ESTIMATING OVERSIZED TRUCK ROUTE CHOICE BEHAVIOR USING GPS TRAJECTORY DATA AND RECURSIVE LOGIT MODELS

Yue MA[1], Jan-Dirk SCHMÖCKER[2], Wenzhe SUN[3],
Satoshi NAKAO[4], Tadashi YAMADA[5]

[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

Oversized trucks are subject to constraints imposed by their size and performance, leading to specific requirements and preferences in their route choices. This study aims to explore the impact of road attributes, such as road type, travel distance, tolls, as well as turns on route choices. GPS trajectory data from oversized trucks in the harbour region of Osaka, Japan, have been collected for this. Trips originating from and destined to the container terminals, truck terminals, and bonded areas in the region are selected. After filtering in the roads used in these trips and applying different spatial buffers, five road network scales are obtained. Recursive Logit (RL) models are employed to capture the realistic decision-making process in route choice behavior, implicitly assuming that drivers make step-by-step route decisions during the trip rather than determining the entire route from the outset. Different preferences for trips with different origins and destinations are observed, explained by their inclinations or aversions towards toll roads, road types, and lane numbers. In particular we find differences in route preferences for trucks originating from terminals and trucks destined to these. The findings offer new perspectives for designing, improving, and managing logistics road systems to meet the diverse requirements of oversized trucks based on their cargo characteristics and travel objectives. Further, we contribute to the choice modelling literature by revealing the sensitivity of the RL model to the road network scale included in the model. We show the effect on parameter estimates when excessive irrelevant alternative roads are included.

Key Words : *oversized truck; route choice; recursive logit model; GPS data; irrelevant alternative effect*

1. INTRODUCTION

The utilization of oversized trucks for cargo transportation has been on the rise. According to statistics published by the Ministry of Land, Infrastructure, Transport, and Tourism of Japan for the year 2020, out of the total 3,786,998 thousand tons of cargo transported by automobiles, 45.9% of it was carried by trucks with a load capacity of 6.5 tons or more (1).

Oversized trucks display notable characteristics distinct from small vehicles or ordinary trucks, including their larger size, weight, braking distance, blind spots, maneuvering challenges, noise, and higher fuel consumption and emissions (2). Furthermore, oversized trucks have a higher propensity to contribute to severe traffic congestion and raise safety concerns (3). Therefore, oversized trucks may have unique preferences and specific requirements

for selecting their routes, such as considering road types and turning radius at intersections. They may prefer larger intersections for turning maneuvers and are constrained by size and weight restrictions imposed by traffic management. Given these special characteristics, dedicated studies on the route choices of oversized trucks become essential and valuable.

When examining route choices with discrete choice approaches such as Multinomial Logit or other Generalized Extreme Value models, firstly the set of considered alternatives needs to be defined. Conventional approaches for choice set creation include using the shortest path according to multiple weights (4, 5), link penalty approaches (6), or random algorithms (7). However, these approaches may result in the inclusion of some routes that are highly unreasonable and irrelevant, leading to potential impacts on model results and computational burden (8). While there are a number of dicrete choice approaches that can address concern of route overlap, they still have limitations in the number of alternatives they can handle (9–11). In practical route choices, there can be a much larger number of alternative paths.

Additionally, in static discrete choice models assume that decision makers can only perceive characteristics related to the trip and determine the entire path from the beginning, which does not align with the reality where drivers can change their route during the trip. In this study, we employ the Recursive Logit (RL) model to simulate the route choice behavior of oversized trucks (12). This approach eliminates the need to pre-generate a choice set of routes and theoretically allows for an infinite number of potential routes. It does, however, still require the analyst to define the network limits and which links are included. In particular if some links are “unreasonable”, because in our case the vehicles are too large, this can lead to impacts on the estimated paramters. We discuss this sensitivity to the network scale which we suggest has not been sufficiently studied in the RL literature.

We investigate how the characteristics of road links influence the route choice of oversized trucks. Trips originating from or destined to different logistic facilities are analzed separately to determine the relationship between the OD type and the heterogeneity in route choices. Additionally, we included link pair features and turnings as potential influencing factors. To manage the computational burden and avoid interference from excessive irrelevant link alternatives to the model result, we discuss the impact of reducing the size of the road network.

The structure of this paper is as follows. In the next section, a literature review on oversized truck route choice is presented, highlighting the contribution of this research. Section 3 introduces the Recursive Logit model employed in this study. Section 4 outlines the data preparation process for our modeling. The results and interpretations are discussed in Section 5. This paper concludes with a summary of findings and recommendations in the final section.

2. LITERATURE REVIEW

The route choice problem of trucks has been widely studied, given the increasing significance of efficient and sustainable freight transportation by trucks and the attention to safety, emissions, and operational effectiveness.

Studying the characteristics and influencing factors of truck route choice serves two main purposes. Firstly, it enhances the understanding and prediction of trip paths with more accuracy, especially when only the trip's origin and destination are available. Traditional route estimation for oversized trucks has relied on the shortest path or required time, overlooking the significant impact of road factors (13). Additionally, trip characteristics such as travel distance significantly impact truck route choices (14). Origin-destination (OD) specific attributes also play a crucial role in route usage decisions, like container trucks favoring large intersections for turns and time-sensitive trucks preferring highways (15). 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 routes (16). Incorporating these factors into the model yields more reasonable explanations and predictions. The second aim is to provide policy recommendations for road network optimization and truck management. Frequent use of roads shared by trucks with other traffic can lead to heavy congestion and accidents. Similar to our research, but not focusing on oversized trucks, Japanese scholars have employed discrete choice models to extract route preferences and then propose strategies for encouraging truck drivers to choose expressways by implementing road transformations on bottleneck sections and optimizing expressway fees (17).

Recursive Logit (RL) models have been introduced as an alternative to standard discrete choice approaches to estimate route choice as a weighted Markovian process where the weights are to be estimated. The advantage of the Recursive Logit (RL) model over traditional discrete choice models lies in its ability to avoid generating a finite set of path or trip choice and its capability to handle step-by-step decision-making processes, which better reflects reality, as various factors such as traffic conditions and trip objectives may lead people to make different route choices during their journey (18). As a result of its flexibility and realism, the RL model is increasingly

used in research on travel patterns, crowd monitoring, and route choice. Through the use of Wi-Fi probes to collect the locations of travelers or crowds and then generate travel chains or trajectories, the RL model is utilized to identify travel patterns (19) and large-scale crowd activity patterns (20). Route choice behavior has also been studied for various modes of transportation, including e-scooters (18), vehicles (21), taxis (22), and trucks (23, 24). Furthermore, some studies have made improvements to the RL model to accelerate estimation speed (25), further relax the IIA property (26, 27), consider utility diminishing effects (22), or deal with incomplete trip observations (28).

Regarding data sources, previous studies have utilized travel route surveys (29, 30), truck travel permit application data, and GPS trajectories (23) from vehicle probe. Among these, GPS trajectories can provide precise time and spatial location of the vehicles, including the characteristics of roads used after map matching, making it easier to introduce more factors for analysis.

While extensive research exists on truck route choice behavior, this study focuses on understanding how specific road characteristics influence truck route choices between multiple origin-destination pairs. Moreover, the model we selected, Recursive Logit (RL) model, allows us to better simulate the decision-making process of oversized trucks and incorporate more types of independent variables into the model.

3. METHODOLOGY

The concept of the Recursive Logit (RL) model is illustrated in Figure 1 (12).

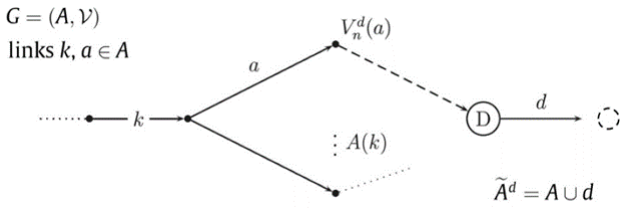


Fig.1 Concept of the Recursive Model, proposed by (12), modified by (19).

G is a directed connected graph consisting of the node set $\mathcal{V}$ and the link set A. Links $k, a \in A$, in route choice problem, where a path is a sequence of links, k is the current link and a is a potential next alternative link given current link k. The set of all alternative links from the sink node of link k is denoted as $A(k)$.

At each link k, the oversized truck n selects the next link a from the set of all alternative links $A(k)$. The selection is based on maximizing the sum of instantaneous utility $u_n(a|k)$ and expected downstream utility. The former is given by Equation 1, and the latter is defined by taking the continuation of this process into account via the Bellman equation (31), as shown in Equation 2.

$$u_n(a|k) = v_n(a|k) + \mu\varepsilon_n(a) \quad (1)$$

$$V_n^d(k) = E\left[\max_{a \in A(k)} (v_n(a|k) + V_n^d(a) + \mu\varepsilon(a))\right], \forall k \in A \quad (2)$$

where $v_n(a|k) = v(x_{n,a|k}; \beta)$ is a linear combination of observed characteristics $x_{n,a|k}$ associated with each link pair, which may include road characteristics, such as road length, or road (link) pair characteristics, such as turning angle. β is an unknown parameter vector to be estimated.

Then the probability of choosing a next link given a current link k is given by the Multinomial Logit model by Equation 3.

$$P(\sigma) = \prod_{i=0}^{I-1} e^{\frac{1}{\mu}(v(k_{i+1}|k_i) + V(k_{i+1}) - V(k_i))} \quad (3)$$

The RL model can also accommodate the Link Size (LS) attribute, which relaxes the Independence of Irrelevant Alternatives (IIA) property of the logistic model by adjusting the utility function for overlapping paths (12), similar to the C-Logit model (10) and Path Size Logit (PSL) model (11). An issue of considering Link Size is only that the computational time increases significantly for largr problems as origin-destination-specific choices instead of destination specific choices need to be estimated (12).

4. DATA

GPS trajectory data was collected in Osaka, Japan during the period of July to September 2020. We limited our study area to a central zone with dimensions of 10km x 10km for two main reasons. Firstly, this area serves as a key harbour region in Osaka, housing numerous logistics facilities for oversized trucks, such as container terminals, truck terminals, and bonded areas (see Figure 2). As a result, a considerable number of trips by oversized trucks can be collected, enabling us to investigate the diverse route choices for trips originating from and destined to different logistics facilities. Secondly, since we consider a detailed road network, choosing larger areas leads to computationally challenging problems, so the chosen study area is manageable and practical.

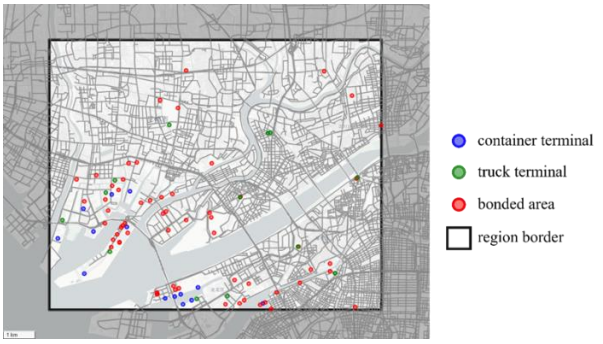


Fig.2 Osaka study area (10km x 10km) and distributin of three types of logistic facilities.

The collected trips have been map matched. Since the GPS records do not contain a marker for the truck having reached its destination and to ensure that consecutive trips are not counted as a single trip, the trajectores are split based on the vehicle speed. For instance, if a truck departs from location A to location B and then returns to location A, it should be treated as two separate trips instead of one continuous trip. If the vehicle speed remains below 5 km per hour continuously for more than 120 s, we consider it as a segmentation point to potentially split the trip. We also examine whether vehicles in low-speed conditions experience directional reversals to avoid misidentifying low-speed states caused by congestion or traffic signals as reaching its destination. Only if both conditions are fulfilled we define the segementation point as a destination. Additionally, trips with a total travel distance less than 5 kilometers or consisting of only 5 or fewer links are discarded to avoid insignificant short trips (These might be maneuvers connected to loading and unloading inside terminals). Finally, 2285 trips are retained, involving 841 oversized trucks, as shown in Figure 3. The numbers of trips and used links are presented in Table 1. Trips originating from or destined to container terminals (CTs), truck terminals (TTs), and bonded areas (BAs) are reported separately. In the RL model, all links should be unidirectional, so we add reversed dummy links for bi-directional roads.



Fig.3 2285 trips generated by 841 oversized trucks (different colors to differentiate between trips).

Table 1 Number of trips, links, and link pairs for trips originating from or destined to three types of logistic facilities.

Number	from CT	to CT	from TT	to TT	from BA	to BA
trips	518	158	210	280	980	882
links	1080	910	1168	1616	2171	2488
links inc. dummy links	1509	1224	1582	2110	3015	3451
link pairs	3253	2640	3378	4600	6673	7764

For each road link, the following information is available to describe its characteristics:

- Administrator, such as whether the road is a national, prefectural or the municipality administered.
- Whether or not the road is a freeway and whether or not it is tolled. We note that the majority of freeways in Japan are tolled as well as a few other roads.
- Whether it is a recommended arterial road or a local road, where we note that “recommended” is not synonymous with “main arterial road”, but rather a concept relative to “local road”. Local roads, which are not preferred or suitable for vehicles, constitute only about 10% of the total road links.
- Further information on road type is available in the form as to whether the road is a bi-directional divided road, bi-directional undivided road, or parallel side road, as exemplified in Figure 4.
- Other discrete attributes are number of lanes and whether or not the road is for motorized traffic only.
- “Continuous variable information” specifiygin the road are length, road width and median barrier width and speed limit
- Traffic volume related information are average traffic volume (within a 12-hour period, based on on-site traffic surveys) and peak hour travel speed.
- Finally, we have information as to whether the road is included in the truck management system. If included in the system, this means that even oversized trucks do not need to obtain permission for using these routes and that emergency management measures are in place.



Fig.4 Examples for road types and toll road.

Clearly, some road attributes are found to have strong correlations, which can affect the accuracy of parameter estimation of the RL model. Therefore, pairs of factors with an absolute Pearson correlation coefficient greater than 0.4 are reduced to one factor, unless both factors are considered essential, in which case the threshold is relaxed to 0.5 (Figure 5). Note that the categorical variable "administrator" is transformed into ordinal variable by 1: national, 2: prefectural, 3: municipality administered, and 4: others, such as private roads.

After adding dummy variables for categorical factors, we have a total of 10 factors to be input into the RL model: whether it is a recommended arterial road, road length, whether the number of lanes is greater than or equal to 4 (total in both directions), median barrier width, whether it is a toll road, whether it is a bi-directional undivided road, whether it is a parallel side road, and whether it is a left turn, right turn, or u-turn. These factors were selected based on their relevance to the route choice behavior of oversized trucks and their potential influence on the utility of road links.

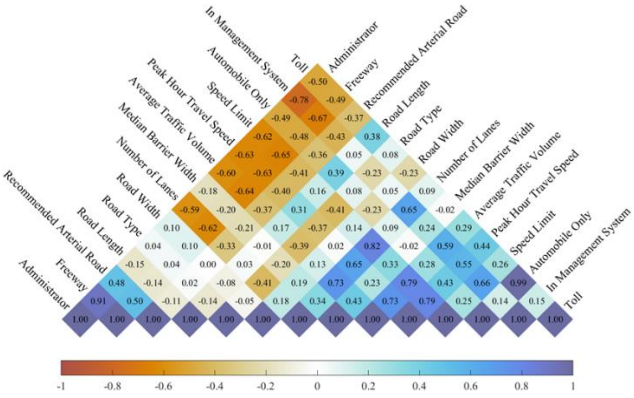


Fig.5 Results of Pearson correlation coefficient analysis between factors.

One of the advantages of using the RL model is that, since the utility is defined for link pairs rather than individual links, we can incorporate variables such as turning types. Figure 6 illustrates how we identified turning types for each link pair. A directional change between the current link and the next link greater than 45 degrees but less than 135 degrees is identified as a left or right turn, while a directional change greater than 135 degrees is identified as a u-turn.

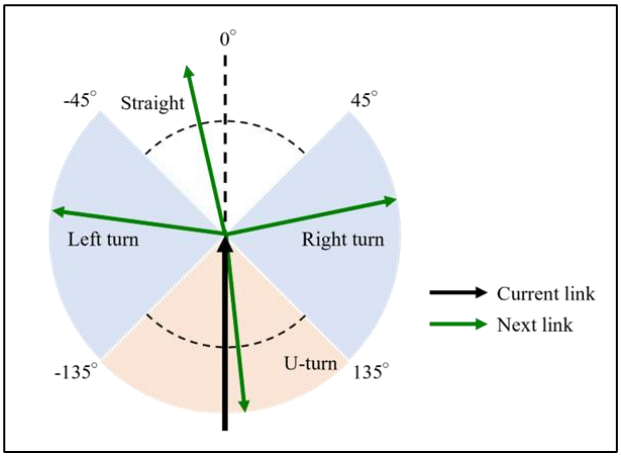


Fig.6 Identification of left turn, right turn, and u-turn.

The final step is to prepare the road network for the RL model. In order to expedite parameter estimation while ensuring that only roads relevant to oversized trucks are included, we first select all roads used in trips within the three-month period and create a "0 buffer" road network. This network includes only links that have been used at least once. It is hence a complete and efficient network with the definition of (32) in that all taken options are included and no never chosen option is included. The term "complete" can, however, be argued, especially for network planning purposes. The model with this choice set explains why some links are more chosen than others but does not explain why some links are never chosen. This is, however, relevant for network planning where one like to include the "consideration set" not just the "chosen set".

Therefore, to get closer to a potential "complete consideration choise set" we expand the network. Based on "0 buffer" road network, we apply different spatial buffers and include roads within the buffer zones into the choice set. We include more similar route options in the choice set. As the buffer size increases, more unused road links are included.

It is important to note that as the road network scale increases, the estimation time for the RL model also increases. Figure 7 illustrates the road networks for 0 buffer, 100m buffer, 300m buffer, 500m buffer, and 1000m buffers. When the buffer size is 1000m, the choice set road network is close to the entire road network. Road characteristics encompassed by different buffer scales, and the non-linear growth in network size for each buffer scale is presented in Table 2.

Table 2 Descriptive Analysis for 10 Factors in 5 Different Road Network Buffer Scales.

Factor	Statistics	0 buffer	100m buffer	300m buffer	500m buffer	1000m buffer
Recommended Arterial Road	%	72.90	73.47	84.06	86.76	87.36
Road Length	mean	154.04	112.88	111.88	115.30	115.42
	std	184.30	167.84	136.28	135.18	135.13
	min	4	2	2	2	2
	max	3815	3841	3841	3841	3841
Number of Lanes ≥ 4	%	40.36	26.87	16.74	15.08	15.07
Median Barrier Width	mean	11.84	11.28	6.64	6.24	5.75
	std	36.17	40.67	30.70	29.02	26.86
	min	0	0	0	0	0
	max	300	300	300	300	430
Toll	%	4.17	4.66	3.15	2.97	2.55
Bi-Directional Undivided Road	%	51.36	48.67	68.30	71.74	73.04
Parallel Side Road	%	0.27	1.77	0.93	0.73	0.66
Left Turn	%	12.04	20.42	28.11	28.66	28.75
Right Turn	%	11.75	30.53	29.50	29.90	29.74
U Turn	%	1.61	2.84	2.28	2.11	2.06
# of trips				518		
# of links		1080	2613	4864	6845	9714
# of links inc. dummy links		1509	3844	7931	11389	16286
# of link pairs		3253	9690	22512	33589	49040

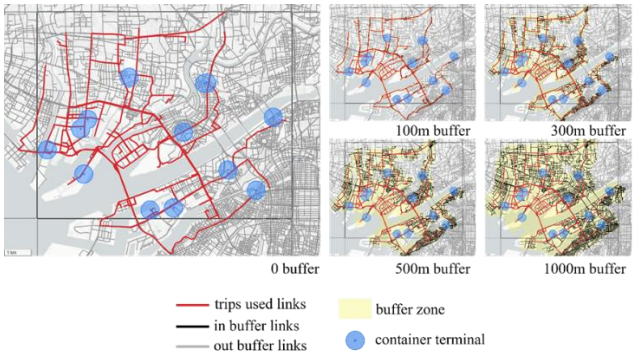


Fig.7 Buffer scales for trips from container terminals.

5. RESULTS AND DISCUSSION

Using the method mentioned in the preceding section, we obtained five different road networks of varying scales, with buffer ranging from 0-meter to 1000-meter. We employed and adapted the MATLAB code from (12) to perform parameter estimation. The estimation results for trips from CT are presented in Table 3.

Among all ten estimated parameters, only the median barrier width factor does not significantly affect link utility, regardless of the road network scale used. The results indicate that oversized trucks show a preference for recommended arterial roads rather than those that are not recommended, such as local roads. Moreover, the road length and 4+ lanes, have a negative impact on utility. It is important to note that in the RL model, link attributes are considered additive. Therefore, it is not indicative of a preference for shorter individual road link (which means a more fragmented road with frequent intersections), but rather a preference for roads that can result in shorter overall travel distance.

We have three types of roads: bi-directional divided (BDD) road, bi-directional undivided (BDU) road, and parallel side (PS) road. Oversized trucks prefer the higher-grade BDD road while showing aversion towards the latter two. Notably, the PS road has a more pronounced negative impact on utility than BDU road, with coefficients of -0.962 and -0.171, respectively, in the 0-buffer case. This can be explained by the fact that PS roads are often parallel to urban expressways or highways (both belonging to the BDD road type). As a result, oversized trucks usually use the PS road only to access the BDD road.

As expected, turning has a significant and strong negative effect on utility, with u-turn showing the most evident impact. In the 0-buffer case, the parameter for u-turn is -3.643, followed by -1.559 for right turn, and -0.802 for left turn. In Japan, where vehicles drive on the left side of the road, right turns require traveling a greater distance at intersections compared to left turns. Moreover, it is worth observing that as the buffer scale increases, the absolute values of the parameters, representing the magnitude of the turn

Table 3 Recursive Logit Model Estimation Using Different Buffer Scales for Trips from CT.

Factor	Note	0 buffer	100m buffer	300m buffer	500m buffer	1000m buffer
Recommended Arterial Road	vs. local road	0.166***	0.159***	0.124***	0.107***	0.079***
Road Length	per 100m	-0.143***	-0.105***	-0.117***	-0.116***	-0.117***
Number of Lanes ≥ 4	vs. <4	-0.227***	-0.235***	-0.247***	-0.222***	-0.180***
Median Barrier Width	per 10m	-0.083	-0.016	0.064	0.047	0.018
Toll	vs. not toll	0.283***	-0.154*	-0.219*	-0.168*	-0.158*
Bi-Directional Undivided Road	base: bi-directional divided road	-0.171***	-0.224***	-0.233***	-0.251***	-0.205***
Parallel Side Road	base: bi-directional divided road	-0.962***	-2.585***	-2.520***	-2.547***	-2.521***
Left Turn	base: straight	-0.802***	-1.561***	-1.667***	-1.681***	-1.694***
Right Turn	base: straight	-1.559***	-2.057***	-2.231***	-2.230***	-2.257***
U Turn	base: straight	-3.643***	-4.095***	-3.645***	-3.662***	-3.425***
initial LL		-13.621	-13.621	-13.621	-13.621	-13.621
max LL		-0.682	-1.199	-1.559	-1.611	-1.684
# of trips		518	518	518	518	518
# of links		1080	2613	4864	6845	9714
# of links inc. dummy links		1509	3844	7931	11389	16286
# of link pairs		3253	9690	22512	33589	49040
Estimation time (s)		437.67	1868.20	5024.90	8386.50	13281.00

*** for a significance level of 0.001, ** for 0.01, and * for 0.05

factors, show noticeable changes.

The significance of the toll parameter also varies with the buffer scale. At 0-buffer, the significance level is 99.9%, but it reduces to 95% for buffer scales of 100m, 300m, 500m, and 1000m. This reflects the RL model's sensitivity to the utilization of different road networks. A proper interpretation requires combining the information of the road characteristics encompassed by different buffer scales, as detailed in Table 2. The proportion of toll roads varies with different buffer scales. Toll roads, which primarily consist of highways and expressways, tend to be spatially concentrated and interconnected. This leads to an increase in the proportion of toll roads when a 100m buffer is applied. Conversely, non-toll roads exhibit a more homogeneous spatial distribution. Therefore, as the buffer is further increased, more non-toll roads are encompassed, while the increase in toll road number is relatively minimal, leading to a decrease in its overall proportion.

We suggest that the inclusion of excessive non-toll roads could lead to what is referred to as the “irrelevant alternative effect”, which may be misleading to the model estimation results. Many of these non-toll roads have never been used by oversized truck trips within the three-month period, meaning they are not realistic alternatives that oversized trucks would genuinely consider. In the RL model, a road link's utility includes downstream effects. Irrelevant alternatives can influence this downstream utility of viable alternatives, consequently impacting the estimation of parameters. The similar irrelevant alternative effect is also evident in some other parameters, although it may not manifest in terms of significance or sign, but rather in the magnitudes they represent. As the buffer scale increases from 0 to 100m, the influence of recommended arterial roads decreases, while the impact of BDU and PS road types increases, as well as the effects of the three types of turns. As the buffer continues to increase, there are still variations, but these become less pronounced. From Table 2, it can be observed that the road proportions for these factors vary significantly from 0 buffer to 100m buffer. Beyond 100m buffer, the changes become minimal.

In summary, the RL model exhibits sensitivity to the road network used. The inclusion of excessive irrelevant alternatives can lead to changes or even reversals in the estimated parameters, as well as variations in the significance levels. When addressing the issue of oversized truck's route choice, it is crucial to use a road network that encompasses viable roads and exclude irrelevant alternatives. We conducted RL model parameter estimation for six different trip types, each with varying origins or destinations, using 0 buffer road networks. These trips include departures/arrivals at container terminals (CT), departures/arrivals at truck terminals (TT), and departures/arrivals at bonded areas (BA). The results are

presented in Table 4.

To address potential overlapping path biases, we included link size (LS) as an additional attribute in a version of the model. As demonstrated in (12), the chosen beta for the model with LS is the optimal beta obtained without considering LS. The estimation results are presented in Table 5.

The coefficients for the LS attribute are all negative, which aligns with our expectations, as overlapping paths can reduce utility. However, in the model with "from TT" trips, the LS attribute is not significant. This suggests that the overlapping paths for these trips are not prominent, or the overlap is evenly distributed in the road network (19).

Moreover, the introduction of the LS attribute does not alter the sign of any significant factors' coefficients, it only affects some significance levels. While the p-values indicate that most of such changing significance level factors are already close to being significant without LS. For instance, in the model without LS, the toll factor's p-value for "to CT" trips is 0.052, the recommended arterial road factor's p-value for "to BA" trips is 0.067. The toll factor's p-value for "to TT" trips is 0.058, and the recommended arterial road factor's p-value is 0.211. In the model considering LS, they become significant. The magnitudes of the LS factors indicate that "to TT" trips indeed experience the most frequent overlapping paths.

The oversized trucks returning to the three types of logistic facilities exhibit a disfavor for toll roads and recommended arterial roads, with a particular aversion observed among those originating from the CT. Trucks originating from the TT and BA are relatively unaffected by these two factors. Only the trucks heading to the CT show a preference for roads with 4+ lanes. Trucks except for those heading to the BA, do not consider the median barrier width as a significant factor in their route choices.

The longer travel distance and lower-grade BDU and PS roads are never favored, and the three types of turns also have a negative impact on the utility. However, it is evident that the trucks departing from CT, TT, and BA strongly dislike u-turns compared to the returning trucks.

6. CONCLUSIONS

In this research, we utilized GPS trajectory data from oversized trucks operating in the Osaka harbour region, Japan, to study route choice criteria. The selected trips are originated from or destined to container terminals, truck terminals, and bonded areas.

Table 4 Recursive Logit Model Estimation for Trips (without link size attribute).

Factor	Note	from CT	to CT	from TT	to TT	from BA	to BA
Recommended Arterial Road	vs. local road	0.166***	-0.106***	-0.039	-0.022	-0.020	-0.021
Road Length	per 100m	-0.143***	-0.087***	-0.137***	-0.119**	-0.128***	-0.088***
Number of Lanes ≥ 4	vs. <4	-0.227***	0.103***	-0.138***	-0.136***	-0.114***	-0.123***
Median Barrier Width	per 10m	-0.083	-0.096	0.084	-0.056	-0.020	-0.034*
Toll	vs. not toll	0.283***	-0.467	0.037	-0.383	0.034	-0.183*
Bi-Directional Undivided Road	base: bi-directional divided road	-0.171***	-0.029	-0.218***	-0.039	-0.062*	-0.125***
Parallel Side Road	base: bi-directional divided road	-0.962***	-0.662**	-0.581***	-1.203***	-1.048***	-0.901***
Left Turn	base: straight	-0.802***	-0.760***	-0.667***	-0.857***	-0.951***	-1.125***
Right Turn	base: straight	-1.559***	-1.486***	-1.471***	-1.468***	-1.702***	-1.659***
U Turn	base: straight	-3.643***	-0.890***	-1.414***	-0.724***	-2.164***	-1.382***
initial LL		-13.621	-37.362	-21.144	-29.456	-20.848	-27.664
max LL		-0.682	-1.620	-1.053	-1.530	-1.435	-2.237
# of trips		518	158	210	280	980	882
# of links		1080	910	1168	1616	2171	2488
# of links inc. dummy links		1509	1224	1582	2110	3015	3451
# of link pairs		3253	2640	3378	4600	6673	7764
Estimation time (s)		437.67	112.77	157.76	314.24	1979.30	1871.70

*** for a significance level of 0.001, ** for 0.01, and * for 0.05

Table 5 Recursive Logit Model Estimation for Trips (with link size attribute).

Factor	Note	from CT	to CT	from TT	to TT	from BA	to BA
Recommended Arterial Road	vs. local road	0.196***	-0.065**	-0.036	-0.037*	0.002	-0.015*
Road Length	per 100m	-0.162***	-0.117***	-0.146***	-0.156**	-0.161***	-0.106***
Number of Lanes ≥ 4	vs. <4	-0.244***	0.124***	-0.148***	-0.121***	-0.127***	-0.124***
Median Barrier Width	per 10m	-0.078	-0.121	0.101	-0.086	-0.020	-0.047***
Toll	vs. not toll	0.274***	-0.563**	0.015	-0.550***	0.041	-0.287***
Bi-Directional Undivided Road	base: bi-directional divided road	-0.191***	-0.051	-0.233***	-0.078	-0.074***	-0.144***
Parallel Side Road	base: bi-directional divided road	-1.041***	-0.732**	-0.608***	-1.322***	-1.164***	-0.984***
Left Turn	base: straight	-0.844***	-0.778***	-0.695***	-0.835***	-1.037***	-1.236***
Right Turn	base: straight	-1.654***	-1.589***	-1.539***	-1.583***	-1.962***	-1.892***
U Turn	base: straight	-3.723***	-0.886***	-1.311***	-0.489**	-2.176***	-1.382***
Link Size	/	-0.060**	-0.096***	-0.037	-0.112***	-0.096***	-0.103***
initial LL		-21.380	-60.742	-39.050	-52.182	-35.636	-45.511
max LL		-0.674	-1.587	-1.047	-1.451	-1.369	-2.103
# of trips		518	158	210	280	980	882
# of links		1080	910	1168	1616	2171	2488
# of links inc. dummy links		1509	1224	1582	2110	3015	3451
# of link pairs		3253	2640	3378	4600	6673	7764
Estimation time (s)		553.63	132.41	205.67	366.22	2475.80	2887.80

*** for a significance level of 0.001, ** for 0.01, and * for 0.05

By filtering in the roads used in these trips and applying varying spatial buffers, we derived five distinct road network scales: 0-buffer, 100m buffer, 300m buffer, 500m buffer, and 1000m buffer. The road network with a 0-buffer contains the fewest road links, and the network with a 1000m buffer includes the most. Employing Recursive Logit (RL) models, we were able to capture the realistic decision-making process of drivers in route choice behavior, where step-by-step route decisions are made during the trip rather than determining the entire route from the beginning.

The RL model is considered capable of handling with infinite alternatives, thus eliminating the need to generate a choice set of routes like traditional Multinomial Logit (MNL) model or Path Size Logit (PSL) model. However, our results show that the choice set problem does not fully vanish when employing RL models. We show that the RL model is sensitive to road network scale, particularly when excessive irrelevant alternative roads are considered. This suggests that a carefully selected choice set of viable links is still necessary, especially in cases like oversized truck route choices, where road choice behavior is constrained.

Our results confirm some of the well-known and expected route choice factors, but we suggest we can add to the literature. In line with expectations we find trucks' preference against long travel distances, low-grade road types, and turning, especially u-turns. Additionally, the results reveal diverse preferences for trips with different origins and destinations. Trucks bound for the three types of logistic facilities generally exhibit a preference for non-toll roads, but for trucks departing from these facilities, there is indifference or inclination towards them. Trucks heading to bonded areas show a preference for narrow median barriers, those destined for container terminals tend to prefer roads with more lanes, and trucks departing from container terminals also demonstrate a preference for recommended arterial roads. Especially for trucks from and to container terminals the differences are pronounced. Trucks from the terminal prefer to take tolled routes whereas on their return they avoid these, probably since there is less time pressure and the fees on Japanese toll roads can be substantial. To furher explore these results, it would clearly be beneficial to have information about the trucks and their type and amount of cargo loaded, but for commercial reasons such information in conjunction with GPS trajectories is not available to us.

Overall, we suggest that the findings provide insights for the design, enhancement, and management of logistics road systems, catering to the varied needs of oversized trucks according to their cargo attributes and travel goals.

ACKNOWLEDGMENTS: This work is 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'. For the Recursive Logit model estimation, we implement a version of code made available by Tien Mai under https://github.com/maitien86/.

REFERENCES

1) The Ministry of Land, Infrastructure, Transport and Tourism. Yearbook of Automobile Transport Statistics'. Statistics of Japan. Jun 25, 2023.
2) Jin, Y., Z. Jia, P. Wang, Z. Sun, K. Wen, and J. Wang. Quantitative Assessment on Truck-Related Road Risk for the Safety Control via Truck Flow Estimation of Various Types. IEEE Access, Vol. 7, 2019, pp. 88799–88810. https://doi.org/10.1109/ACCESS.2019.2924699.
3) Peeta, S., P. Zhang, and W. Zhou. Behavior-Based Analysis of Freeway Car–Truck Interactions and Related Mitigation Strategies. Transportation Research Part B: Methodological, Vol. 39, No. 5, 2005, pp. 417–451. https://doi.org/10.1016/j.trb.2004.06.002.
4) Hunt, D. T., and A. L. Kornhauser. Assigning Traffic over Essentially-Least-Cost Paths. Transportation Research Record, Vol. 1556, No. 1, 1996, pp. 1–7. https://doi.org/10.1177/0361198196155600101.
5) Bekhor, S., and C. G. Prato. Effects of Choice Set Composition in Route Choice Modeling: 11th Conference of the International Association for Travel Behaviour Research. Proceedings of the 11th IATBR Conference, 2006.
6) Prato, C. G., and S. Bekhor. Applying Branch-and-Bound Technique to Route Choice Set Generation. Transportation Research Record, Vol. 1985, No. 1, 2006, pp. 19–28. https://doi.org/10.1177/0361198106198500103.
7) Nielsen, O. A. A Stochastic Transit Assignment Model Considering Differences in Passengers Utility Functions. Transportation Research Part B: Methodological, Vol. 34, No. 5, 2000, pp. 377–402. https://doi.org/10.1016/S0191-2615(99)00029-6.
8) Prato, C. G. Route Choice Modeling: Past, Present and Future Research Directions. Journal of Choice Modelling, Vol. 2, No. 1, 2009, pp. 65–100. https://doi.org/10.1016/S1755-5345(13)70005-8.
9) Bliemer, M. C. J., and P. H. L. Bovy. Impact of Route Choice Set on Route Choice Probabilities. Transportation Research Record, Vol. 2076, No. 1, 2008, pp. 10–19. https://doi.org/10.3141/2076-02.
10) Cascetta, E., A. Nuzzolo, F. Russo, and A. Vitetta. A Modified Logit Route Choice Model Overcoming Path Overlapping Problems. Specification and Some Calibration Results for Interurban Networks. Transportation and Traffic Theory, 1996, 1996, pp. 697–711.
11) Ben-Akiva, M., and M. Bierlaire. Discrete Choice Methods and Their Applications to Short Term Travel Decisions. In Handbook of Transportation Science (R. W. Hall, ed.), Springer US, Boston, MA, pp. 5–33.
12) Fosgerau, M., E. Frejinger, and A. Karlstrom. A Link Based Network Route Choice Model with Unrestricted Choice Set. Transportation Research Part B: Methodological, Vol. 56, 2013, pp. 70–80. https://doi.org/10.1016/j.trb.2013.07.012.
13) Hyodo, T., S. Schreiner, and Y. Takahashi. Large Size Truck's Route Choice Modeling by Tokyo Metropolitan Freight Survey. INFRASTRUCTURE PLANNING REVIEW, Vol. 24, 2007, pp. 405–412. https://doi.org/10.2208/journalip.24.405.
14) OKA, H., M. Chikaraishi, J. TANABE, D. Fukuda, and T. Oguchi. An Analysis of Trucks' Route Choice Behavior Considering Vehicle Type and Trip Length. Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management), Vol. 74, 2018, p. I_657-I_666. https://doi.org/10.2208/jscejipm.74.I_657.
15) Hirotaka, S., M. Shunsuke, and F. Makoto. International Logistics, International Marine Container, Truck, Route Selection.
16) Miwa, T., and T. Morikawa. Analysis on Route Choice Behavior Based on Probe-Car Data. Proceedings of the 10th World Congress on Intelligent Transport Systems, 2003. https://doi.org/10.2208/journalip.21.553.
17) Ozawa, T., K. Kishi, H. Nomiyama, T. Nakanishi, T. Ikarashi, and T. Horiguchi. Predicting the Effects of Promotion of Using Expressway and Its Route Choice Model Based on The Freight Car Probe Data. Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management), Vol. 73, 2017, p. I_651-I_660. https://doi.org/10.2208/jscejipm.73.I_651.
18) Zhang, W., R. Buehler, A. Broaddus, and T. Sweeney. What Type of Infrastructures Do E-Scooter Riders Prefer? A Route Choice Model. Transportation Research Part D: Transport and Environment, Vol. 94, 2021, p. 102761. https://doi.org/10.1016/j.trd.2021.102761.
19) Gao, Y., and J.-D. Schmöcker. Estimation of Walking Patterns in a Touristic Area with Wi-Fi Packet Sensors. Transportation Research Part C: Emerging Technologies, Vol. 128, 2021, p. 103219. https://doi.org/10.1016/j.trc.2021.103219.
20) van Oijen, T. P., W. Daamen, and S. P. Hoogendoorn. Estimation of a Recursive Link-Based Logit Model and Link Flows in a Sensor Equipped Network. Transportation Research Part B: Methodological, Vol. 140, 2020, pp. 262–281. https://doi.org/10.1016/j.trb.2020.08.003.
21) Knies, A., J. Lorca, and E. Melo. A Recursive Logit Model with Choice Aversion and Its Application to Transportation Networks. Transportation Research Part B: Methodological, Vol. 155, 2022, pp. 47–71. https://doi.org/10.1016/j.trb.2021.10.011.
22) Oyama, Y., and E. Hato. A Discounted Recursive Logit Model for Dynamic Gridlock Network Analysis. Transportation Research Part C: Emerging Technologies, Vol. 85, 2017, pp. 509–527. https://doi.org/10.1016/j.trc.2017.10.001.
23) Kaneko, N., H. Oka, M. Chikaraishi, H. Becker, and D. Fukuda. Route Choice Analysis in the Tokyo Metropolitan Area Using a Link-Based Recursive Logit Model Featuring Link Awareness. Transportation Research Procedia, Vol. 34, 2018, pp. 251–258. https://doi.org/10.1016/j.trpro.2018.11.039.
24) Oka, H., D. Fukuda, and T. Shinohara. Tour Pattern Choice Modelling and Simulation of Freight Trucks in the Tokyo Metropolitan Area. Procedia Computer Science, Vol. 170, 2020, pp. 708–713. https://doi.org/10.1016/j.procs.2020.03.167.
25) Mai, T., F. Bastin, and E. Frejinger. A Decomposition

Method for Estimating Recursive Logit Based Route Choice Models. EURO Journal on Transportation and Logistics, Vol. 7, No. 3, 2018, pp. 253–275. https://doi.org/10.1007/s13676-016-0102-3.
26) Mai, T., F. Bastin, and E. Frejinger. On the Similarities between Random Regret Minimization and Mother Logit: The Case of Recursive Route Choice Models. Journal of Choice Modelling, Vol. 23, 2017, pp. 21–33. https://doi.org/10.1016/j.jocm.2017.03.002.
27) Mai, T., M. Fosgerau, and E. Frejinger. A Nested Recursive Logit Model for Route Choice Analysis. Transportation Research Part B: Methodological, Vol. 75, 2015, pp. 100–112. https://doi.org/10.1016/j.trb.2015.03.015.
28) Mai, T., T. V. Bui, Q. P. Nguyen, and T. V. Le. Estimation of Recursive Route Choice Models with Incomplete Trip Observations. Transportation Research Part B: Methodological, Vol. 173, 2023, pp. 313–331. https://doi.org/10.1016/j.trb.2023.05.004.
29) Meyer de Freitas, L., H. Becker, M. Zimmermann, and K. W. Axhausen. Modelling Intermodal Travel in Switzerland: A Recursive Logit Approach. Transportation Research Part A: Policy and Practice, Vol. 119, 2019, pp. 200–213. https://doi.org/10.1016/j.tra.2018.11.009.
30) Zimmermann, M., O. Blom Västberg, E. Frejinger, and A. Karlström. Capturing Correlation with a Mixed Recursive Logit Model for Activity-Travel Scheduling. Transportation Research Part C: Emerging Technologies, Vol. 93, 2018, pp. 273–291. https://doi.org/10.1016/j.trc.2018.05.032.
31) Bellman, R. Dynamic Programming. Science, Vol. 153, No. 3731, 1966, pp. 34–37. https://doi.org/10.1126/science.153.3731.34.
32) Shelat, S., O. Cats, N. van Oort, and H. van Lint. Calibrating Route Choice Sets for an Urban Public Transport Network Using Smart Card Data. Presented at the 2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2019.