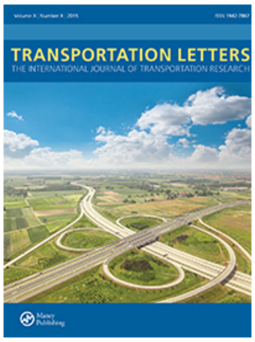

Taylor & Francis
Taylor & Francis Group

Transportation Letters

The International Journal of Transportation Research

ISSN: 1942-7867 (Print) 1942-7875 (Online) Journal homepage: https://www.tandfonline.com/loi/ytrl20

Modeling discretionary lane-changing behavior on urban streets considering drivers' heterogeneity

Hongmei Zhou, Ye Sun, Xiao Qin, Xiujuan Xu & Ronghan Yao

TRANSPORTATION LETTERS
https://doi.org/10.1080/19427867.2019.1581465

Modeling discretionary lane-changing behavior on urban streets considering drivers' heterogeneity

Hongmei Zhou[a], Ye Sun[a], Xiao Qin[b], Xiujuan Xu[c] and Ronghan Yao[a]

[a]School of Transportation and Logistics, Dalian University of Technology, Dalian, China; [b]Civil and Environmental Engineering, University of Wisconsin-Milwaukee,Milwaukee, WI, USA; [c]School of Software Technology, Dalian University of Technology, Dalian, China

ABSTRACT

Lane changes are particularly frequent on urban streets, which not only impact traffic operations, but also cause negative effects on safety. This study uses vehicle trajectory data collected from Southwest Road in Dalian, China, to investigate the influential factors of discretionary lane changing in an urban road environment. Both the standard logit model and mixed logit model were fitted to the data, which evidently identified several key factors. The mixed logit model outperformed the standard logit model in terms of model fit. The results suggest that driver heterogeneity is present on the speed differentials of the subject vehicle and the leading vehicles, and the distance gap in the target lane. A sensitivity analysis was further conducted to quantify the degree of influence of the statistically significant variables. The findings support the notion and purpose of discretionary lane changing, i.e. seeking a speed advantage and/or a more satisfactory driving environment.

KEYWORDS

Discretionary lane changing; urban street; driver heterogeneity; mixed logit model; sensitivity analysis

Introduction

Changing lanes is a basic driving task, but it is also one of the most important decisions on the part of the driver. Improper lane-changing behavior not only impacts traffic flow characteristics but also threatens traffic safety. Studies have shown that lane changing is one of the main causes for escalating local traffic oscillations to large-scale disturbances that lead to traffic congestion or breakdowns (Ahn and Cassidy 2007; Zheng, Ahn, and Monsere 2010; Sun, Zhang, and Zhang 2014). During lane changes, drivers need to process a large amount of information in a short period of time, making them prone to errors and accidents. It is necessary to conduct comprehensive and in-depth research on drivers' lane changing decisions and behaviors in order to improve traffic operations and safety.

Lane changes generally can be divided into two categories based on purpose: mandatory lane changing (MLC) and discretionary lane changing (DLC). The former refers to the driver's lane-changing behavior when he/she has to leave the current lane in order to reach the planned destination, whereas the main purpose of DLC is to gain an advantage in speed or improve driving conditions (e.g. greater driving space), and changing lane is not necessary (Zheng 2014). Compared with MLC, DLC is more complex due to the flexibility and uncertainty. In some cases, it is difficult to distinguish between the two; for example, a driver may change to the adjacent left lane to prepare for a left-turn at the downstream intersection long before the actual left-turn is executed. A more accurate way to distinguish between the two is to model them separately, as drivers are expected to think and act differently due to different purposes.

As one of the most vital components in traffic flow theory, the lane-changing model has been fully incorporated in the microscopic traffic simulation (Yang, Koutsopoulos, and Ben-Akiva 2000; Choudhury et al. 2007; Laval and Leclercq 2008). Though more research on the lane-changing model is needed, progress is limited because lane-change behavior is more complicated than other behaviors such as car-following. Moreover, large-scale data is lacking for analyzing, modeling and calibrating such behaviors, and the collection and compilation of a large trajectory dataset for model development is costly and time-consuming (Moridpour, Sarvi, and Rose 2010; Rahman et al. 2013; Zheng, Suzuki, and Fujita 2014). Another issue with lane-changing modeling is that the heterogeneity of driver behaviors is often neglected. In many existing models, drivers are assumed to behave in the same way for a given situation, when in reality the factors affecting the lane changing behavior can possibly be given different importance by different drivers. For example, an aggressive driver may perceive the speed advantage to be more important, so he/she is more likely to execute a lane change. By incorporating such effects into modeling, the results should more accurately reflect the actual driving behaviors.

The objective of this paper is to develop a lane-changing model to quantitatively analyze the factors affecting drivers' lane-changing decisions, specifically focusing on discretionary lane changes of cars in an urban road environment. Models in this paper were based on vehicle trajectory data collected from urban streets. To account for a complex traffic environment, probabilistic decision models were developed based on the random utility theory. A mixed logit model was fitted and compared with the standard logit model to accommodate heterogeneity among drivers.

The remainder of this paper is organized as follows. Section 2 is a literature review on lane-changing models, followed by the methodology in Section 3. Section 4 comprises data collection and analysis. Model estimation and validation are presented in Section 5 and the impacts of factors on the lane changing are evaluated and discussed in Section 6. Finally, conclusions and recommendations for future work are provided in Section 7.

Literature review

According to Rahman et al. (2013), there are four categories of lane-changing models: ruled-based, discrete-choice-based, artificial

CONTACT Ronghan Yao cyanyrh@dlut.edu.cn School of Transportation and Logistics, Dalian University of Technology, Room 510, No. 4 Lab Building, 2 Linggong Rd., Dalian 116024, China

intelligence, and incentive-based. Among the existing models, rule-based and discrete-choice-based models appear to be the most popular ones.

The ruled-based model treats the lane-changing process as a decision tree with a series of fixed conditions. The result is a binary choice (change or not change), but the model parameters are difficult to calibrate. As one of the earliest rule-based models, Gipps' model included several important factors, such as the existence of a safety gap, the location of permanent obstructions, the intent of turning movement, the presence of heavy vehicles, and the speed advantage (Gipps 1986). Drivers decided whether or not to execute a lane-changing manoeuver by considering the physical possibility or safety, the necessity, and the desirability to change lanes. Although Gipps' model has been applied in several microscopic traffic simulations, it has not been validated using microscopic traffic and driver behavior data. In Deutsch's research, the cellular automata model was developed, assuming that a vehicle makes lane changes after assessing speed conditions in the current and target lanes as well as the availability of sufficient space to execute the lane change (Deutsch et al. 1995). Kita (1999) proposed a game theory model based on the give-way behavior in a merging situation.

Hidas (2002, 2005) classified lane-changing maneuvers as either free, forced, or cooperative, depending on the gaps between the lead and lag vehicles in the target lane. A free lane change was considered to be feasible if the target lead and lag space gaps were both greater than the desired critical space gaps. A cooperative lane change relied on the willingness and feasibility of the lag driver to slowdown. The forced lane change was different from a cooperative lane change in the maximum speed decrease and deceleration assumed by the subject vehicle driver. Hidas implemented and tested his model using the ARTEMiS microscopic traffic simulator, but there was no framework for calibrating model parameters. In addition, some factors that may influence lane changes, such as a heavy vehicle, were not considered.

The discrete-choice-based model assumes lane-changing decisions are made on the basis of maximum utility, and the output is a probabilistic result (Moridpour, Sarvi, and Rose 2010; Rahman et al. 2013). It was first proposed to describe the lane changing by Ahmed et al. (1996). In this lane-changing model, the utility function of the current lane and adjacent lane would be estimated if the driver was not satisfied with the current driving condition. The probability function to determine the lane-changing decision could then be calculated. The lead and lag gaps were checked to determine whether a lane change should be executed. Ahmed estimated the parameters of his model for a special case using vehicle trajectories: merging to the left lane from a freeway on-ramp but in reality, it is difficult to determine the utility functions for the various decision choices for Ahmed's model.

Toledo et al. (2003, 2009) developed an integrated probabilistic lane changing model based on utility theory considering mandatory and discretionary lane changing at the same time. Variables underlying lane-changing decisions were considered, such as gaps, speeds, distance from the intended exit off-ramp, avoiding the nearest lane to the shoulder, driving styles driving capabilities, etc. The general framework for estimating the probability of a lane change and its execution was similar to Ahmed's model. Sun et al. (2010) conducted a focus group study in which driver types, the likelihood of attempting a given discretionary lane change, and the factors affecting the execution of a given lane changing maneuver were revealed. An 'in-vehicle' experiment was also conducted. Probability functions for each lane change scenario were established, incorporating the driver characteristics which most previous models did not take into account (Sun and Elefteriadou 2012).

Based on the literature, while many studies have involved the modeling of lane-changing behavior, few have accounted for the effects of drivers' heterogeneity on the lane-changing process. In fact, behavioral variation does exit among different drivers. While observed characteristics such as age and gender may to some extent explain the variation, differences are caused by drivers' various perceptions and preferences in the choice-making process, which must be further researched.

Methodology

The rule-based model is more difficult to calibrate than the discrete-choice-based model. In this paper, using the random utility theory, the discrete-choice-based model was developed to explore the influential factors of the lane-changing behavior and quantitatively evaluate their impacts.

Random utility theory

Random utility theory is based on utility maximization hypothesis, in which the individual will always choose the alternative that he/she thinks has the largest utility. Assuming that there are J independent alternatives in a choice set A_{n}, and each alternative corresponds to a certain utility, according to this theory, the individual n will choose the alternative i if the following condition holds true:

$$U_{in} > U_{jn};\ i \neq j;\ i, j \in A_n \tag{1}$$

where U_{in}, U_{jn} are the utilities for the individual n of alternatives i and j, respectively.

The utility U_{in} of the alternative i for the individual n consists of two components, the observed utility V_{in} and the unobserved error term ε_{in}, given by:

$$U_{in} = V_{in} + \varepsilon_{in} \tag{2}$$

The V_{in} is observable in that it is a function of the observable characteristics of the individual n and the alternative i. A linear utility function has often been assumed, as shown in Equation (3).

$$V_{in} = \beta_0 + \sum_{k=1}^{N} \beta_k X_{ink} \tag{3}$$

where X_{ink} is the k$^{\mathrm{th}}$ attribute variable for individual n and the alternative i; N is the total number of attributes studied; β_0 is the constant term; β_k is the coefficient for the k$^{\mathrm{th}}$ attribute variable.

Based on the utility maximization hypothesis, the probability of any alternative i being selected by the individual n from the choice set A_{n} (denoted as P_{in}) is given by:

$$\begin{aligned} P_{in} &= \mathrm{Prob}(U_{in} > U_{jn};\ i \neq j;\ i, j \in A_n) \\ &= \mathrm{Prob}(V_{in} + \varepsilon_{in} > V_{jn} + \varepsilon_{jn};\ i \neq j;\ i, j \in A_n) \end{aligned} \tag{4}$$

$$0 \leq P_{in} \leq 1\,,\ \sum\nolimits_{i \in A_n} P_{in} = 1 \tag{5}$$

The error term ε_{in} is a random variable described by a probability distribution. Assuming it to be independently and identically distributed, following the double exponential distribution, it yields the logit model, and the following equation can be derived:

$$P_{in} = \frac{\exp(V_{in})}{\sum_{j \in A_n} \exp(V_{jn})}\,,\ i \in A_n \tag{6}$$

Mixed logit model

In a standard logit model, the heterogeneity among individuals is not considered because the parameters in the utility functions are all fixed. A mixed logit model obviates the limitations of the standard logit model by allowing for random taste variation and correlation in unobserved factors over time (Train 2003). The mixed logit probabilities are the integrals of the standard logit probabilities over a density of parameters, of which some may be fixed and some may be randomly distributed. A mixed logit model can be expressed as:

$$P_{in} = \int L_{in}(\beta) f(\beta|\theta) d\beta \tag{7}$$

where $L_{in}(\beta)$ is the logit probability as expressed in equation (6), evaluated at parameters β; $f(\beta|\theta)$ is the density function of β, and θ refers collectively to the parameters to be estimated of the distribution (e.g. the mean and the variance of β). If observable utility V_{in} is linear in β, the mixed logit model can be written as

$$P_{in} = \int \frac{\exp(\beta X_{in})}{\sum_{j\in A_n} \exp(\beta X_{jn})} f(\beta|\theta) d\beta \tag{8}$$

The mixed logit probability is a weighted average of the logit formula evaluated at different values of β, with the weights given by the density $f(\beta|\theta)$. If the βs are all fixed, the mixed logit reduces to the standard logit. Thus, the heterogeneity across individual decision makers can be introduced by specifying the distribution of β. Normal, lognormal, uniform and triangular distributions are usually used.

The choice probability in (7) or (8) cannot be calculated exactly because the integral does not have a closed form in general. Alternatively, the integral is approximated through simulation. For any given value of θ, a value of β is drawn from $f(\beta|\theta)$,and $L_{in}(\beta)$ can be calculated. Repeating the process for many times, the average of the results is taken as the simulated probability:

$$\tilde{P}_{in} = \frac{1}{R} \sum_{r=1}^{R} L_{in}(\beta^r) \tag{9}$$

where r is the number of draws. $\tilde{P}_{in}$ is an unbiased estimator of P_{in}.

Model development

In a lane-changing maneuver, the vehicle that intends to change lanes (referred to as the subject vehicle) will interact with its surrounding vehicles, especially the immediate leading and following vehicles (if there are any) in the lane in which it is traveling (current lane) and the lane to which it's going to change (target lane). The lane-changing scenario is depicted in Figure 1, which involves at least four vehicles: the subject vehicle (denoted as S), the vehicles in front of the subject vehicle in the current lane and in the target lane (denoted as CL and TL, respectively), and the vehicles following behind the subject vehicle in the target lane (denoted as TF). The speeds of these vehicles and the distances between them are factors that may affect the driver's decision to change lanes, so they are considered as the variables in the models. Specifically, in the target lane we use the distance between the TL and TF (D_{TLF}) to capture drivers' gap acceptance instead of considering the distances between the TF and S, and S and TL separately. Based on our observations, a large D_{TLF} usually stimulates a driver to change lanes, even if TF is parallel to the subject vehicle. Chinese drivers will accelerate and cut in front of TF with very small back gaps. Thus, using D_{TLF} may be more appropriate to model drivers' lane-changing behaviors in this case because drivers seem to be more sensitive to D_{TLF} than to the distances with the TL and TF. Besides, the influence of buses in urban areas should not be neglected because of their frequent stops, slow speed, and large size that can block the view of the vehicles following them.

Based on random utility theory, a driver's satisfaction of driving in a lane can be represented by the utility of the lane, which obeys the utility maximization hypothesis. In other words, the driver will generate the desire to change lanes once the satisfaction of driving in the current lane is lower than the adjacent lane. There are two choices: to stay in the current lane or change to the target lane; thus, the logit model is reduced to a binary logit model. The observable utilities for staying in the current lane (V_C) and for changing to the target lane (V_T) can be expressed as linear functions of the variables discussed above.

$$V_C = \beta_0 + \sum_{k=1}^{N1} \beta_k X_{Ck} \tag{10}$$

$$V_T = \sum_{k=1}^{N2} \beta_k X_{Tk} \tag{11}$$

where X_{CK} and X_{TK} are the variables included in the utility functions of the current lane and the target lane, respectively; N_1 and N_2 are the number of variables in the utility functions of the current lane and the target lane, respectively; β_0, β_k were the coefficients to be estimated. Then, the probability of changing into the target lane (P_T) can be calculated as:

$$P_T = \frac{\exp(V_T)}{\exp(V_C) + \exp(V_T)} \tag{12}$$

Accounting for drivers' heterogeneity, the probability of changing into the target lane (P_T) is:

$$P_T = \int \frac{\exp(V_T)}{\exp(V_C) + \exp(V_T)} f(\beta|\theta) d\beta \tag{13}$$

where $f(\beta|\theta)$ is the density function of β, as defined previously.

The variables used in this study are specified in Table 1. Two dummy variables, BUS and P_S, were created for the existence of a bus in front of the subject vehicle in the current lane and for the position of the subject vehicle.

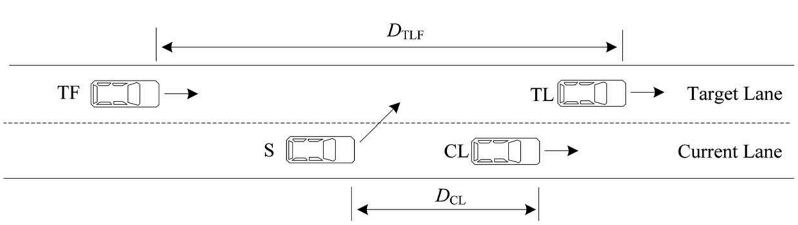

Figure 1. Schematic diagram of a lane changing.

Table 1. Variable specification.

Variable	Definition	Unit
LC	Choice by the subject vehicle to either change a lane (T) or remain in the current land (C)	–
V_S	speed of the subject vehicle (S)	km/h
V_{CL}	speed of the leading vehicle in current lane (CL)	km/h
V_{TL}	speed of the leading vehicle in target lane (TL)	km/h
V_{TF}	speed of the following vehicle in target lane(TF)	km/h
D_{CL}	distance between vehicles S and CL	m
D_{TLF}	distance between vehicles TL and the vehicle TF	m
$\varDelta V_{CL}(= V_{CL} - V_S)$	speed difference between the vehicle CL and the vehicle S	km/h
$\varDelta V_{TL}$ $(= V_{TL} - V_S)$	speed difference between the vehicle TL and the vehicle S	km/h
$\varDelta V_{TF}$ $(= V_S - V_{TF})$	speed difference between the vehicle S and the vehicle TF	km/h
BUS	Existence of a bus in front of the vehicle S in current lane. BUS = 1, if there is a bus in front of S; BUS = 0, otherwise	–
P_S	The position of the subject vehicle S. P_S = 1, if S is in the rightmost lane of the study lanes; P_S = 0, otherwise	–

Data collection and processing

Selecting the study sites for data collection is the first and crucial step. Study sites were screened following principles such as availability of a commanding position nearby (e.g. a footbridge or a tall building which provides good views of the study site for video recording), minimum lateral interference to the traffic flow, and distinguishable mandatory and discretionary lane changes through video data. One segment of Southwest Road in Dalian, China was chosen, as illustrated in Figure 2. The study area is about 180 m in the length. The segment has four lanes in each direction, and data were collected on the direction from southwest to northeast (the study direction). The four lanes in this direction are denoted as 1–4 from the rightmost lane. The Huanghe interchange (drawn in the blue color in Figure 2) is in front of the study area in that direction. All through vehicles take the interchange by Lane 3 or 4, and vehicles going other directions travel under the interchange by lane 1 or 2. However, because vehicles in lane 1 were observed to be greatly affected by bus stops, only lane changes occurring between Lane 3 and lane 4 were considered discretionary lane changes.

Data were collected in October and November of 2015. Video recording was conducted from 7am to 10am on four weekdays with clear weather and good visibility conditions. Two video cameras were placed on the footbridge to record the traffic flow on both sides of it. It should be noted that during the time period, traffic flow in the study direction was relatively large but not congested, thus making it easier to observe a sufficient number of lane changes.

The videos were processed to extract vehicle trajectories. Detailed trajectories of vehicles with or without lane changes were extracted. First, vehicles conducting a lane change between Lane 3 and 4 were identified as a subject vehicle. Only the vehicles making one lane change were considered; so if a vehicle changed from Lane 4 to Lane 3 and then further changed to Lane 2 (two lane changes), this vehicle would not be counted as a subject vehicle. Next, the surrounding vehicles (CL, TL and TF) were identified. The time and the position of each vehicle could be extracted from the video, based on which the speeds of these vehicles and the distances between them could be obtained. It was noted if a bus was positioned in front of the subject vehicle. In addition to collecting the lane-changing vehicles, vehicles making no lane changes were also selected from Lane 3 and 4 as well as their surrounding vehicles.

George 2.1, a video image processing software developed by Nagoya University, was used to collect vehicle trajectory data. The software needs to be calibrated before use. At least five reference points are added on the video screen and the coordinates of the points measured from the real world are entered to transform the coordinate system. The software calculates eight coordinate conversion parameters, and uses them to estimate the real coordinates of the reference points. The estimated values and observed values are compared, and R2 and adjusted R2 are used to evaluate the goodness-of-fit. If the accuracy is low, there may be an error in the location of some reference points, and we need to identify them and modify or remove them. New reference points can be also

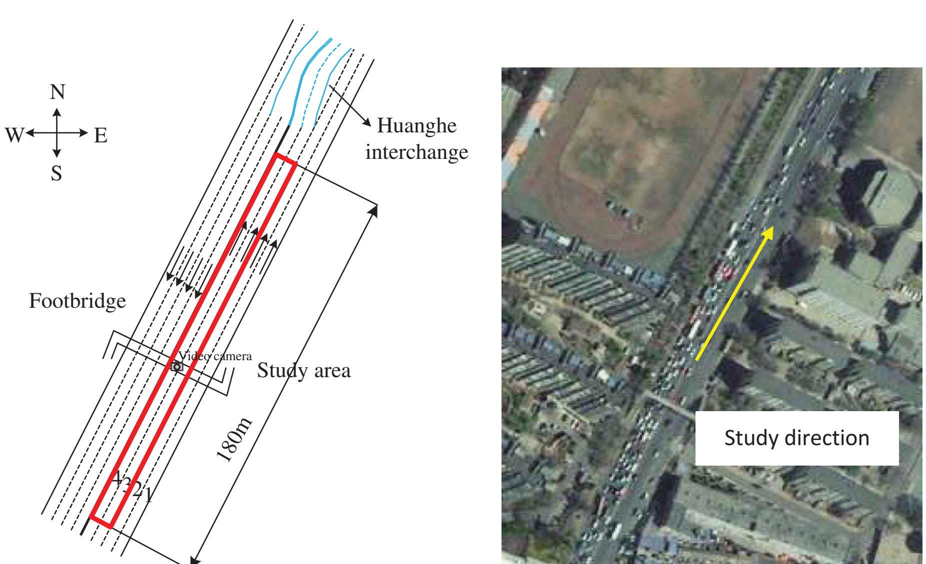

Figure 2. The schematic illustration of the study site.

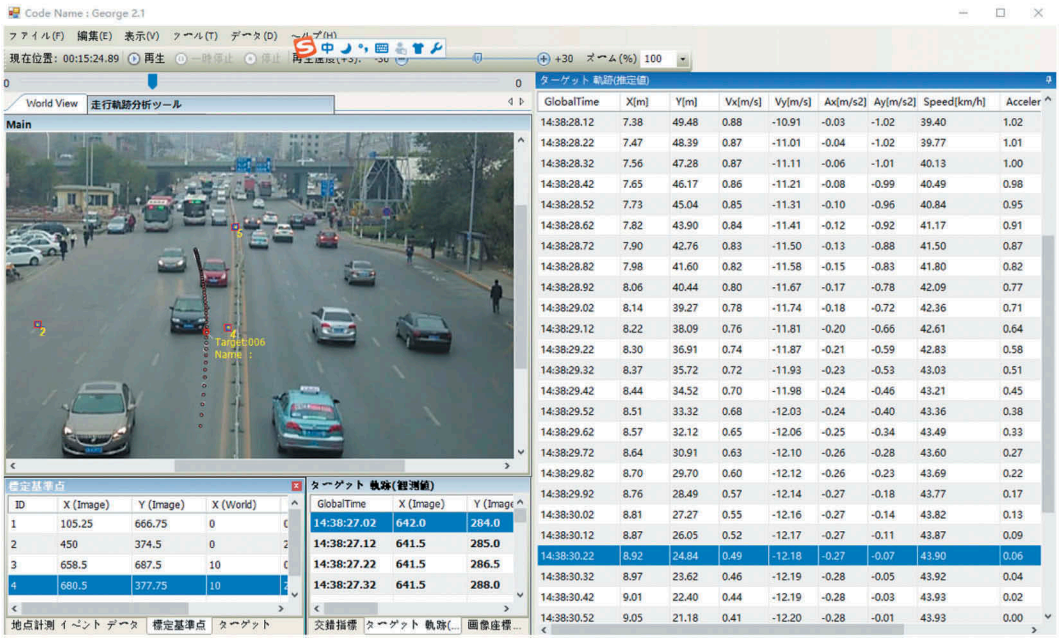

Figure 3. The interface of the software George.

added. The process will be repeated till the results are satisfactory. In our study, we used the adjusted R2 larger than 0.99 as the criteria to ensure the accuracy of the extracted trajectories. The target vehicles can be manually traced in every one-tenth of a second, and the coordinates of the vehicle trajectory are obtained, based on which the distances and speeds are computed, as shown in Figure 3. The performance and accuracy of the software have been evaluated for practical use (Suzuki and Nakamura 2006).

Trajectories for 1632 vehicles were collected, which constituted a total of 408 samples, including 145 lane-changing samples and 263 no-lane-changing samples. For each lane-changing sample, the vehicles including S, CL, TL, and TF are traced from 2 to 3 seconds prior the start of the lane change (i.e. the vehicle starts to turn) to 2–3 seconds after the completion of the lane change (i.e. the vehicle completely enters the target lane and goes straight). For each no-lane changing sample, the time interval is about 2–3 seconds because the vehicles speeds are generally stable.

Field speed measurements in the study area were also conducted on the footbridge while doing the video recording and speeds of 518 vehicles in the study area of Lane 3 and 4 were recorded using radar guns. We assume that if a vehicle is not disturbed by others, it will keep the same speed within the study area. As shown in Figure 4, the distribution of the extracted speeds was plotted and verified to be generally consistent with that of the measured speeds. The descriptive statistics of the variables are listed in Table 2.

Model results and discussion

Model estimation

The dataset was randomly divided into two subsets: dataset 1, including 104 lane-changing samples and 202 no-lane-changing samples, was used for model estimation, and dataset 2, including 41 lane-changing samples and 61 no-lane-changing samples, was used for model validation.

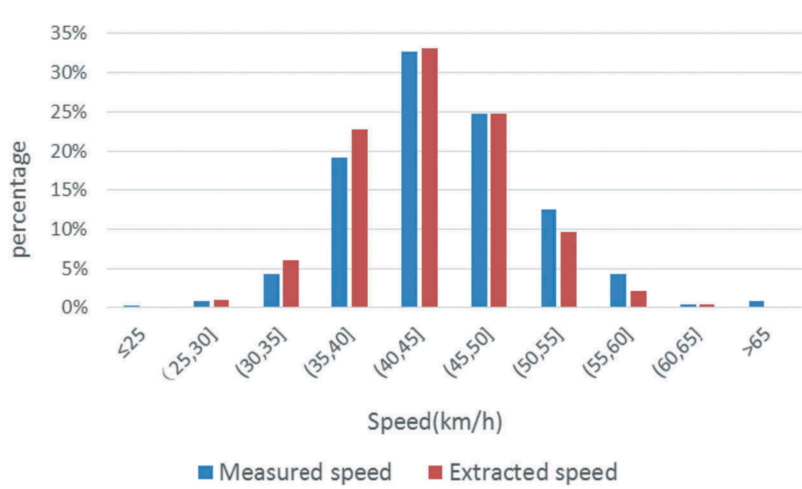

Figure 4. The distribution of measured and extracted speeds in the study area.

Table 2. The descriptive statistics of the variables.

	Variables	Unit	Min.	Max.	Average	Standard deviation
Measured speed	V_R	km/h	25.00	68.00	44.82	6.48
All samples	V_S	km/h	24.07	59.37	41.77	5.63
	V_{CL}		27.46	60.00	42.27	5.92
	V_{TL}		29.29	64.63	44.15	5.62
	V_{TF}		14.39	64.18	39.37	7.65
	ΔV_{CL}		−20.76	16.77	0.46	5.18
	ΔV_{TL}	km/h	−18.55	25.64	2.38	7.01
	ΔV_{TF}		−25.08	34.08	2.41	8.63
	D_{CL}	m	8.95	55.11	22.79	7.76
	D_{TLF}		16.67	144.57	46.26	18.81
Lane-changing samples	V_S	km/h	24.07	59.37	40.94	6.99
	V_{CL}		27.46	59.3	39.82	5.98
	V_{TL}		29.29	64.63	45.57	5.68
	V_{TF}		14.39	61.69	38.00	8.02
	ΔV_{CL}	km/h	−20.76	12.04	−1.12	6.10
	ΔV_{TL}		−18.55	25.64	4.63	7.24
	ΔV_{TF}		−25.08	34.08	2.94	9.88
	D_{CL}	m	8.95	55.11	20.18	8.74
	D_{TLF}		20.53	144.57	49.32	20.85
No-lane-changing samples	V_S	km/h	26.37	57.47	42.23	4.67
	V_{CL}		29.58	60	43.62	5.45
	V_{TL}		30.53	63.73	43.37	5.43
	V_{TF}		19.88	64.18	40.12	7.34
	ΔV_{CL}	km/h	−9.51	16.77	1.34	4.36
	ΔV_{TL}		−14.4	21.68	1.13	6.58
	ΔV_{TF}		−19.45	21.11	2.11	7.87
	D_{CL}	m	10.77	47.71	24.24	6.76
	D_{TLF}		16.67	100.05	44.58	17.41

Firstly, standard logit models were fitted to the data, in which the lane utilities V_C and V_T are expressed as linear functions of the variables that are expected to have an influence on them. All variables listed in Table 1 were tested in the models. When it comes to the variables concerning the speeds, either the speeds of the individual vehicles (V_S, V_{CL},, V_{TL}, V_{TF}) or the relative speeds between the subject vehicle and its surrounding vehicles (ΔV_{CL}, ΔV_{TL}, ΔV_{TF}) were included in the model. The models were estimated using the maximum likelihood method. Based on the t-values of the coefficient estimates, the speed and position of the subject vehicle (V_S and P_S) are not significant at the confidence level of 95%. Excluding the insignificant variables, the 'best' model was selected, using the relative speeds instead of the absolute speeds based on the model fit. The utility functions took the forms of

$$V_C = \beta_0 + \beta_1 \Delta V_{CL} + \beta_2 D_{CL} + \beta_3 BUS \tag{14}$$

$$V_T = \beta_1 \Delta V_{TL} + \beta_4 \Delta V_{TF} + \beta_5 D_{TLF} \tag{15}$$

where V_C and V_T were the observable utilities of the current lane and the target lane, respectively; β_S (s = 0 ~ 5) were the coefficients to be estimated; other variables were as defined previously.

Similar procedures were followed in the fitting of the mixed logit models. The differences were that in this case the βs might no longer be fixed values, but can be randomly distributed. In this paper, normal, lognormal, and uniform distributions were considered and tested for the coefficients to find the most suitable one. The mixed logit models were estimated with simulation-based maximum likelihood methods. Five hundred Halton draws were used to accomplish the simulation because Halton draws were more efficient and involved far fewer draws to achieve convergence as opposed to random draws. Based on the results, the variables V_S and P_S were not significant, which was consistent with the result of the standard logit model. It was also found that by assuming the coefficient of speed difference of the subject vehicle and the leading vehicle in the current/target lane (β_1) a uniform distribution and the coefficient of the distance gap in the target lane (β_5) a lognormal distribution, the model fit was better than others.

The two 'best' models selected from the above process are listed in Table 3. All variables except the constant term in the model are significant at the confidence level of 95% or 90%, indicating that all have a significant impact on a driver's lane-changing behavior.

Discussion

The speed of the subject vehicle, V_S, is not statistically significant when other speed-related variables are included in the model. This is probably because the subject vehicle cannot arbitrarily determine its speed but must adjust the speed according to that of the leading vehicle, which results in a high correlation between the two. And drivers may be more concerned with the relative speed than the absolute speed. It also shows that the position of the subjective vehicle, i.e. whether it is in Lane 3 or 4, does not have an influence on the utility of the current lane.

In the standard logit model, the positive coefficients for ΔV_{TL}, ΔV_{TF}, and D_{TLF} suggest that the driver tends to change into the target lane when the speed of the leading vehicle in the target lane is higher than the subject vehicle, when the speed of the subject vehicle is higher than the vehicle following in the target lane, and/or when the distance gap in the target lane increases. The positive coefficients for ΔV_{CL} and D_{CL} indicate that the driver's satisfaction with the current driving state enhances as the speed and distance of the vehicle in front of it increases. The negative coefficient for *BUS* signals the existence of a bus in front of the vehicle will reduce the utility of the current lane.

Odds ratios (ORs) were computed for the interpretation of the results. The ORs of ΔV_{TL} and ΔV_{TF} are 1.117 (95% CI [1.069, 1.166]) and 1.060 (95% CI [1.025, 1.095]), respectively. This suggests that the odds of executing a lane change will increase by 1.117 and 1.060 when ΔV_{TL} and ΔV_{TF} increase by 1 km/h, respectively. The OR of D_{TLF} is 1.017 (95% CI [1.002,1.033]), meaning that the odds is 1.017 times larger with 1 m increase in the distance gap in the target lane. Meanwhile, the ORs of ΔV_{CL} and D_{CL} are 0.896 (95% CI [0.858, 0.935]) and 0.915 (95% CI [0.877, 0.954]), both of which are smaller than 1, indicating that that lane-changing is less likely to occur with an increase of speed and distance of the leading vehicle in the current lane.

In the mixed logit model, the estimates of the fixed coefficients are similar to those in the standard logit model. The standard deviations of two random coefficients are significant, indicating the existence of drivers' heterogeneity with regard to these two factors: (1) the speed difference with the leading vehicle in the current/target lane (ΔV_{CL}, ΔV_{TL}), and (2) the distance gap in the target lane (D_{TLF}). The estimation of the random coefficients are shown in Table 4. When looking at the speed difference with the leading vehicle in the current/target lane (ΔV_{CL}, ΔV_{TL}), the result shows that the coefficient (β_1) is smaller than 0.0963, 0.2973 and 0.4983 for 25%, 50% and 75% of the drivers, respectively. The maximum value is 0.6693, which represents those drivers who see the speed advantage as more important than other drivers. These drivers may make more risky lane changes in pursuit of increase speed. It should be noted that there are also a proportion of drivers who have negative coefficients for the speed difference with the minimum value of −0.1047. In real driving process, there are drivers who are very conservative, not pursuing the high speed, but willing to drive at a comfortable speed. There are also drivers who change to the adjacent lane only to seek acceleration

Table 3. Estimation results of the models.

	Standard logit model			
Variables	Coefficient estimate	Std. Error	t-value	Pr(>|t|)
Constant	0.0570	0.5629	0.1013	0.9193
$\varDelta V_{CL}$, $\varDelta V_{TL}$	0.1103	0.0221	4.9866	<0.0001**
$\varDelta V_{TF}$	0.0580	0.0169	3.4335	0.0006**
D_{CL}	0.0891	0.0215	4.1514	<0.0001**
D_{TLF}	0.0171	0.0077	2.2321	0.0256**
BUS	−3.4070	1.1432	−2.9802	0.0029**
Log-Likelihood		−153.38		
ρ^2 (adjusted)		0.2180		
	Mixed logit model			
Variables	Coefficient estimate	Std. Error	t-value	Pr(>|t|)
Constant	−0.3699	1.2704	−0.2912	0.7709
$\varDelta V_{CL}$, $\varDelta V_{TL}$	0.2973	0.0758	3.9242	<0.0001**
$\varDelta V_{TF}$	0.1434	0.0503	2.8490	0.0044 **
D_{CL}	0.2198	0.0636	3.4590	0.0005 **
D_{TLF}	−4.0845[a]	0.9657 [a]	−4.2297	<0.0001**
BUS	−6. 8079	2.4501	−2.7786	0.0055**
SD. $\varDelta V_{CL}$($\varDelta V_{TL}$)	0.4020	0.2241	1.7941	0.0728*
SD. D_{TLF}	1.6544	0.5555	2.9780	0.0029 **
Log-Likelihood		−148.63		
ρ^2 (adjusted)		0.2422		

**: significant at the confidence level of 95%; *: significant at the confidence level of 90%; SD.: the standard deviation; [a] the parameter estimates of ln(β_5).

Table 4. Estimation of the random coefficients.

	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
$\varDelta V_{CL}$, $\varDelta V_{TL}$	−0.1047	0.0963	0.2973	0.2973	0.4983	0.6993
D_{TLF}	0	0.0055	0.0168	0.0661	0.0513	Infinite

space to overtake the vehicle in front, so they may temporarily ignore the speeds of the leading vehicles. For the two types of drivers discussed above, they can possibly exhibit a negative coefficient for the speed difference. In the standard logit model, this coefficient is positive and fixed, indicating that the larger the speed difference of the subject vehicle and the leading vehicle in the current/target lane, the larger the utility of the current/target lane is. Comparatively, the mixed logit model estimates are more consistent with the actual driving behavior.

The random coefficient for the distance gap in the target lane (D_{TLF}) (β_5) is a lognormal distribution. In other words, the logarithm of β_5 follows a normal distribution with the mean u and the variance σ^2. The estimated values of u and σ of this normal distribution are shown in Table 3. The mean and variance of β_5 can then be derived by $\exp(u+\ \sigma^2/2)$ and $\exp(2u+\ \sigma^2)(\exp(\sigma^2)\text{-}1)$, respectively. As shown in Table 4, the positive coefficient implies the larger the gap in the target lane, the greater the possibility of a lane change. The median of the coefficient is 0.0168, which is close to the coefficient 0.0171 in the standard logit model. However, in the mixed logit model the coefficient is not fixed but smaller than 0.0168 for 50% of the drivers and larger than that for the other 50%. By assuming the coefficient a random parameter, it offers more flexibility to capture drivers' heterogeneity in gap acceptance, and helps understand the effects of gap acceptance on drivers' lane-changing choice.

The coefficient of McFadden (ρ^2) is used to evaluate the overall goodness of fit of the model and a higher value of ρ^2 means better fit. Comparing the two models, the mixed logit model with a ρ^2 of 0.2422 fits the data better than the standard logit with a ρ^2 of 0.2180.

Model validation

The mixed logit model with better model fit was used for model validation. Firstly, the calibrated model was used to compute the probability of changing into the target lane or remaining in the current lane for dataset 1. In order to evaluate the accuracy of the results, the probability values must be transformed into options (i.e. changing lanes or not changing lanes). When the calculated probability is smaller than a value, the driver would choose one option, and if greater, the other. Different split values, such as 30%, 40% or 50% were tested. In this case, a value of 30% means that if the probability of changing lanes is greater than or equal to 30%, a lane change is predicted. The percentage of correct predictions for the training dataset is listed in Table 5. Results suggest that choosing a smaller split value will lead to higher accuracy for the lane-changing samples but lower accuracy for the no-lane changing samples, and vice versa. The percentage of correct predictions for all samples is larger when the split value is 40%. For model validation, 40% is selected as the split value.

Using the split point of 40%, the percentage of correct predictions for dataset 2 (the validation dataset) is listed in Table 5. The correct prediction for the lane-changing samples and non-lane-changing samples are 73.2% and 82.0%, indicating that the calibrated model can be used to predict the lane-changing behavior for the new dataset.

Sensitivity analysis

Although the parameter estimates of the logit model can provide whether or not the variables significantly affect the driver's lane-changing behavior and the weight of each variable in the utility function, it is difficult to judge the degree to which the variables affect the probability of different options. In order to quantitatively evaluate the influence of each variable, analyses were conducted to evaluate the sensitivity of an individual's choice probability to a change in the value of a variable.

The probabilities were predicted by the mixed logit model with 500 Halton draws. Figure 4 shows the effects of $\varDelta V_{CL}$ and $\varDelta V_{TL}$

Table 5. Percentage of correct predictions.

	Split value	Lane-changing samples	No-lane-changing samples	All samples
Dataset 1	≥30%	75.0%	73.8%	74.2%
	≥40%	62.5%	85.2%	77.5%
	≥50%	53.9%	89.1%	77.1%
Validation (Dataset 2)	≥40%	73.2%	82.0%	78.4%

on the probability of lane changing, assuming that the other variables are fixed and take the average values of all samples in Table 2. The values of $\varDelta V_{CL}$ and $\varDelta V_{TL}$ range from −30 to 30. The increase of the speed difference between the leading vehicle in the current lane and the subject vehicle ($\varDelta V_{CL}$) will reduce the lane-changing probability On the contrary, if the speed difference between the leading vehicle in the target lane and the subject vehicle ($\varDelta V_{TL}$)„ the lane-changing probability will increase accordingly. The probability surface (with the maximum and minimum equal to 0.09 and 0.83, respectively), is divided into different regions with clear boundaries. When $\varDelta V_{TL}$ is not larger than $\varDelta V_{CL}$, the probability of a lane change is less than 0.2. Similarly, when the differences of $\varDelta V_{TL}$ and $\varDelta V_{CL}$ are 0–6, 6–12 and 12–40 km/h, the probabilities are in the ranges of 0.2–0.4, 0.4–0.6, and 0.6–0.8, respectively. The results indicate that seeking a speed advantage is one of the main reasons for a lane change.

Figure 5 shows the probability of a lane change for $\varDelta V_{TL}$, $\varDelta V_{CL}$, and D_{CL} in scenarios with different values for D_{TLF}, $\varDelta V_{TF}$ and *BUS*. For each scenario, the other variables are set to the average values of all samples in Table 2. The figure indicates how D_{TLF}, $\varDelta V_{TF}$, and *BUS* influence the decision of whether or not to make a lane change. For example, in Figure 6(a), the lane-changing probability increases with $\varDelta V_{TL}$, but when D_{TLF} and $\varDelta V_{TF}$ take different values, the probabilities are not the same. In general, the higher the value of D_{TLF} or $\varDelta V_{TF}$, the greater the probability. In Figure 6(b,c), although the lane-changing probability decreases with the increase of $\varDelta V_{CL}$ and D_{CL}, the effects of D_{TLF} and $\varDelta V_{TF}$ on the lane-changing choice are the same as in Figure 6(a). Take D_{TLF} for instance, when $\varDelta V_{TL}$ and $\varDelta V_{CL}$ values between −10 and 10 km/h, and D_{CL} between 20 and 40 m, the impact of D_{TLF} on the probability is greater, particularly when D_{TLF} is small. When $\varDelta V_{TL}$, $\varDelta V_{CL}$ and D_{CL} are not within the above range, the effect of D_{TLF} is relatively small. This indicates that a driver may not change lanes for a small gain in speed even when D_{TLF} is enough; when the speed advantage is high, a driver may be highly likely to accept a small gap. Similarly, when $\varDelta V_{TL}$, $\varDelta V_{CL}$ and D_{CL} are in the ranges of 0–20 km/h, −20–0 km/h and 0–40 m, respectively, $\varDelta V_{TF}$ has a significant impact on the probability of a lane change. The results suggest that in these cases, $\varDelta V_{TF}$ plays a key role in drivers' lane-changing behavior. In other cases, drivers are more sensitive to $\varDelta V_{TL}$, $\varDelta V_{CL}$, and D_{CL} when making their lane-changing decisions. In addition, the presence of buses affect the probability significantly. When $\varDelta V_{TL}$ is greater than 0, $\varDelta V_{CL}$ is less than 0, or D_{CL} is less than 25 m, the probability is close to 1with the existence of a bus in the current lane.

Conclusions

In this paper, the discretionary lane-changing behavior was collected from an urban street segment on Southwest Road in Dalian, China. Logit models were developed using data extracted from vehicle trajectories to predict the lane-changing decisions. The mixed logit model has a better goodness of fit for the data than the standard logit model, and provides more information on drivers' lane-changing preferences. Drivers exhibit heterogeneity in (1) the speed difference between the leading vehicle in the current/target lane and the own vehicle and (2) the gap in the target lane. For most drivers, the larger the difference of speed difference between the leading vehicle in the current/target lane and the subject vehicle, the larger the utility of the current/target lane is. However, there are still a portion of drivers who may do the opposite. The gap in the target lane guarantees a safe lane-changing maneuver. It is natural for drivers to have different driving styles (conservative or adventurous), and to attach different importance to this factor. The distance between the leading vehicle and the subject vehicle reflects the comfort level in terms of the driving space, which affects the driver's satisfaction on the current lane. In addition, the existence of a bus in front of the vehicle has a large influence on the utility of the current lane. The

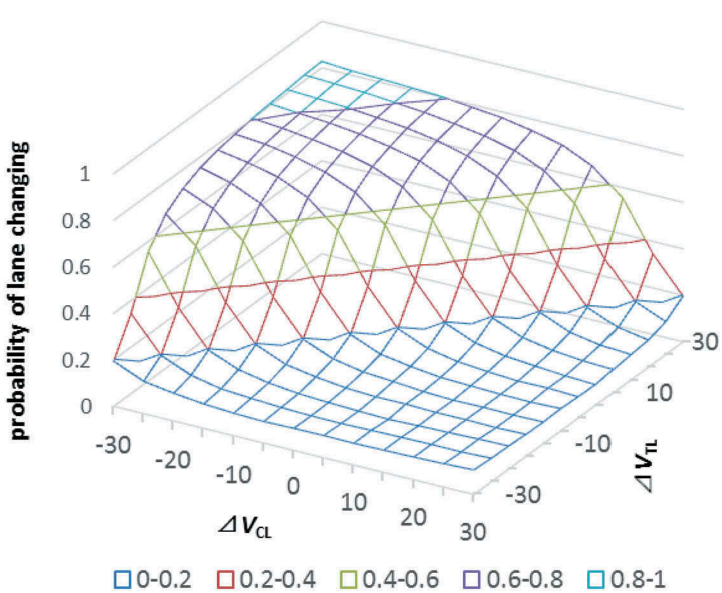

Figure 5. Lane-changing probability for different $\varDelta V_{CL}$ and $\varDelta V_{TL}$.

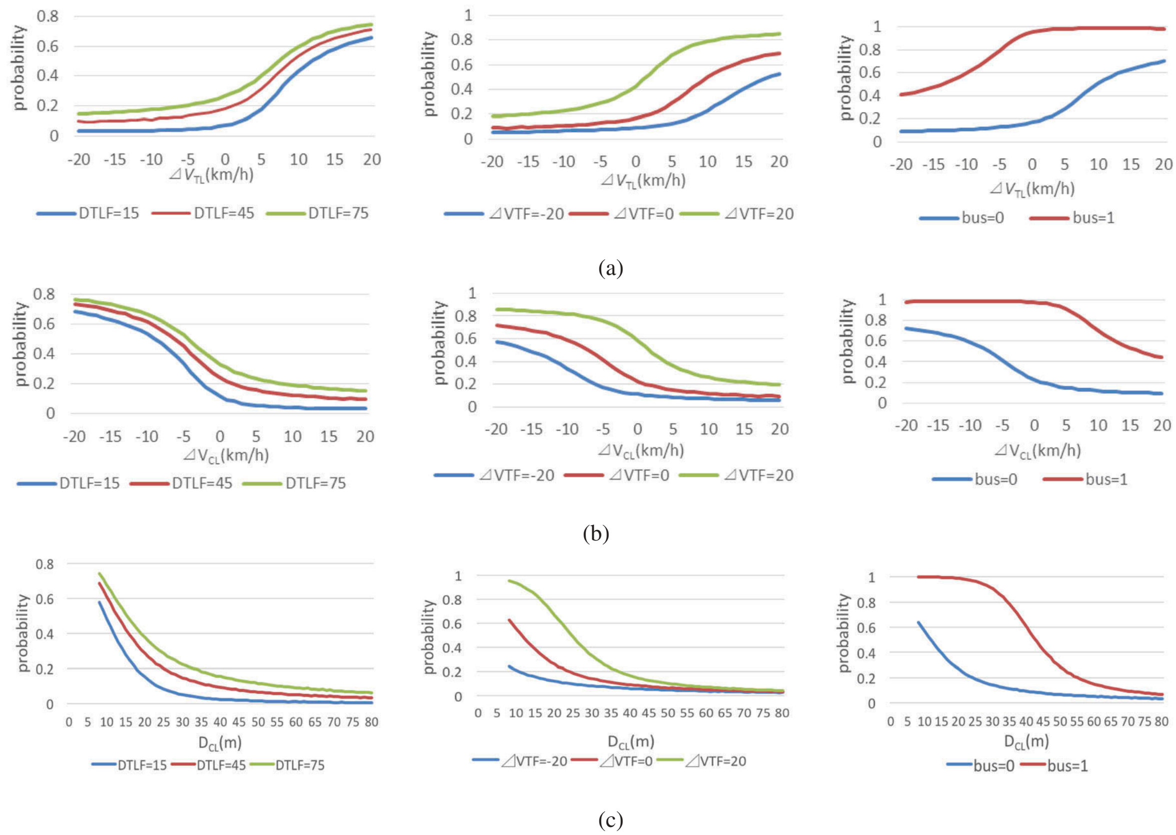

Figure 6. Lane-changing probabilities vs. (a) $\varDelta V_{\mathrm{TL}}$, (b) $\varDelta V_{\mathrm{CL}}$, (c) D_{CL}, with different values of D_{TLF}, $\varDelta V_{\mathrm{TF}}$ and *BUS*.

validation of the model has yielded 73.2% correct prediction for lane-changing samples and 82.0% for non-lane-changing samples in the validation dataset. The prediction accuracy is acceptable, which demonstrates the transferability of the model results to new data.

Sensitivity analyses were introduced to quantify the degree of influence of different factors on the probability of a lane change. Although all factors in the model are statistically significant, their degrees of influence are different. $\varDelta V_{\mathrm{TL}}$, $\varDelta V_{\mathrm{CL}}$ and D_{CL} have strong influences because the motivation of changing lanes is to seek better driving conditions in terms of speed and space. D_{TLF} or $\varDelta V_{\mathrm{TF}}$ are related to the safety of the lane changing, which have strong influences within certain ranges of $\varDelta V_{\mathrm{TL}}$, $\varDelta V_{\mathrm{CL}}$, and D_{CL}. It is not surprising to find that the presence of a bus is a strong factor in the decision to change lanes.

The results of this study help to better understand the effects of influential factors and the drivers' heterogeneity with regard to lane-changing decisions, which in turn provide a strong reference for improving the accuracy of microscopic traffic simulation models. Heterogeneity is specifically evident in the pursuit of speed and the gap acceptance. This finding can be further applied in simulation to assess the safety effects of lane changes and examine the risk level of individual drivers. Another application of the finding is that when developing safety countermeasures. Highly adventurous drivers can reduce the safety risks of lane changing through targeted and effective safety education. Driving assistance systems should be able to assist with the gap acceptance and prevent risky lane changes.

Further research can incorporate driver characteristics into the lane-changing model, and examine the impact of a lane change on the surrounding traffic, which matters to the improvement of traffic operations and safety. The relative position of the subject vehicle with respect to the leading and following vehicles in the target lane as well as the reactions of them (e.g. the following vehicle may be forced to brake) are not captured in this study. Incorporating them into the model along with the acceleration/deceleration of the subject vehicle during the lane changes may help to understand the process more deeply. Another limitation of the study is that only single-lane changes were included. Multiple lane changes may have an even greater impact on traffic flow and safety, and should thus be studied further. Lastly, a larger sample size would improve the accuracy of the lane-changing model.

Acknowledgments

The research reported in this paper was supported by the National Natural Science Foundation of China [Grant Number: 51578111 and 61502069], and the Fundamental Research Funds for the Central Universities [Grant Number: DUT18JC39]

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Foundation of China [61502069];Fundamental Research Funds for the Central Universities [DUT18JC39];National Natural Science Foundation of China [51578111].

ORCID

Xiao Qin http://orcid.org/0000-0003-0073-3485
Ronghan Yao http://orcid.org/0000-0002-6614-1960

References

Ahmed, K. I., M. E. Ben-Akiva, H. N. Koutsopoulos, and R. G. Mishalani. 1996. "Models of Freeway Lane Changing and Gap Acceptance Behavior." Paper presented at the 13th International Symposium on the Theory of Traffic Flow and Transportation, Lyon, France

Ahn, S., and M. J. Cassidy. 2007. "Freeway Traffic Oscillation and Vehicle Lane Changing Maneuvers." Paper presented at the 17th International Symposium on Transportation and Traffic Theory, London, England

Choudhury, C. F., M. E. Ben-Akiva, T. Toledo, G. Lee, and A. Rao. 2007. "Modeling Cooperative Lane-Changing and Forced Merging Behavior." Paper presented at the 86th Transportation Research Board Annual Meeting, Washington, D.C.

Deutsch, J. C., C. R. Santhosh-Kumar, M. Rickert, K. Nagel, M. Schreckenberg, and A. Latour. 1995. "Two Lane Traffic Simulations Using Cellular Automata." *Physica A: Statistical Mechanics and Its Applications* 231 (4): 534–550.

Gipps, P. G. 1986. "A Model for the Structure of Lane Changing Decisions." *Transportation Research Part B: Methodological* 20 (5): 403–414. doi:10.1016/0191-2615(86)90012-3.

Hidas, P. 2002. "Modelling Lane Changing and Merging in Microscopic Traffic Simulation." *Transportation Research Part C: Emerging Technologies* 10 (5): 351–371. doi:10.1016/S0968-090X(02)00026-8.

Hidas, P. 2005. "Modeling Vehicle Interactions in Microscopic Simulation of Merging and Weaving." *Transportation Research Part C: Emerging Technologies* 13 (1): 37–62. doi:10.1016/j.trc.2004.12.003.

Kita, H. 1999. "A Merging-Give-Way Interaction Model of Cars in A Merging Section: A Game Theoretic Analysis." *Transportation Research Part A: Policy & Practice* 33 (3–4): 305–312.

Laval, J. A., and L. Leclercq. 2008. "Microscopic Modeling of the Relaxation Phenomenon Using a Macroscopic Lane Changing Model." *Transportation Research Part B: Methodological* 42 (6): 511–522. doi:10.1016/j.trb.2007.10.004.

Moridpour, S., M. Sarvi, and G. Rose. 2010. "Lane Changing Models: A Critical Review." *Transportation Letters: the International Journal of Transportation Research* 2 (3): 157–173. doi:10.3328/TL.2010.02.03.157-173.

Rahman, M., M. Chowdhury, Y. Xie, and Y. He. 2013. "Review of Microscopic Lane-Changing Models and Future Research Opportunities." *IEEE Transactions on Intelligent Transportation Systems* 14 (4): 1942–1956. doi:10.1109/TITS.2013.2272074.

Sun, J., J. Zhang, and H. M. Zhang. 2014. "Investigation Of The Early-onset Breakdown Phenomenon at Urban Expressway Bottlenecks in Shanghai." *Transportmetrica B: Transport Dynamics* 2 (3): 215-228. doi: 10.1080/21680566.2014.932262.

Sun, J., and L. Elefteriadou. 2010. "Lane-Changing Behavior on Urban Streets: A Focus Group-Based Study. ." *Applied Ergonomics* 42 (5): 682–691. doi:10.1016/j.apergo.2010.11.001.

Sun, J., and L. Elefteriadou. 2012. "Lane-Changing Behavior on Urban Streets: An "In-Vehicle" Field Experiment-Based Study." *Computer-Aided Civil and Infrastructure Engineering* 27 (7): 525–542. doi:10.1111/mice.2012.27.issue-7.

Suzuki, K., and H. Nakamura. 2006. "Development and Performance Evaluation of a Video Image Processing System for Traffic Flow analysis–Traffic Analyzer." *Civil Society Proceedings* 62: 276–287. in Japanese.

Toledo, T., H. N. Koutsopoulos, and M. E. Ben-Akiva. 2003. "Modeling Integrated Lane-Changing Behavior." *Transportation Research Record: Journal of the Transportation Research Board* 1857 (1): 30–38. doi:10.3141/1857-04.

Toledo, T., H. N. Koutsopoulos, and M. E. Ben-Akiva. 2009. "Estimation of an Integrated Driving Behavior Model." *Transportation Research Part C: Emerging Technologies* 17 (4): 365–380. doi:10.1016/j.trc.2009.01.005.

Train, K. E. 2003. *Discrete Choice Methods with Simulation*. Berkeley: University of California, Berkeley and National Economic Research Associates.

Yang, Q., H. N. Koutsopoulos, and M. E. Ben-Akiva. 2000. "Simulation Laboratory for Evaluating Dynamic Traffic Management Systems." *Transportation Research Record: Journal of the Transportation Research Board* 1710 (1): 122–130. doi:10.3141/1710-14.

Zheng, J., K. Suzuki, and M. Fujita. 2014. "Predicting Driver's Lane-Changing Decisions Using a Neural Network Model." *Simulation Modelling Practice & Theory* 42 (1): 73–83. doi:10.1016/j.simpat.2013.12.007.

Zheng, Z. 2014. "Recent Developments and Research Needs in Modeling Lane Changing." *Transportation Research Part B: Methodological* 60 (1): 16–32. doi:10.1016/j.trb.2013.11.009.

Zheng, Z., S. Ahn, and C. M. Monsere. 2010. "Impact of Traffic Oscillations on Freeway Crash Occurrences." *Accident Analysis &Prevention* 42 (2): 626–636. doi:10.1016/j.aap.2009.10.009.