamental diagrams should be studied. Therefore, in further work we would like to focus on a study of MFD variance at free flow and congestion regime. Then, comparison between those regimes could be considered. In addition, combination analysis would be done to acknowledge the existence of tentative factors' relation between themselves.

ACKNOWLEDGMENT

We would like to acknowledge the Tokyo Metropolitan Police Department for providing us the data.

REFERENCES

1) Daganzo, C.F. : Urban gridlock: Macroscopic modeling and mitigation approaches, *Transportation Research Part B41*, pp. 49-62, 2007
2) Daganzo, C.F. and Geroliminis N. : An analytical approximation for the macroscopic fundamental diagram of urban traffic, *Transportation Research Part B42*, pp. 771-781, 2008
3) Buisson, C. and Ladier, C. : Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2124, pp. 127-136, 2009
4) Nagel, K., Wolf, D.E., Wagner, P. and Simon, P.: Two-lane traffic rules for cellular automata: A systematic approach. *Physical Review E*, *58*(2), pp.1425, 1998.
5) Geroliminis, N. and Sun, J. : Hysteresis phenomena of a macroscopic fundamental diagram in freeway network, *Transportation Research Part A*, 45(9), pp. 966-979, 2011
6) Saberi, M. and Mahmassani H.S. : Exploring the properties of network-wide flow-density relation in freeway networks, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2315, pp. 153-163, 2012
7) Saeedmanesh, M. and Geroliminis, N. : Clustering of Heterogeneous Networks with Directional Flows Based on "Snake" Similarities, *Transportation Research Board 94th Annual Meeting* No. 15-1354, 2015
8) Ward Jr, J. H. : Hierarchical grouping to optimize an objective function, *Journal of the American statistical association*, 58(301), pp. 236-244, 1963