

ANALYSIS OF INFLUENCING FACTORS IN OVERSIZED TRUCK ROUTE CHOICE USING GPS TRAJECTORY CLUSTERING

Yue MA¹, Jan-Dirk SCHMÖCKER², Wenzhe SUN³, Satoshi NAKAO⁴, Tadashi
YAMADA⁵, Ali Gul QURESHI⁶

¹ Doctoral Candidate, Graduate School of Engineering, Kyoto University
(Kyoto daigaku-katsura, Nishikyo-ku, Kyoto 615-8530, Japan)
E-mail: ma.yue.55t@st.kyoto-u.ac.jp

² Member of JSCE, Associate Professor, Graduate School of Engineering, Kyoto University
(Kyoto daigaku-katsura, Nishikyo-ku, Kyoto 615-8530, Japan)
E-mail: schmoecker@trans.kuciv.kyoto-u.ac.jp (Corresponding Author)

³ Member of JSCE, Postdoctoral Researcher, Graduate School of Engineering, Kyoto University
(Kyoto daigaku-katsura, Nishikyo-ku, Kyoto 615-8530, Japan)
E-mail: wz.sun@trans.kuciv.kyoto-u.ac.jp

⁴ Member of JSCE, Assistant Professor, Graduate School of Engineering, Kyoto University
(Kyoto daigaku-katsura, Nishikyo-ku, Kyoto 615-8530, Japan)
E-mail: nakao@trans.kuciv.kyoto-u.ac.jp

⁵ Member of JSCE, Professor, Graduate School of Management, Kyoto University
(Kyoto daigaku-katsura, Sakyo-ku, Kyoto 606-8317, Japan)
E-mail: yamada.tadashi.2x@kyoto-u.ac.jp

⁶ Member of JSCE, Associate Professor, Graduate School of Engineering, Kyoto University
(Kyoto daigaku-katsura, Nishikyo-ku, Kyoto 615-8530, Japan)
E-mail: qureshi.aligul.4c@kyoto-u.ac.jp

In contrast to ordinary vehicles, the mobility of oversized trucks is often limited by their large size. They thus may have additional requirements on road types, road widths, and the turning radius at the intersection when traveling. Therefore, this study explores the unique needs and preferences of oversized trucks in route choices with a focus on trip and road network characteristics. GPS trajectory data that come from the Osaka-Kobe region having numerous ports and freight stations are used in this research. The relatively concentrated origins and destinations give us more opportunities to analyze them with an aggregated perspective. Using GPS trajectory data of large and extra-large trucks, we group the trajectories by their origins and destinations at a spatial level of square meshes with 500 m side length. For the trajectories traveling over the same pair of origin-destination (OD) meshes, several clusters are obtained by a QuickBundles algorithm based on geographical configuration. Sampling techniques are used to equalize the number of input points of each vehicle trajectory to the algorithm. The optimal number of clusters is based on the Silhouette Coefficient and automatically determined by the algorithm. Finally, a binomial logit model is used to identify the factors that significantly influence the route choice of oversized trucks and quantify their impacts. The results indicate the preference of oversized trucks for normal roads without central belts with fewer turns, fewer intersections, and high average speed. The findings of this study will help to understand the unique preferences of oversized trucks and may also provide insights on freight road network design and optimization.

Key Words : oversized truck, route choice, GPS data, trajectory clustering

1. INTRODUCTION

The use of oversized trucks for cargo transportation is becoming increasingly frequent. Statistics released by the Ministry of Land, Infrastructure,

Transport and Tourism show that in Japan in 2020, of the 3,786,998 thousand tons of cargo transported by automobiles, 1,739,446 thousand tons were transported by large trucks or extra large trucks with a

maximum load capacity of 6.5 tons or more, accounting for 45.9% of the total¹⁾.

Compared with small vehicles or ordinary trucks, oversized trucks have some apparent characteristics, such as larger size, larger weight, larger braking distance, more blind spots, difficulty in turning, more noise, and higher fuel consumption and emissions²⁾. In addition, oversized trucks are more likely to cause severe traffic congestion and traffic safety problems on their routes³⁾. Oversized trucks have more concentrated trip origins and destinations. Their physical and operational characteristics differentiate them from ordinary vehicles in route choice. They may have additional unique preferences or requirements for the path to be used, such as road types, road widths, and the turning radius at the intersection. Oversized trucks may prefer wider roads rather than narrower ones, perhaps more inclined to turn at larger intersections than small ones. They are also more limited to size and weight restrictions by traffic management. The special characteristics of oversized trucks make it necessary and meaningful to conduct dedicated studies on their route choices.

For the classification of truck types, different agencies have different standards. Oversized trucks mentioned in this study include large trucks and extra large trucks. Large trucks refer to freight vehicles with 1) gross vehicle weight of 8 tons or more, or 2) maximum load capacity of 5 tons or more and three axles or less, or 3) gross vehicle weight of 25 tons or less and four axles, or 4) semi-trailer tractors with no attached vehicle and three axles. Extra large trucks refer to freight vehicles with four axles or more, other than those classified as large trucks. The above is also how the expressway management company, East Nippon Expressway Company Limited, categorizes large and oversized trucks to determine different expressway toll rates⁴⁾.

The remainder of this paper is organized as follows. The next section provides a literature review of studies related to oversized truck route choice and points out the contribution of this research. Section 3 introduces the trajectory and road network data used in this study and conducts a descriptive analysis. Section 4 provides an introduction to the binomial logit model and discusses its applicability to this study. Results are given and interpreted in Section 5. Finally, conclusions and recommendations are given in the last section.

2. LITERATURE REVIEW

With frequent freight transportation by trucks, the attendant issues of efficiency, safety, and emission is-

ues have attracted attention. Among the many research topics on trucks, the route choice problem of trucks has been widely studied.

There are two reasons for studying truck route choice's characteristics and influencing factors. The first is for explanation and prediction. When only the origin and destination of the trip are known, if one wants to infer a more accurate trip path, it is necessary to conduct more in-depth research on the route choice of the truck. Traditionally, route estimation for oversized trucks is based on the shortest distance or the shortest required time. However, road factors also significantly impact the route choices of trucks or truck drivers⁵⁾. In addition, trip characteristics can also affect the route choice, such as travel distance⁶⁾. Some studies have also verified the influence of vehicle characteristics on route usage. Container trucks are more inclined to use large-sized intersections for turns, while time-sensitive trucks are more inclined to use highways⁷⁾. Expected travel time, use of different road types (e.g., expressway), and the number of turns are also essential factors for truck drivers to consider when choosing their routes⁸⁾. More reasonable explanations and predictions can be obtained when these factors are incorporated into the model.

The second and follow-on reason is to provide policy recommendations, i.e., road network optimization and truck management methods, on the premise that the behavior of trucks can be explained and predicted. Heavy traffic jams and accidents may occur if trucks frequently use normal roads. Some researchers established a discrete choice model for this problem, hoping to promote trucks to choose expressways through road transformation on bottleneck sections and expressway fee optimization⁹⁾. Geometric factors of the road network, such as road width, the number of lanes, and the size of intersections, significantly impact truck route choice¹⁰⁾. In the case of road height restrictions, weight restrictions, and intersection turning restrictions, the route selection of trucks will also change significantly¹¹⁾.

Regarding data sources, travel route survey⁵⁾, truck trip permit application data¹¹⁾, and GPS trajectory from vehicle probe⁶⁾ are the most commonly used data acquisition methods. GPS trajectories allow the analyst to know the vehicle's precise location in time and space, including the type of road used after map matching.

There has been extensive research on trucks' route choice behavior. However, this research is more concerned about how the vehicle, trip, and road characteristics affect a truck's route choice when faced with a set of unique routes between an OD (origin and destination) pair. And different from the data sources of previous studies, we used data from the Kansai region of Japan.

3. DATA

GPS trajectory data used in this study were obtained from the Osaka-Kobe region. From September to November 2020, the trajectories of 21,400 trucks were recorded, of which 82.76% were large trucks and 17.24% were extra large trucks. A total of 2,289,114 trips were collected, of which 14.08% were trips by extra large trucks. Table 1 and Table 2 show the number of vehicles and number of trips by vehicle type and month.

Table 1 Number of vehicles in GPS trajectory data.

Vehicle Type	Number	Percentage
Large Vehicle	17710	82.76%
Extra-large Vehicle	3690	17.24%
Total	21400	100.00%

Table 2 Number of trips by vehicle type and month.

Month in 2020	Large Vehicle	Extra-large Vehicle	Total
September	639483	101463	740946
October	683173	112338	795511
November	644116	108541	752657
Total	1966772	322342	2289114

The trips are concentrated in the central Kansai urban area, as shown in Fig.1.

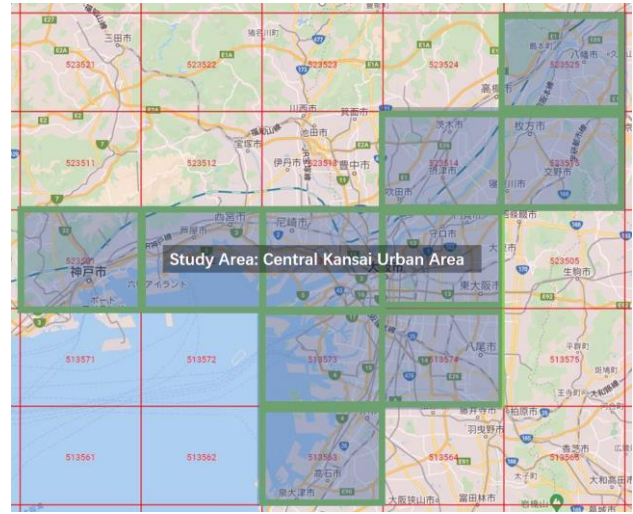


Fig.1 Osaka-Kobe region and central study area.

The trip distance and trip duration are widely distributed. As can be seen from Fig.2 and Fig.3, the distance of most trips is within 200km, and most of the duration of a single trip is within 240 minutes.

Based on the start time of trips, Fig.4 and Fig.5 respectively show the number of trips for all vehicles by date and the number of trips for all vehicles by hour.

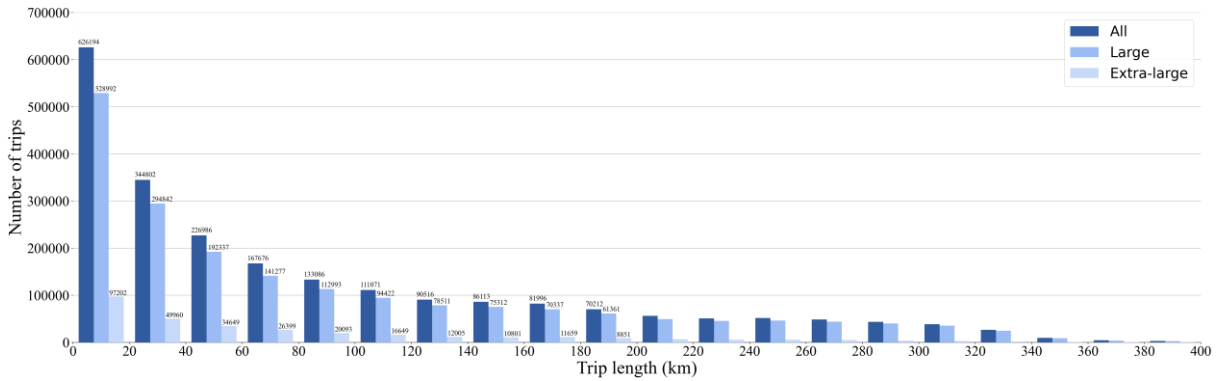


Fig.2 Trip length distribution.

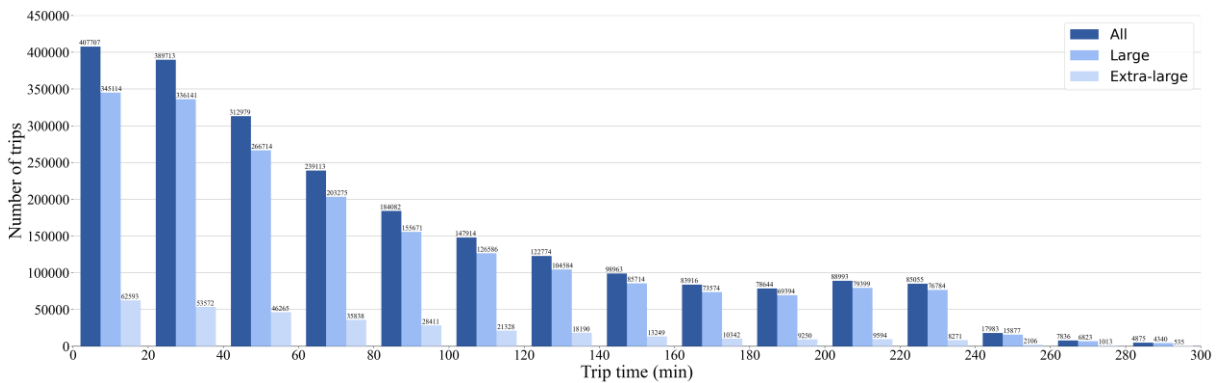


Fig.3 Trip duration distribution.

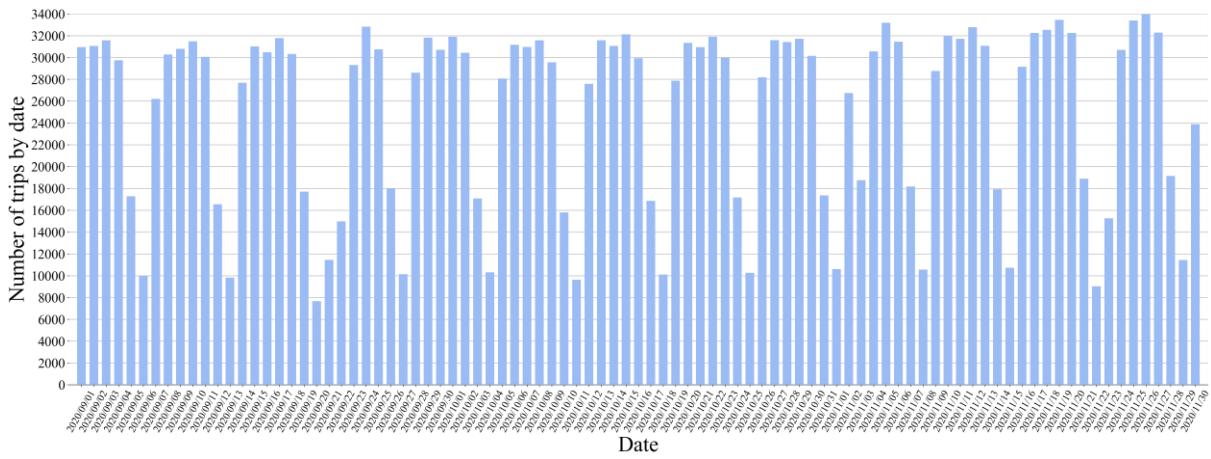


Fig.4 Number of trips for all vehicles by date.

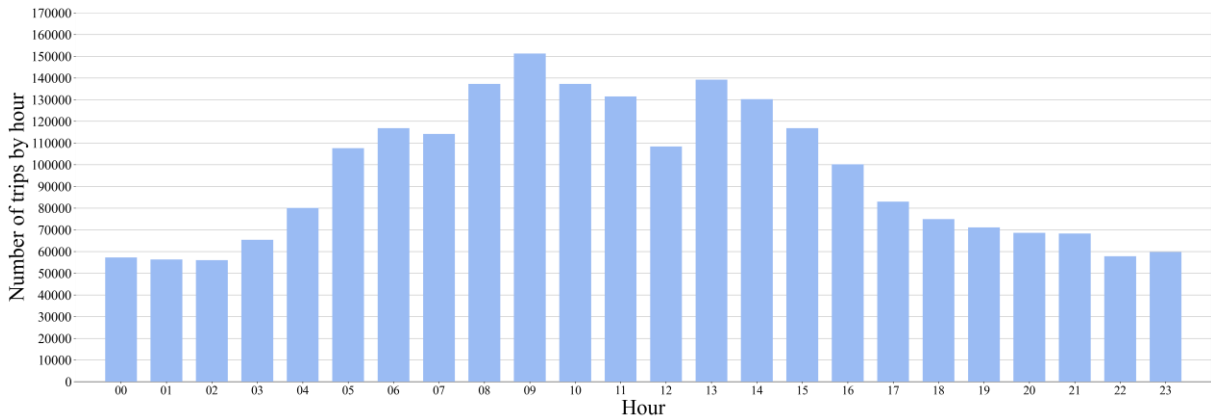


Fig.5 Number of trips for all vehicles by hour (based on the start time of trips).

For the start time of trips, relatively few trips are generated on weekends and holidays, and week-long periodically change is formed in the number of trips. Within a day, there are trips generated each hour, but the number of trips during the day is relatively large, and the obvious trip peaks at 8-10 am, and 1-3 pm can be seen in Fig.5.

Using GPS trajectory data of large and extra large trucks, we group the trajectories by their origins and destinations at a spatial level of 500 m mesh. The relatively concentrated origins and destinations give us more opportunities to analyze them with an aggregated perspective. From Fig.6, the heat map of origin distribution and Fig.7, the heat map of destination distribution, it can also be found that the trips are more concentrated between coastal ports and cargo warehouses and transshipment terminals.

In order to use a more heterogeneous dataset in the subsequent analysis of clustering and route choice influencing factors, we filtered out some ODs using the following rules, keep 500 m mesh ODs for which:

1. there are at least 50 observations (trips) with this OD;
2. the maximum trip length is no longer than 30 km, and the minimum trip length is no shorter than 2 km;
3. at least 5% of trips are generated by extra

large vehicles;

4. there are trips that start at at least six different hours of the day.
5. there are at least three route kinds of percentage of the highway;

The above rules are to ensure that, when studying the factors influencing route choice, there are enough trips with varying attributes in the dataset. The trips used are not too short or too long that the routes are too similar. Moreover, ensure that there are enough extra large truck trips, as well as different departure times and highway usage rates to be input into the model.



Fig.6 Trip origin distribution heatmap



Fig.7 Trip destination distribution heatmap

A total of 96 ODs and 9,889 trips were left as the dataset for subsequent clustering.

The road network data in the Osaka-Kobe region was also collected. Each road record includes attributes such as responsible organisation, road type, road length, toll or not, number of lanes, road width, and restriction information. Fig.8 is our road network map. Using road network data, all GPS trajectories are matched to road sections.



Fig.8 Road network in Osaka-Kobe region.

Table 3 shows roads of different road types and the times they are used. Although highways account for only 2.75% of all roads, the number of times they are used accounted for 55.42% of the total number of times all roads were used.

4. METHODOLOGY

We use the binomial logit model to explore the impact of various vehicle, trip, and road characteristics on oversized truck's route choices, i.e., whether to choose the most used route or an alternative route. After the ODs of the trips are aggregated to 500 m meshes, each OD pair contains more than 50 trips. The trips usually do not overlap perfectly but have large or small spatial differences between each other. This also means that if we directly use these trips as the input of the binomial logit model, the dependent variable of the model, the option, will have too many values for us to obtain valid parameter results. Therefore, trips need to be clustered according to their spatial distance between each other to form several clusters. The route options will be reduced from dozens or hundreds to a reasonable number through clustering. Then by identifying whether a cluster is the most used cluster, the dependent variable becomes a binary variable, with 1 representing the most used cluster and 0 representing the other clusters, which are some special routes.

For clustering, a method is needed to calculate the distance of any two trajectories. In this study, the spatial distance of two routes with the same number of nodes is defined as the average value of the pairwise distances of the corresponding nodes from their start to the end:

$$D_{AB} = \frac{\sum_N D_{A_i B_i}}{N} \quad (1)$$

Table 3 Road types, number of roads and road use.

Road Type	Number of Road	Percentage	Road Use (times)			Percentage
			Large Vehicle	Extra-large Vehicle	Total	
National highway	29441	2.32%	143431007	17531565	160962572	46.98%
Urban highway	5474	0.43%	24150202	4770730	28920932	8.44%
National road	187497	14.79%	74109988	13227215	87337203	25.49%
Main local road (prefecture)	161465	12.73%	29081438	5623183	34704621	10.13%
Main local road (city)	6760	0.53%	3429505	822804	4252309	1.24%
Prefectural road	159474	12.58%	5645199	977185	6622384	1.93%
Designated city road	12378	0.98%	3161068	636694	3797762	1.11%
Others	705427	55.64%	12481183	3519336	16000519	4.67%
Total	1267916	100.00%	295489590	47108712	342598302	100.00%

In Equation 1, D_{AB} is the spatial distance between route A and route B; N is the number of nodes; $D_{A_iB_i}$ is the pairwise distance of the i -th node of route A and route B.

Fig.9 and Equation 2 give an example of the spatial distance between two routes, both with five nodes (origin node and destination node included).

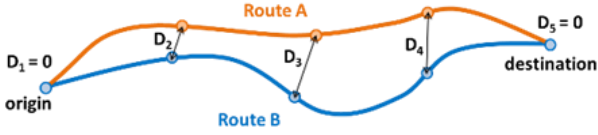


Fig.9 Spatial distance between two routes.

$$D_{AB} = \frac{(D_1 + D_2 + D_3 + D_4 + D_5)}{5} \quad (2)$$

The GPS trajectories of all trips are uniformly resampled, so that each trip has the same number of nodes, which is 256 in this study. Then using the above method, we can get the pairwise distances between all trips.

We tried some common clustering algorithms, including k-means and DBSCAN, as well as a method for neuroimaging clustering. We found that QuickBundles (QB) can obtain better clustering results and has better efficiency¹²⁾. The process of the QuickBundles algorithm is as follows:

- select the first route r_1 and place it in the first cluster c_1 . For each remaining route in turn $i = 2, \dots, N$,
- a) calculate the minimum average direct-flip (MDF) distance, i.e., spatial distance between route r_1 and the centroid routes v_e of all the current clusters c_e ;
- b) if any of the MDF values m_e are smaller than the threshold β , add route i to the cluster e with the minimum value for m_e ;
- otherwise create a new cluster.

The evaluation of clustering results is based on the Silhouette coefficient, which is to evaluate the goodness of a clustering technique. It can be understood as an indicator describing the outline clarity of each category after clustering. The Silhouette coefficient s of clustering can be calculated using Equation 3.

$$s = \frac{b - a}{\max(a, b)} \quad (3)$$

where a is the mean distance between a trip and all other trips in the same cluster; b is the mean distance between a trip and all other trips in the next nearest cluster. The value of Silhouette coefficient ranges from -1 to 1. The larger the value, the better the clustering result.

The threshold is a parameter to be selected in QuickBundles algorithm. If more clusters are needed, the threshold should be set lower. Otherwise, if fewer clusters are needed, it should be set higher. By adjusting the threshold, multiple clustering results with different numbers of clusters can be obtained. Then the Silhouette coefficient is used to select the best results and the number of clusters. Select the result with the largest Silhouette coefficient, or the previous one when the Silhouette coefficient no longer increases significantly with the number of clusters, as the best clustering result. Fig.10 and Fig.11 show the variation of the Silhouette coefficient with the number of clusters and the clustering results of an OD pair containing 125 trips.

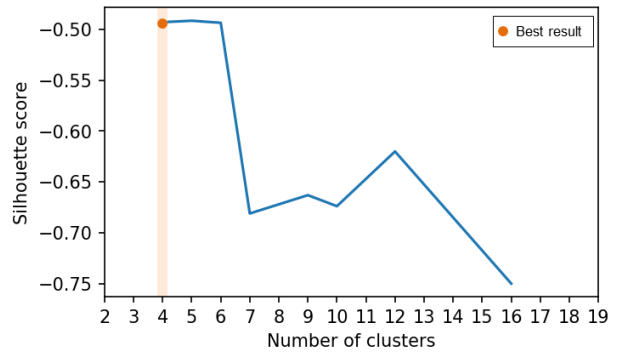


Fig.10 Select the best results and the number of clusters based on the Silhouette Coefficient.

After obtaining the various vehicle, trip, and road characteristics of all trips and their clusters, discrete choice models are used to analyze the influencing factors for route choice. In our initial analysis we employ a binomial choice model to focus on understanding why the most common route cluster is not chosen by some trucks. Clearly in further work, this can be advanced to multinomial choice models to understand in more detail the differences in route choice.

In the binomial logit model, the choice probability

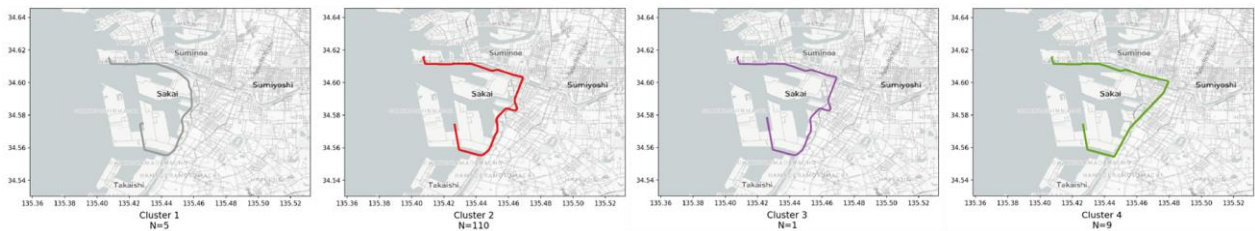


Fig.11 Clustering result of an OD pair.

Table 4 Binomial logit model result for whether to choose the most used cluster.

Factor	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.07596	0.2055	-0.37	0.7116	
left turn	-0.215	0.02065	-10.414	< 2e-16	***
right turn	-0.08739	0.01808	-4.832	1.35e-06	***
u turn	-0.1658	0.06519	-2.544	0.011	*
merging area	-0.9442	0.1066	-8.86	< 2e-16	***
diverging area	0.9078	0.1213	7.487	7.06e-14	***
main-main intersection	-0.02218	0.004319	-5.135	2.82e-07	***
main-side intersection	-0.04998	0.005594	-8.935	< 2e-16	***
side-side intersection	-0.1541	0.006794	-22.677	< 2e-16	***
average speed	0.1605	0.0205	7.829	4.94e-15	***
average lane number	1.339	0.1114	12.012	< 2e-16	***
central belt percentage	-0.7377	0.2041	-3.615	0.0003	***
highway percentage	-2.267	0.1871	-12.116	< 2e-16	***

for alternative i for person n is shown in Equations 4 and 5.

$$P_n(i) = \frac{e^{V_{in}}}{e^{V_{in}} + e^{V_{jn}}} \quad (4)$$

$$\begin{aligned} V_{in} &= \beta_0 + \beta_1 X_{in1} + \beta_2 X_{in2} + \dots + \beta_K X_{inK} \\ &= \beta' \mathbf{X}_{in} \end{aligned} \quad (5)$$

In Equations 4 and 5, V_{in} is the observed utility that truck n obtains from alternative i . Each observed utility V_{in} is a function of a vector of independent observable variables \mathbf{X}_{in} , which include vehicle, trip, and road characteristics in this research. $\beta_0, \beta_1, \beta_2, \dots, \beta_K$ are the parameters of the binomial logit model. β' is the vector of these parameters.

5. RESULTS AND DISCUSSION

The following 14 alternative influence factors were entered into the model, including the number of left turn, right turn, u turn, merging area, diverging area, main-main intersection, main-side intersection, and side-side intersection. trip length, average speed, average lane number, central belt percentage, highway percentage, and vehicle type are also included.

After stepwise screening, factors with multicollinearity and those that were not significant were removed.

The result of the binomial logit model is shown in Table 4, 12 influence factors kept. The results show that in trip characteristics, the number of left turn, right turn, u turn, merging area, diverging area, main-main intersection, main-side intersection, side-side intersection, and average speed have significant effects on route choice of oversized trucks. Among road characteristics, average lane number, average

road width, central belt percentage and highway percentage have significant effects. The effect of vehicle type on route choice was not significant.

Regarding trip characteristics, except for the diverging area, other turns or intersections have a negative effect on the choice of the most used cluster by the truck. Instead, trucks prefer alternative routes with fewer turns or intersections. Similarly, trucks will prefer an alternative route if the average speed of the alternative route is higher.

Regarding road characteristics, trucks seem to prefer roads with a high number of lanes and few central belts, as well as a tendency to choose normal roads.

6. CONCLUSIONS

In this study, GPS trajectory data of oversized trucks in the Osaka-Kobe region were aggregated by their origins and destinations at a spatial level of 500m x500m meshes. ODs with significant differences in trip and used road characteristics were selected. The binomial logit model was used to analyze the factors that influence oversized trucks' choice of the most used route or an alternative route.

The results show that oversized trucks do not tend to choose routes with many turns and intersections, as well as roads with central belts and highways.

The findings of this study will help to understand the unique preferences of oversized trucks and may also provide insights on freight road network optimization and design.

In the current clustering, only spatial distances are considered, and features about roads are not considered. For the next step of the research, not only the spatial distance of the trips but also the trip time, length, speed and road type will be used in the clustering process to classify the trips within the OD into

different types more accurately.

ACKNOWLEDGMENT: This work was supported by the Ministry of Land, Infrastructure, Transport and Tourism under the project 'Research and development of technologies that contribute to improving the quality of road policy'.

REFERENCES

- 1) The Ministry of Land, Infrastructure, Transport and Tourism : Yearbook of Automobile Transport Statistics, Statistics of Japan, <https://www.e-stat.go.jp/dbview?sid=0003443758>, 2021.
- 2) Jin, Y., Jia, Z., Wang, P., Sun, Z., Wen, K. and Wang, J. : Quantitative assessment on truck-related road risk for the safety control via truck flow estimation of various types, *IEEE Access*, 7, pp.88799-88810, 2019.
- 3) Peeta, S., Zhang, P. and Zhou, W. : Behavior-based analysis of freeway car-truck interactions and related mitigation strategies, *Transportation Research Part B: Methodological*, 39(5), pp.417-451, 2005.
- 4) East Nippon Expressway Company Limited : Vehicle classification on expressways, <https://www.drive-plaza.com/search/division>, 2022.
- 5) 兵藤哲朗, シジニシユライナー and 高橋洋二 : 東京都市圏物資流動調査を用いた大型貨物車走行経路のモデル分析, *土木計画学研究・論文集*, 24, pp.405-412, 2007.
- 6) 岡英紀, 力石真, 田名部淳, 福田大輔 and 大口敬 : 車種及び距離帯を考慮した貨物車経路選択行動のモデル分析, *土木学会論文集 D3 (土木計画学)*, 74(5), pp.I_657-I_666, 2018.
- 7) 関谷浩孝, 上坂克巳, 松本俊輔 and 古川誠 : 国際海上コンテナ車の走行経路の特性, *土木技術資料*, 52(11), pp.16-19, 2010.
- 8) 三輪富生 and 森川高行 : プローブカーデータを利用した経路選択行動に関するモデル分析, *土木計画学研究・論文集*, 21, pp.553-560, 2004.
- 9) 小澤俊博, 岸邦宏, 野見山尚志, 中西哲也, 五十嵐達哉 and 堀口拓未 : 貨物車プローブデータを用いた高速道路選択モデルの構築と高速道路利用促進施策の効果予測, *土木学会論文集 D3 (土木計画学)*, 73(5), pp.I_651-I_660, 2017.
- 10) 関谷浩孝, 萩野保克, 剣持健, 前田雅人 and 田名部淳 : 道路構造に着目した一般道における大型貨物車の経路選択モデル, *土木学会論文集 D3 (土木計画学)*, 73(5), pp.I_527-I_536, 2017.
- 11) 萩野保克, 兵藤哲朗 and 宮原ゆい : 特車申請電子データ及び道路情報便覧データを用いた海上コンテナ車の経路選択特性, *土木学会論文集 D3 (土木計画学)*, 67(5), pp.67_I_599-67_I_609, 2011.
- 12) Garyfallidis, E., Brett, M., Correia, M.M., Williams, G.B. and Nimmo-Smith, I. : Quickbundles, a method for tractography simplification, *Frontiers in neuroscience*, 6, p.175, 2012.

(Received September 30, 2022)