Measurement 182 (2021) 109707

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Comprehensive survival analysis of lane-changing duration

Yang Li [a], Linbo Li [a,*], Daiheng Ni [b], Yue Zhang [a]

[a] Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, China, 4800 Cao'an Road, Shanghai 201804, China
[b] Civil and Environmental Engineering, University of Massachusetts Amherst, MA 01003, USA

ARTICLE INFO

Keywords:
Lane-changing duration
Lane-changing behavior
Survival analysis
HighD dataset

ABSTRACT

Lane-changing duration (LCD) measures the total time it takes for a vehicle to travel from the current lane to the target lane. However, there is still a paucity research on LCD up to now. In this paper, we present a comprehensive analysis of LCD from the perspective of survival analysis. Naturalistic vehicle trajectory HighD dataset is employed in this paper, which contains of 16.5 h of measurement and over 11,000 vehicles. Both comparative univariate and regression analysis has been conducted to research the characteristic of the whole survival function and the influencing factors of LCD. Results indicate that Generalized Gamma distribution has high degree of coincidence with the non-parametric method in estimating the survival function, and LoglogisticAFT model exhibits more credible results than other regression models. Furthermore, the results and modeling implications have been discussed. We hope this paper could contribute to our further understanding of LCD and LC behaviors.

1. Introduction

Along with car-following (CF) behavior, lane-changing (LC) behavior is also an indispensable component of traffic flow theories [1]. LC behavior describes the lateral movement of the vehicles from the current lane to the target lane while proceeding forward. Over the past decades, a considerable amount of works has been made to research the LC behaviors. Generally speaking, the research theme of LC can be roughly classified into: the decision-making process of LC [2,3], the implementation of LC [4–9], and the impacts of LC on the surroundings [1].

As the schematic diagram of LC shown in Fig. 1, between the starting and ending point, the lateral speed of the subject vehicle gradually increases from zero, and then gradually decrease to zero. This process could be regarded as the execution of LC, or could be viewed as the implementation of LC. This process may involve the research of LC trajectory planning and tracking [4,6,10], LC trajectory predicting [11] and LC duration [12–15]. This paper focuses on researching the LC duration. For the convenience of subsequent research, we denote LC duration as LCD. LCD measures the total time spent of the subject vehicle in the execution of LC (from the starting point to the ending point). Up to now, tremendous studies demonstrated that LCD roughly ranges from 1 s to 16 s [12–19]. At the same time, results show that various factors might affect LCD, like traffic density [12,14], vehicles types [12,15,17,20], driver characteristics [13,15], the direction of LC [12,14,21], different time periods [20], road types [22], the interactions with surrounding vehicles [12,13,15,19].

Toledo and Zohar [12] made the first attempt to model LCD using multiple linear regression model. The dataset is the well-known NGSIM dataset, which contains a set of trajectory data at a fine time resolution. The results indicate that traffic density, by the direction of the change, and by other vehicles around the subject vehicle may influence the LCD. Different from this method, Wu, Zhang, Singh and Qin [20] employed the semi-parametric proportional hazard-based model to analyze the mandatory LCD. The data is collected from an unmanned aerial vehicle in a freeway maintenance construction area. Results demonstrate that there is no significant evidence showing that different vehicle types have an effect on LCD, but there is a significant difference in LCD during different time periods. Thereafter, Vlahogianni [13] modeling the overtaking duration in two-lane highways using the parametric hazard-based duration. Three parametric survival regression methods (Weibull, Loglogistic and Lognormal distributions) have been introduced to model the hazard function of the duration time. Results demonstrate that the Loglogistic exhibits better performance than other models. Meanwhile, the speed difference relative to the lead vehicle, the speed of opposing traffic, the spacing from the lead and opposing traffic, and the driver's gender influence the LCD. In the same year, Cao, Young and Sarvi [19] analyzed the LCD using the data collected from the video camera mounted on a high building adjacent to the road. Result reveals that the

* Corresponding author.
E-mail addresses: cc960719@tongji.edu.cn (Y. Li), llinbo@tongji.edu.cn (L. Li), ni@engin.umass.edu (D. Ni), zhangyue18@tongji.edu.cn (Y. Zhang).

https://doi.org/10.1016/j.measurement.2021.109707
Received 4 January 2021; Received in revised form 20 April 2021; Accepted 2 June 2021
Available online 10 June 2021
0263-2241/

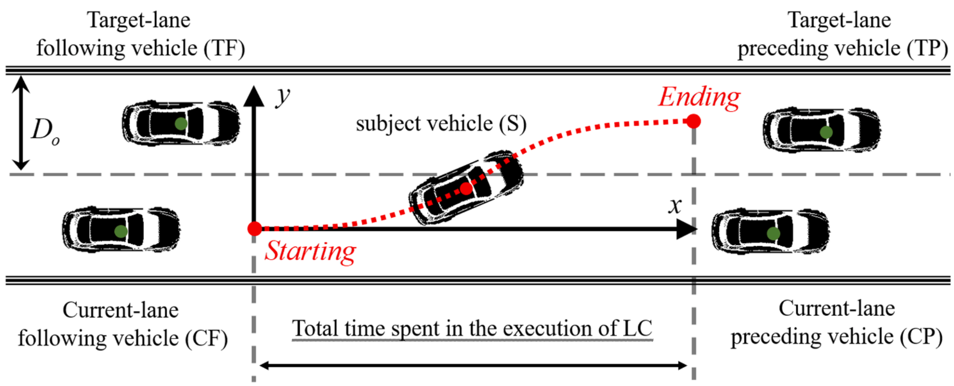


Fig. 1. The schematic diagram of LC behavior (dot curve represents the LC trajectory from the starting point to the ending point).

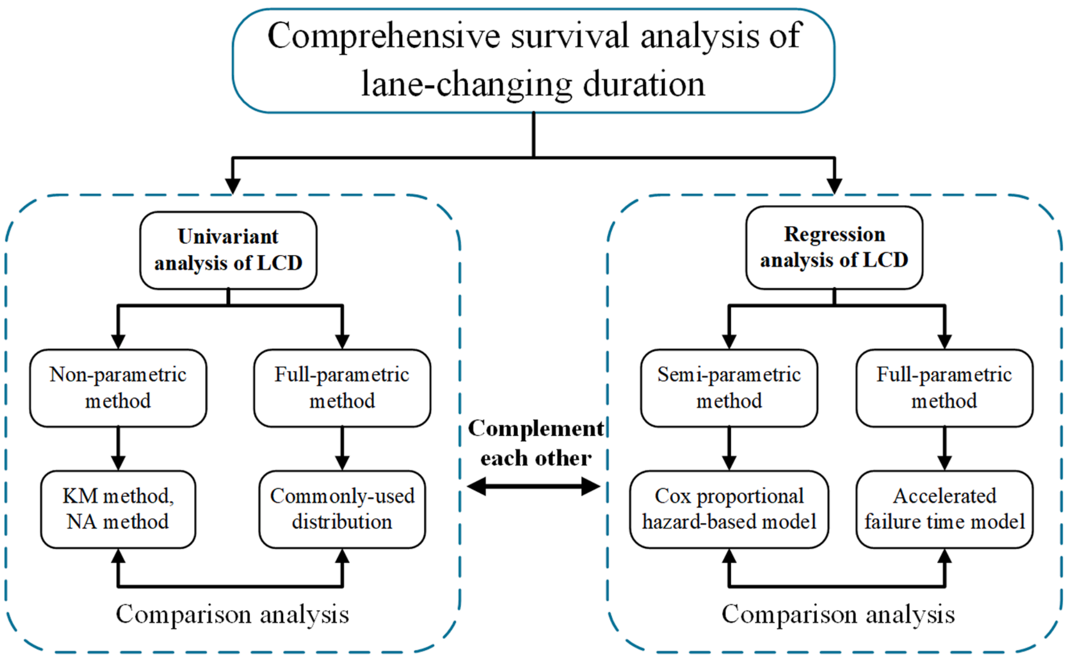
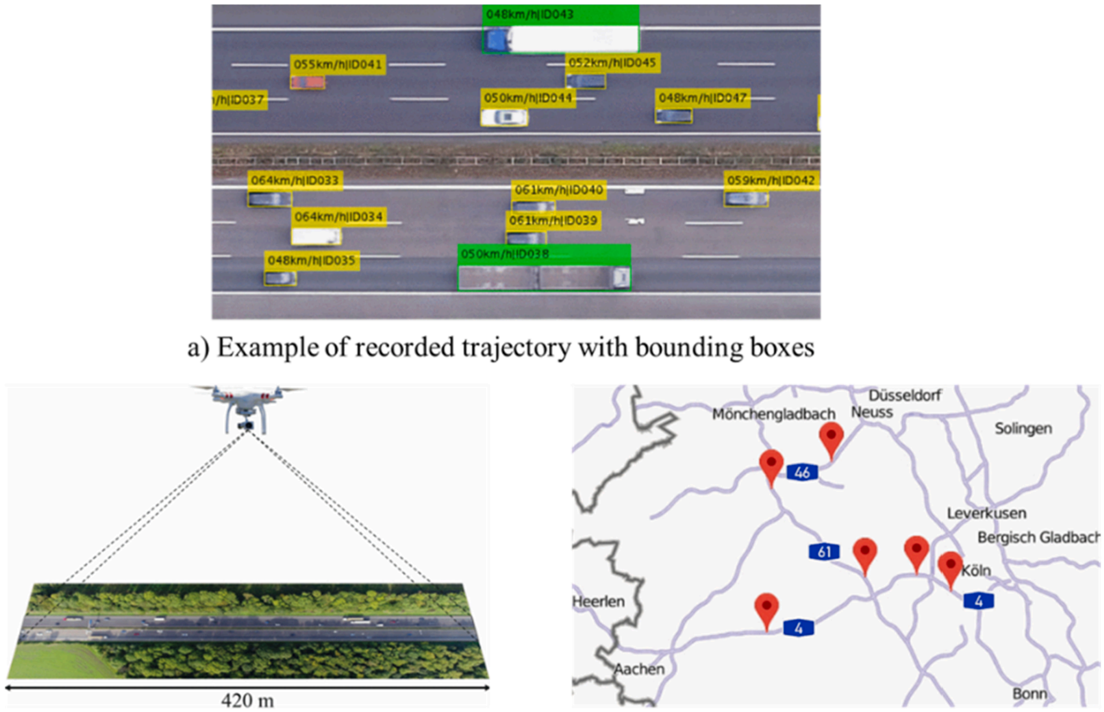


Fig. 2. The established survival analysis framework of LCD.

traffic conflict during a LC event also influences LCD. Furthermore, Wang, Li and Li [14] found that there is no significance difference between the left-to-right LCD, whose result is quite different from the result in Toledo and Zohar [12]. They also conjectured that the duration times will reach a saturation value when the velocity becomes even higher. Recently, Yang, Wang and Quddus [15] developed a three-level mixed-effects linear regression model to explore the variables affecting LCD, and the research results are consistent with results in Toledo and Zohar [12].

Unlike previous research, this paper attempts to present a comprehensive analysis of LCD from the perspective of survival analysis. The reason why we choose this perspective is due to its merit and popularity in mining the information behind the traffic data [23]. Over the past decades, this method has been widely used in transportation data analysis, including but not limited to the reliability travel time analysis [24,25], overtaking duration analysis [26], the duration between adjacent trips [27–29], time incident duration analysis [30,31] and so on. Although existing research have achieved certain progress, there is still a paucity research on LCD so far [12,15], let alone using the survival analysis methods, since most efforts in LC have been devoted to modeling the decision-making process and the impact of LC [1]. At the same time, existing research mainly focuses on the regression analysis part, while ignoring the analysis of the overall distribution characteristics of LCD. Furthermore, a systematic comparative univariate and regression analysis of LCD is still lacking. What the differences between different types of survival models in analyzing LCD are, and whether these differences would affect our comprehending of the mechanism of LCD remain to be explored. These issues would inevitably hinder us from having a deeper understanding of LC behavior, thus limiting our exploration of microscopic traffic flow characteristics.

To address these needs, this paper takes a further step to investigate the LCD. The objective of this paper is threefold: (a) establishment of a comprehensive survival analysis framework of LCD; (b) study of the characteristic of the survival function of LCD through the comparative univariate analysis; (c) investigation of the influencing factors of LCD through the comparative regression analysis. The main contribution of this study is that a complete analysis of LCD from the perspective of survival analysis for the first time has been conducted for the first time. In the univariate part, non-parametric and parametric methods are employed, and in the regression part, semi-parametric and parametric methods are introduced. It is worth noting that the differences between the above-mentioned methods are further explored. Finally, we discussed the obtained results and the modeling implications in detail. Such kind of synthesis analysis not only helps us understand LC behavior or



a) Example of recorded trajectory with bounding boxes

b) The bird' s eye view on the highway from a drone

c) Six different highway recording locations in this dataset

Fig. 3. The brief introduction of the HighD dataset including the bounding boxes of each vehicle, the collecting method and the recording locations.

Table 1
Data format of the Tracks file.

Name	Description
frame	The current frame.
id	The Track's id.
x	The × position of the upper left corner of the vehicle's bounding box.
y	The y position of the upper left corner of the vehicle's bounding box.
width	The width of the bounding box of the vehicle.
height	The height of the bounding box of the vehicle.
xVelocity	The longitudinal velocity in the image coordinate system.
yVelocity	The lateral velocity in the image coordinate system.
xAcceleration	The longitudinal acceleration in the image coordinate system.
yAcceleration	The lateral acceleration in the image coordinate system.
frontSightDistance	The distance to the end of the recorded highway section in driving direction from the vehicle's center.
backSightDistance	The distance to the end of the recorded highway section in opposite direction from the vehicle's center.
dhw	The distance headway.
thw	The time headway.
ttc	The time-to-collision.
precedingXVelocity	The longitudinal velocity of the preceding in the image coordinate system.
precedingId	The id of the preceding vehicle in the same lane.
followingId	The id of the following vehicle in the same lane.
leftPrecedingId	The id of the preceding lane on the left in the direction of travel.
leftAlongsideId	The id of the adjacent lane on the left in the direction of travel.
leftFollowingId	The id of the following lane on the left in the direction of travel.
RightPrecedingId	The id of the preceding lane on the right in the direction of travel.
RightAlongsideId	The id of the adjacent lane on the right in the direction of travel.
RightFollowingId	The id of the following lane on the right in the direction of travel.
laneId	The IDs start at 1 and are assigned in ascending order.

microscopic traffic flow characteristics more deeply, but also helps to provide data analysis support for the design and development of the ADAS systems. At the same time, the HighD dataset is utilized as the measurement data in this paper, which has strengths in terms of naturalistic driving behavior, static and dynamic scenario description as well as data privacy protection [32]. The remainder of this paper is organized as follows. Section 2 presents the introduction of survival analysis and the analysis framework employed in this paper. Section 3 presents the description and processing of the HighD dataset, and preliminary analysis of LCD. Section 4 presents the comprehensive survival analysis of LCD, including the univariate analysis and regression analysis part. Section 5 presents the discussion. Finally, the conclusion is presented in Section 6.

2. The survival analysis framework of LCD

2.1. The illustration of the analysis framework

In general, survival analysis is a collection of statistical procedures of data analysis for which the outcome variable of interest is time until an event occurs [33]. The main goal of survival analysis is to estimate the survival and hazard function of the duration data, and to assess the relationship of explanatory variables to duration time [33]. The methods in survival analysis could be divided into: non-parametric, parametric, and semi-parametric [23]. Non-parametric methods make no assumptions about the statistical distribution of the duration data, but using this method may result in lower interpretability of the influence of the covariates on the duration data. Higher interpretability could be realized if adopting the parametric methods, which assumes the specific distribution in advance. However, this method might not be robust enough if the duration data does no obey the distribution. In order to achieve the balance between the interpretability of the influence of the covariates and the reduction of the model hypothesis, the semi-parametric method is proposed [23]. The explanatory properties and high interpretability have been both considered in this model. In

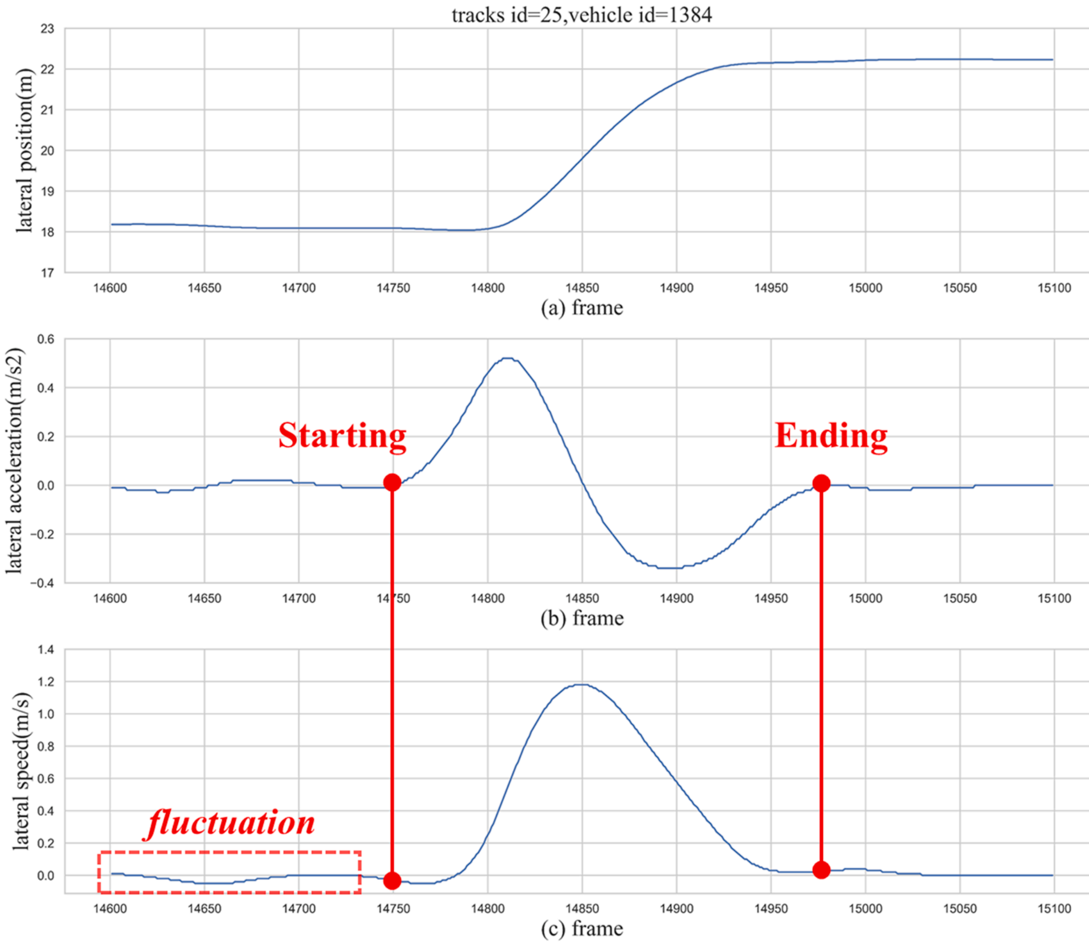


Fig. 4. Example of determining the starting and ending point of the LC trajectory.

fact, the models in survival analysis are not limited to these three types. Due to the page limit, what we mentioned here are the most representative three types of models.

Fig. 2 shows the proposed analysis framework, including the univariate analysis and regression analysis part. The univariate analysis part is mainly composed of two parts: (a) investigation of the charateristic of the whole survival distribution of LCD; (b) comparative analysis between the non-parametric and the parametric methods. The KM (Kaplan–Meier) [34], NA (Nelson Aalen) estimator, and five commonly-used distributions are employed in this part. The regression analysis part is mainly composed of two parts: (a) exploration of the influencing factors of the LCD; (b) comparative analysis between the semi-parametric and the parametric methods. The CPH (Cox Proportional Hazard-based) model and the AFT (Accelerated Failure Time) models are employed in this part. The contents of these two parts complement each other and improve each other. For instance, while studying the differences between different regression models, we may utilize certain methods and ideas in univariate analysis (such as the work in Section 4.2). It is worth noting that due to page limit, it is impossible to exhaust all models in this paper, so the models employed in each part are the most classic and most commonly-used.

2.2. Introduction of survival analysis

Survival analysis begins with the probability density function $f(t)$, which is rate of death events per unit time. $F(t)$ is the cumulative density function of $f(t)$. The survival function $S(t)$ is the probability that the time of death is later than some specified time, which is a non-increasing function of t. The relationship between the $S(t)$, $F(t)$ and $f(t)$ is given below.

$$S(t) = P(T > t) = 1 - F(t) = 1 - \int_0^t f(u)du = \int_t^\infty f(u)du \tag{1}$$

Where T represents the lifetime (LCD). The conditional probability of an event occurring at time t given that the event has not occurred until time t is given below:

$$\lim_{\delta t \to 0} P(t \leqslant T \leqslant t + \delta t | T > t) \tag{2}$$

The formula result goes to 0 as δt approaches 0. Deriving from the above formula by the interval δt, we obtain the hazard function $h(t)$ at time t as follows:

$$h(t) = \lim_{\delta t \to 0} \frac{P(t \leqslant T \leqslant t + \delta t | T > t)}{\delta t} \tag{3}$$

$$h(t) = \frac{1}{P(T > t)} \lim_{\delta t \to 0} \frac{P(t \leqslant T \leqslant t + \delta t)}{\delta t} = \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)} \tag{4}$$

The above formulas constitute the basic concepts in survival analysis. It is worth noting that as long as we know one of the above formulas, the other formulas could be derived consequently.

2.3. Univariant survival model

KM [34] and NA estimator method are the most widely used non-parametric univariant model to estimate the survival and hazard function. The formulas are given below:

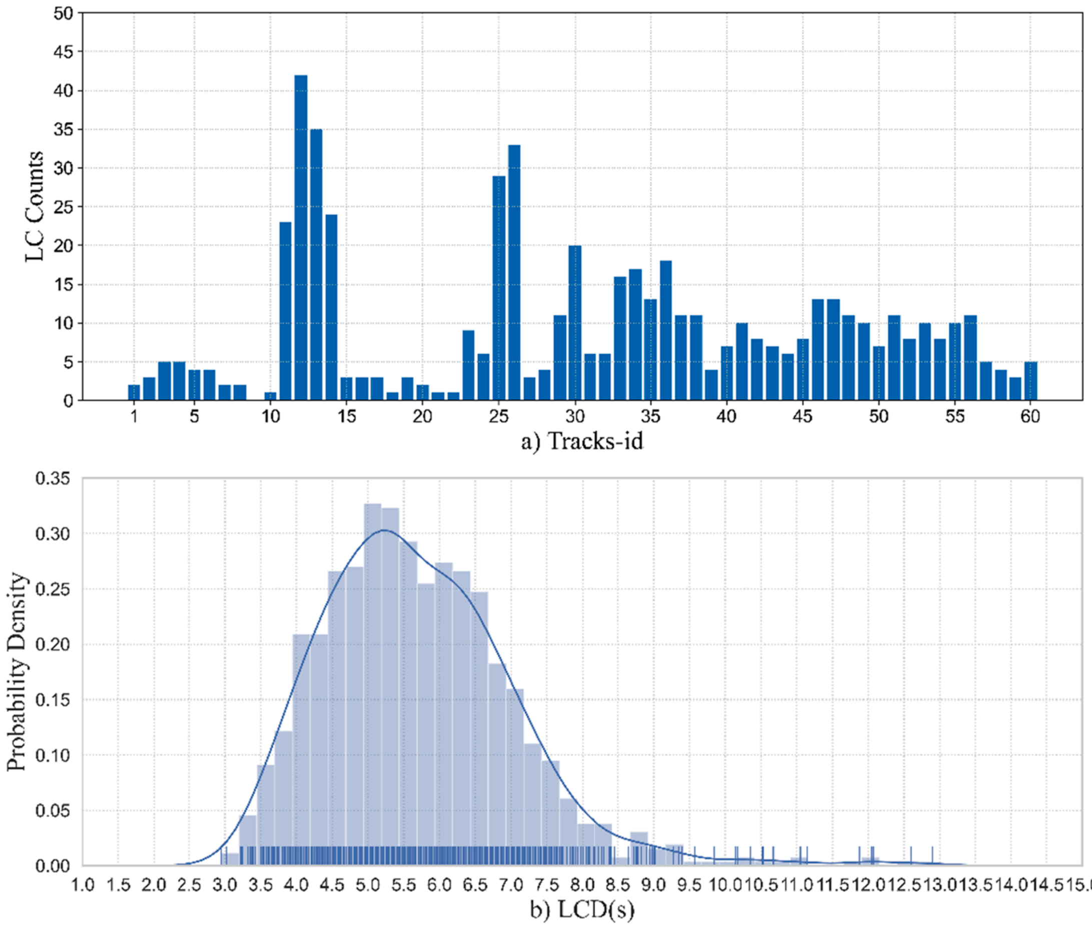


Fig. 5. LC counts for each tracks file and the distribution of the LCD.

$$\widehat{S}(t) = \prod_{t_i<t}\frac{n_i - d_i}{n_i} \tag{5}$$

$$\widehat{H}(t) = \sum_{t_i\leqslant t}\frac{d_i}{n_i} \tag{6}$$

Where d_i are the number of death events at time t, n_i is the number of subjects at risk of death just prior to time t.

As for the parametric model, five commonly-used survival distribution functions are selected: Exponential function, Weibull function, Lognormal function, Loglogistic function and Generalized-Gamma function. Let T denotes the continuous non-negative random variable representing survival time, and we could transform T into the following form:

$$Y = \log T = \alpha + \sigma W \tag{7}$$

The exponential distribution has the constant hazard value, which means the conditional probability of an event is constant over time.

$$h(t) = \lambda \tag{8}$$

Where λ is the constant hazard. According to Eq. (4), we could derive $S(t) = exp(-\lambda t), f(t) = \lambda exp(-\lambda t), E(T) = 1/\lambda, Var(T) = 1/\lambda^2$

The Weibull distribution assumes the hazard obeys the Weibull distribution, which is capable of allowing for positive, negative, or even no duration dependence [23].

$$h(t) = p\lambda^p t^{p-1} \tag{9}$$

Where p and λ are the parameters that controls the shape of $h(t)$. The hazard is rising if $p > 1$, constant if $p = 1$, and declining if $p < 1$. We could derive $S(t) = \lambda exp(-(\lambda t)^p), f(t) = p\lambda^p t^{p-1} exp(-(\lambda t)^p)$.

The Lognormal distribution assumes the W has a standard normal distribution.

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma t} exp\left(-\frac{1}{2}\left(\frac{logt - \mu}{\sigma}\right)^2\right) \tag{10}$$

The $h(t)$ increases from 0 to reach a maximum and then decrease monotonically, approaching 0 as $t\rightarrow\infty$. $F(t) = \frac{1}{2} + \frac{1}{2}exp\left(\frac{logt-\mu}{\sigma\sqrt{2}}\right)$, $S(t) = \frac{1}{2} - \frac{1}{2}exp\left(\frac{logt-\mu}{\sigma\sqrt{2}}\right)$.

The Loglogistic distribution assumes the W has a standard logistic distribution.

$$f_W(w) = \frac{e^w}{(1+e^w)^2} \tag{11}$$

Where we could rewrite, $\alpha = -log\lambda, p = 1/\alpha$, Taking the logs we obtain the integrated hazard, and we find the survival and hazard function.

$$S(t) = \frac{1}{1+(\lambda t)^p} \tag{12}$$

$$h(t) = \frac{\lambda p(\lambda t)^{p-1}}{1+(\lambda t)^p} \tag{13}$$

Where the hazard itself monotone decreasing from ∞ if $p < 1$, monotone decreasing from λ if $p = 1$, and similar to lognormal if $p > 1$.

The Generalized-Gamma distribution assumes the W has generalized

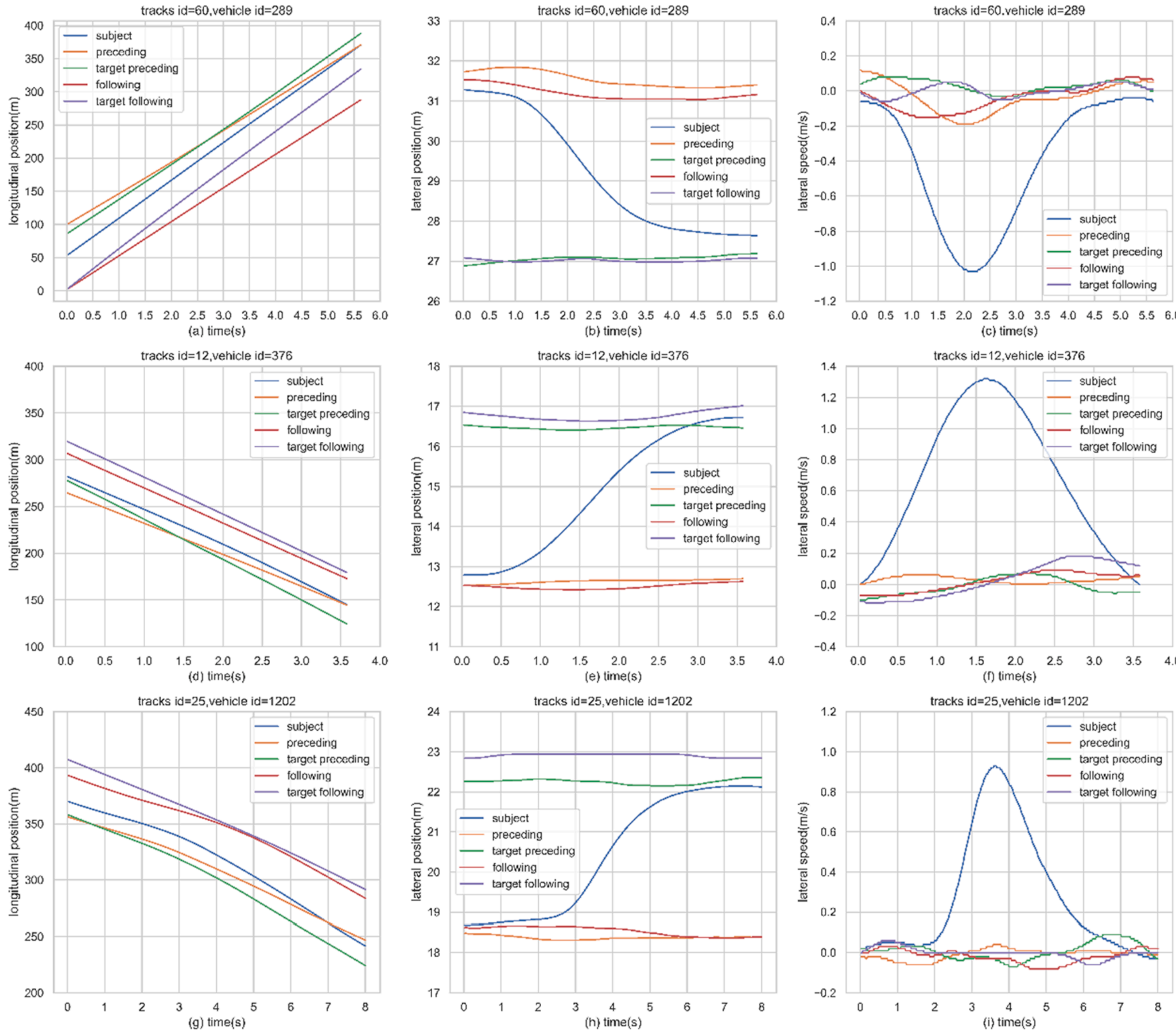


Fig. 6. Three examples of LC trajectories (longitudinal position, lateral position and the lateral acceleration).

extreme value distribution with parameter k. The density of the Generalized Gamma distribution is formulated as:

$$f(t) = \frac{\lambda p (\lambda t)^{pk-1} e^{-(\lambda t)^p}}{\Gamma(k)} \tag{14}$$

Where $p = 1/\sigma$. The Generalized Gamma includes the following interesting special cases: gamma when $p = 1$, Weibull when $k = 1$, Exponential when $p = k = 1$, Lognormal when $k \rightarrow \infty$.

2.4. Regression survival model

The CPH model consists of the baseline hazard (non-parametric) and the partial hazard (parametric). The baseline hazard part indicates the pattern of time dependence that is assumed to be common to all units, and the partial hazard part scales the baseline hazard function common to all units up or down.

$$h(t|x_i) = \overbrace{\underbrace{h_0(t)}}^{baseline\ hazard} \underbrace{exp(x_i^T \beta)}_{partial\ hazard} = h_0(t)\lambda \tag{15}$$

Where $h_0(t)$ represents the baseline hazard, λ represents the partial hazard which is non-negative, x represents the covariates, β represents the regression coefficients.

The form of the AFT model is expressed as below. Unlike the CPH model, the AFT model directly models the survival time of the duration data. The AFT model assumes the linear relationship between the survival time and the covariates. The AFT model can be expressed as:

$$Y_i = x_i^T \beta + W_i \tag{16}$$

Where $Y = logT$, W_i represents the independent residuals, which is an error term with different density function. Then, we could rewrite the above Equation as:

$$T_i = T_0 exp(x_i^T \beta) \tag{17}$$

Where $T_0 = exp(W_i)$. When the $j-th$ dimension changes Δ_j, the survival time would change $exp(\Delta_j \beta_j)$.

The assumption of the distribution about W_i determine which sort of AFT model describes the distribution of the survival time T. In this paper, we introduce three most commonly-used AFT models in this paper: WeibullAFT model, LognoarmalAFT model and LoglogisticAFT

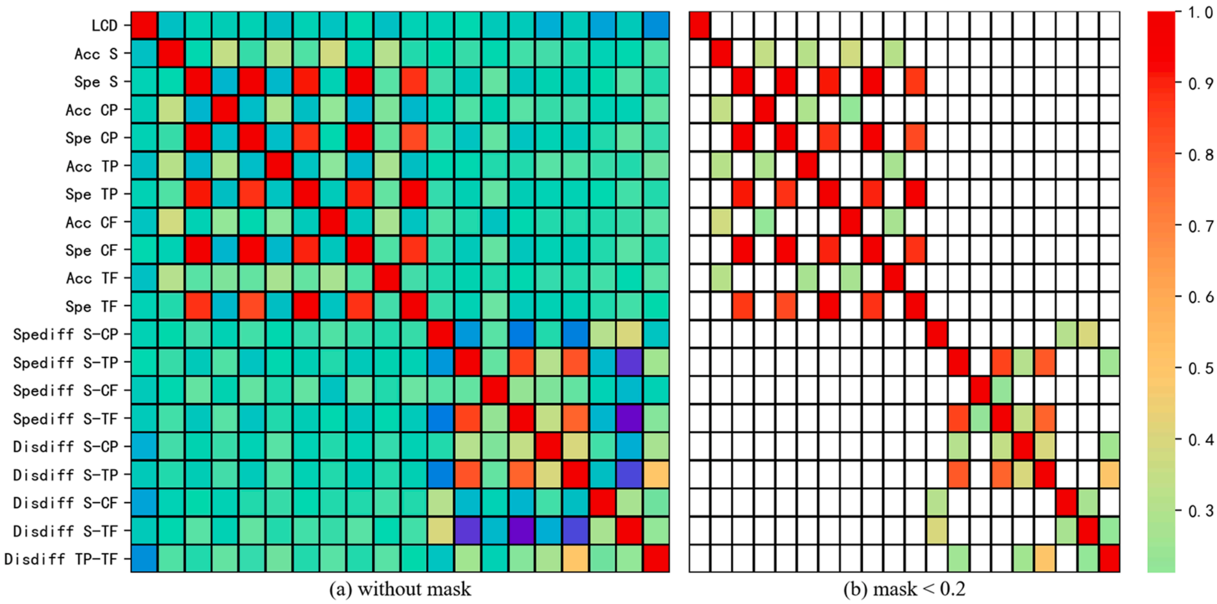


Fig. 7. The heatmap of the Spearman Coefficient of the variables.

Table 2
Variables terms for each LC trajectory at the initial state of LC.

Vehicle	Variable	Definition
Vehicle S	*LCD*	Lane-changing duration (s)
	Status	Whether vehicle complete LC (binary variable, 0 or 1)
	$\dot{x}_S$	The speed of the subject vehicle
	$\ddot{x}_S$	The acceleration of vehicle
Vehicle CP	$\ddot{x}_{CP}$	The acceleration of vehicle CP
	$\triangle x_{S-CP}$	Distance between vehicle S and vehicle CP
	$\triangle \dot{x}_{S-CP}$	Speed difference between vehicle S and vehicle CP
Vehicle CF	$\ddot{x}_{CF}$	The acceleration of vehicle CF
	$\triangle x_{S-CF}$	Distance between vehicle S and vehicle CF
	$\triangle \dot{x}_{S-CF}$	Speed difference between vehicle S and vehicle CF
Vehicle TP	$\ddot{x}_{TP}$	The acceleration of vehicle TP
	$\triangle x_{S-TP}$	Distance between vehicle S and vehicle TP
	$\triangle \dot{x}_{S-TP}$	Speed difference between vehicle S and vehicle TP
Vehicle TF	$\ddot{x}_{TF}$	The acceleration of vehicle TF
	$\triangle x_{S-TF}$	Distance between vehicle S and vehicle TF
	$\triangle \dot{x}_{S-TF}$	Speed difference between vehicle S and vehicle TF
	$\triangle x_{TP-TF}$	Gap distance on the target lane (distance between TP and TF)

model, and the corresponding error term distributions of W_i are extreme value distribution, normal distribution and the logistic distribution.

3. Lane-changing trajectory extraction

The HighD dataset is employed in this paper, which is a new dataset of naturalistic vehicle trajectories recorded on German highways during 2017 and 2018 [32]. This dataset contains of 16.5 h of measurement, 45,000 km of total driven distance and over 11,000 vehicles. These trajectories are recorded in 4 k (4096*2160) resolution from six different locations near Cologne, Germany (as shown in Fig. 3). The total driven distance of this dataset is about 45,000 km and the total recorded hour is about 16.5. At the same time, the positioning error of each trajectory is typically less than ten centimeters. Compared with NGSIM dataset [35], this dataset is more suitable for a system-level validation of highly automated driving systems [32]. For more details, please refer to Krajewski, Bock, Kloeker and Eckstein [32]. The data format of the HighD dataset is given in Table 1.

As shown in Fig. 1, we mainly focus on extracting the LC trajectory which involves five surrounding vehicles: subject vehicle, current-lane preceding vehicle, current-lane following vehicle, target-lane preceding vehicle, target-lane following vehicle (For the convenience of the subsequent research, we denote these vehicles as S, CP, CF, TP and TF). The process of extracting the LC trajectories are summarized below:

Step 1: We mainly focus on researching the LC of the private vehicles. We manually filter out the LC trajectories in which some of the above five vehicles are missing.

Step 2: We determine the beginning and ending point of LC according to the lateral speed of the subject vehicle. It is reasonable for us assume that the lateral speed of the subject vehicle equals to zero at these two points.

Step 3: After the above two steps, we roughly got a preliminary data processing result. Finally, we manually verify the final LC trajectory one by one. This is because the trajectory information may be missing at some time in the process of LC.

Fig. 4 exhibits an example of the process of determining the starting and ending points of the LC trajectory. It is worth noting that when we determine the starting and ending points, we cannot rely solely on the variable speed, but also need to combine acceleration and position information. This is because the speed and acceleration of the vehicle have been vibrating around zero as shown in Fig. 4. Each specific trajectory needs specific analysis and determination. Finally, we have obtained a total number of 560 LC trajectories. The LC counts for each tracks file id (there are total 60 Excel files), and the distribution of LCD is presented in Fig. 5. Among these 60 tracks files, the 12th to 14th tracks and the 25th to 27th tracks have the most LC tracks. The overall LCD roughly concentrates in the 5 s ~ 6 s interval. The average LCD is about 5.2 s, and the standard deviation is about 1.32 s. Fig. 6 presents three examples of LC trajectories. Fig. 7 presents the heatmap of the Spearman Coefficient of the variables. There is a strong correlation between the speed of these five vehicles, and the Spearman Coefficient values all exceed 0.8. Therefore, we introduce the speed difference variables. For the convenience of subsequent research, the variables terms are defined in Table 2.

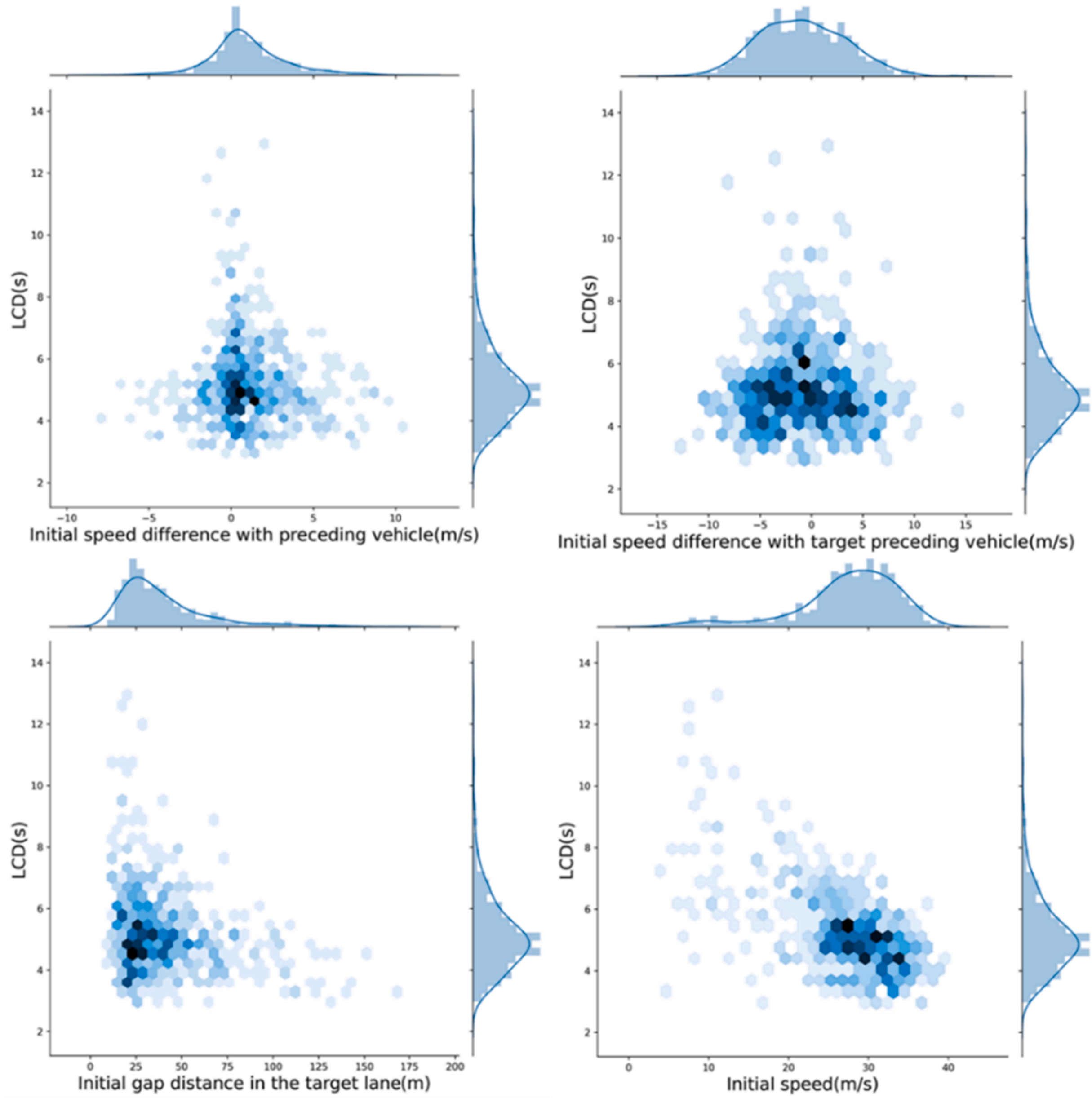


Fig. 8. Joint distribution between LCD with $\triangle\dot{x}_{S-CP}$, $\triangle\dot{x}_{S-TP}$, $\triangle x_{TP-TF}$, $\dot{x}_S$

Fig. 8 presents the joint distribution between the LCD with $\triangle\dot{x}_{S-CP}$, $\triangle\dot{x}_{S-TP}$, $\triangle x_{TP-TF}$ and $\dot{x}_S$. It can be found that the initial speed of the subject vehicle roughly concentrates at 25 m/s ~ 35 m/s, and the initial gap distance on the target lane is about 25 m ~ 50 m. Meanwhile, the initial speed difference with the target preceding vehicle roughly concentrates at −10 m/s ~ 10 m/s, the initial speed difference with the preceding vehicle roughly concentrates at −3m/s ~ 3 m/s. This indicates that the speed difference with the current-lane preceding vehicle is smaller than that of the target-lane preceding vehicle.

In order to investigate the difference of LCD under different LC behaviors, we divide the LC trajectories into two groups: vehicles moving from the fast lane to the slow lane and vehicles moving from the slow lane to the fast lane (as shown in Fig. 9). Fig. 10 exhibits the distribution of LCD for these two groups. It can be found that there was no significant difference in LCD between these two groups. The median value of LCD is lower than the average value, which indicates that half of the vehicles finish the LC within five seconds, and the remaining half vehicles pull up the average LCD of all the vehicles.

4. Survival results analysis

4.1. Total duration of LC

The estimation of the survival and cumulative hazard function of LCD is presented in Fig. 11. The subgraphs (a) and (b) in Fig. 11 are the results of non-parametric methods. It can be found that the survival function decreased rapidly in 3 s ~ 8 s, while decreased gently in 8 s ~ 12 s. This indicates that most vehicles complete LC within 3 s ~ 8 s. MST (Median Survival Time) is defined as the time where on average 50% of the duration has expired. Using the non-parametric method, the MST value is about 4.98 s. The corresponding lower confidence value is 4.9 s, and the upper confidence value is 5.05 s. This indicates that each vehicle has a 50% chance of completing its LC maneuver when LCD equals to 4.98 s. The subgraphs (c) and (d) in Fig. 11 are the results of five most commonly-used parametric methods. The corresponding MST, AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) values are given in Table 3. The AIC and BIC are both metrics of assessing model fit penalized for the number of estimated parameters.

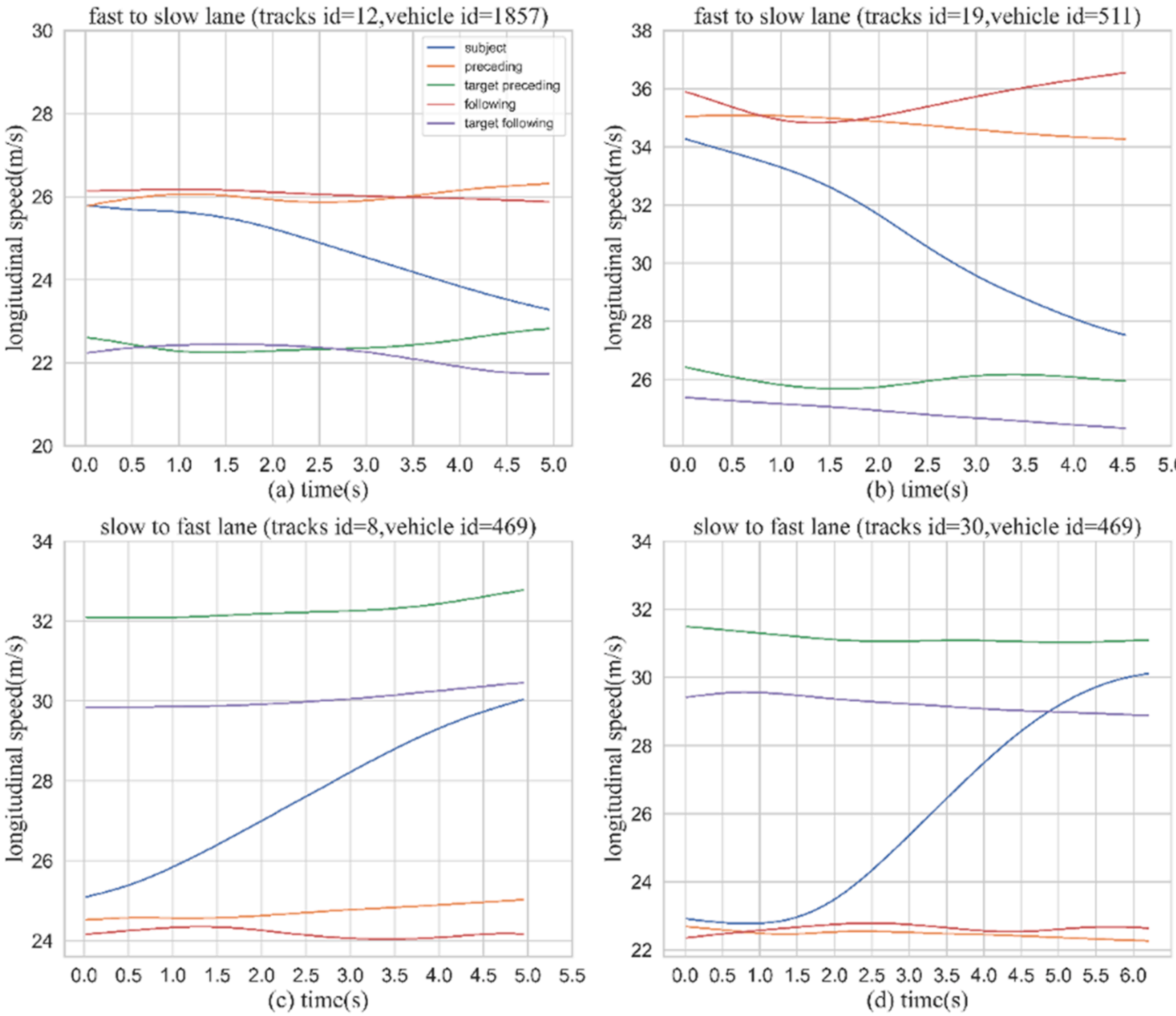


Fig. 9. The examples of vehicles moving from the fast lane to the slow lane and moving from the slow lane to the fast lane (the speed of the vehicles on the current are higher or lower than that of on the target lane)

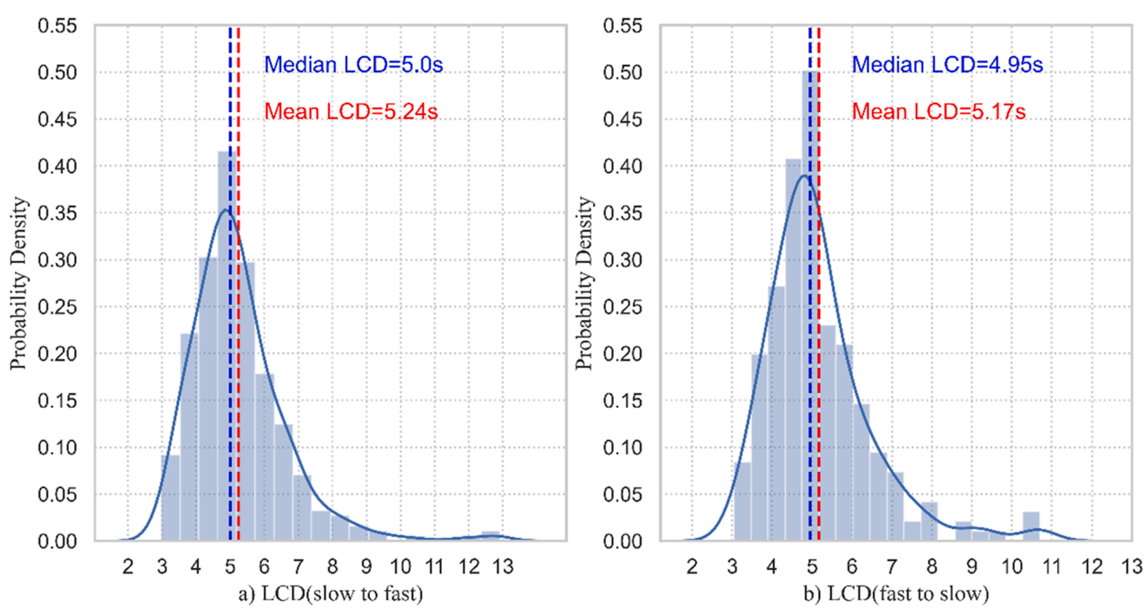


Fig. 10. The distribution of the LCD under different LC behaviors

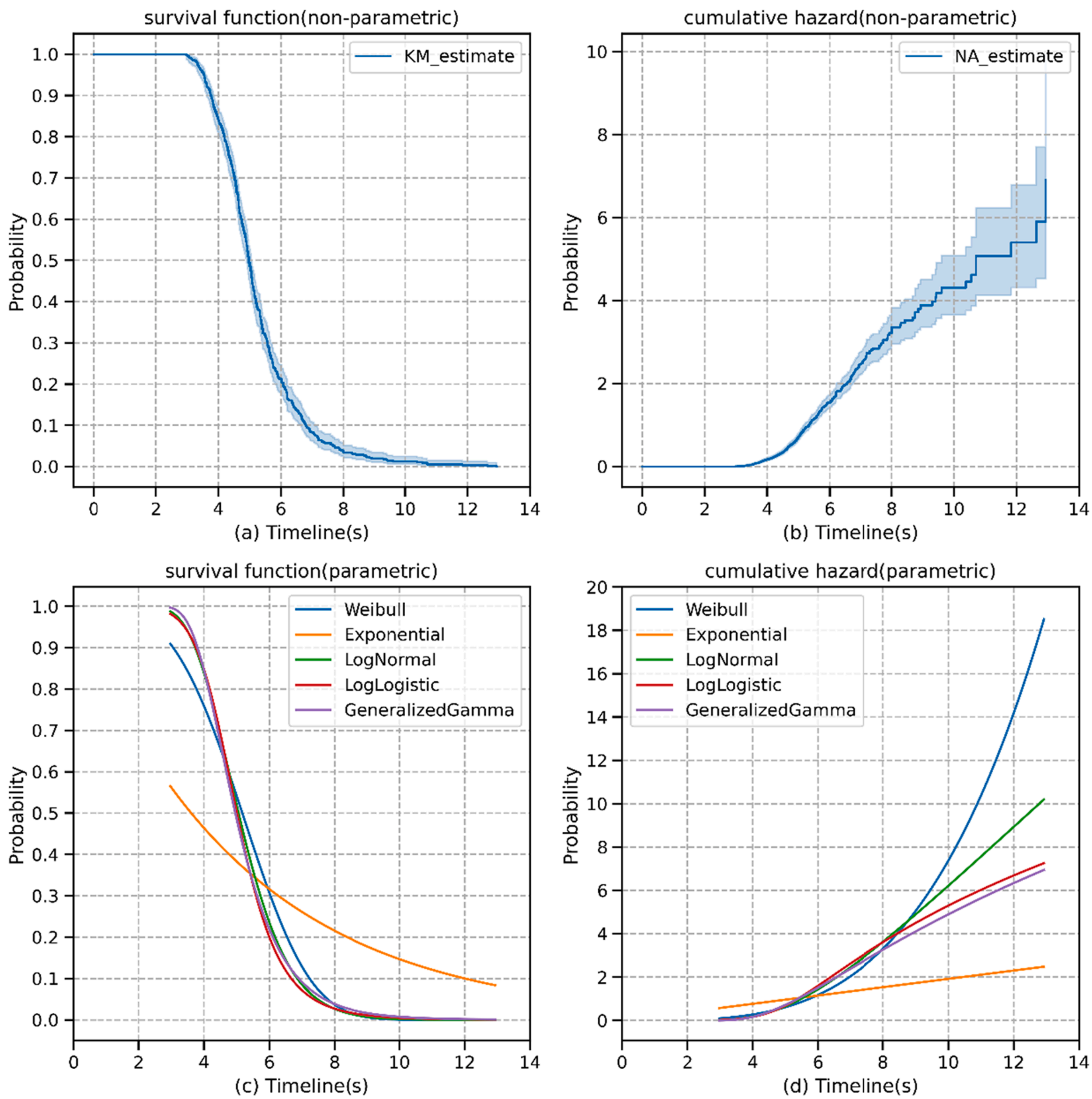


Fig. 11. The survival and cumulative hazard function of LCD using the non-parametric method (KM and NA) and the parametric method (five commonly used distribution)

Table 3
The survival function fitting results of five commonly-used distribution form.

Parametric estimator	AIC	BIC	MST
Weibull	2009.8	12693.94	5.17 s
Exponential	2971.03	18789.85	3.61 s
Lognormal	1795.48	11340.36	5.06 s
Loglogistic	1781.16	11249.78	5.00 s
Generalized Gamma	1769.03	11162.35	4.96 s

BIC penalizes model more for free parameters, and the AIC prefers a more complex over a simpler model. AIC has the danger of over fitting and BIC has the danger of under fitting. Therefore, both AIC and BIC are recommended when choosing the best parameters. We could employ these two metrics to evaluate the performance of the five different distributions. It could be found that the Exponential distribution exhibits the worst performance, while the Generalized Gamma distribution outperforms than other distributions both in AIC and BIC. This might be because the Generalized Gamma distribution is a more generalized distribution function.

The comparison of the survival and cumulative hazard function between the non-parametric univariate and parametric univariate method is presented in Fig. 12. The survival curves of the two methods have a high degree of coincidence, while there is a certain degree of deviation after 8 s in the cumulative hazard function. The corresponding percentage of at-risk vehicles is presented in Table 4. It can be found that nearly 80% of vehicles complete LC within 6 s, nearly 94% of vehicles complete LC within 8 s, nearly 99% of vehicles complete LC within 10 s. The number of vehicles that complete LC is the most during 4 s to 6 s.

4.2. *Regression analysis of LCD*

The regression analysis of LCD is presented in this subsection, where both CPH and the AFT model are employed. The results of the CPH model is presented in Table 5. It could be found that $\dot{x}_S$, $\triangle x_{S-CP}$, $\triangle x_{S-CF}$, $\triangle x_{S-TP}$, $\triangle \dot{x}_{S-TP}$, ,$\triangle x_{S-TF}$, $\triangle \dot{x}_{S-TF}$, $\triangle x_{TP-TF}$ significantly

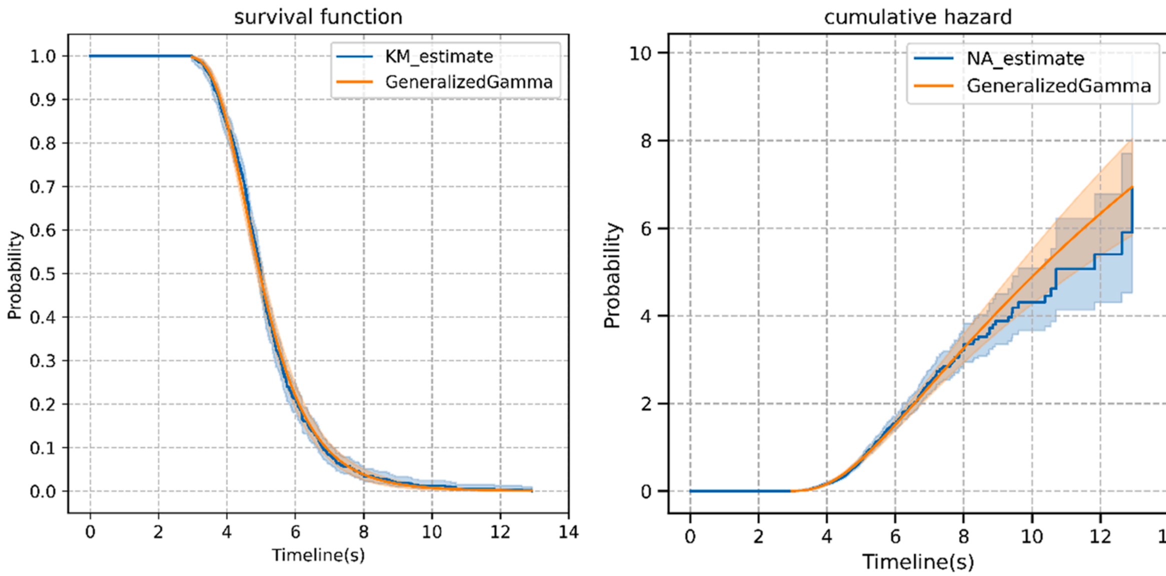


Fig. 12. Comparison between the non-parametric and parametric estimator

Table 4
The percentage of at-risk vehicles at each time points.

	LCD						
Percentage	0 s 100.00%	2 s 100.00%	4 s 84.46%	6 s 21.25%	8 s 3.93%	10 s 1.25%	12 s 0.36%

Table 5
The summary of the CPH model and the PH test.

Variable	CPH model regression result						PH test results	
	Coef	Exp (coef)	Exp(coef) lower 95%	Exp(coef) upper 95%	z	p value	z	p value
$\dot{x}_S$	0.11	1.12	1.09	1.14	10.29	**0.00**	0.07	0.80
$\ddot{x}_S$	0.23	1.25	0.96	1.64	1.66	0.10	1.53	0.22
$\ddot{x}_{CP}$	−0.22	0.80	0.61	1.06	−1.53	0.13	0.04	0.84
$\triangle x_{S-CP}$	0.00	1.00	1.00	1.01	2.01	**0.05**	1.25	0.26
$\triangle \dot{x}_{S-CP}$	−0.05	0.95	0.91	1.00	−1.90	0.06	0.07	0.79
$\ddot{x}_{CF}$	0.00	1.00	0.76	1.31	−0.01	0.99	4.38	0.06
$\triangle x_{S-CF}$	0.01	1.01	1.00	1.01	4.56	**0.00**	0.41	0.52
$\triangle \dot{x}_{S-CF}$	0.01	1.01	0.97	1.06	0.51	0.61	3.32	0.07
$\ddot{x}_{TP}$	−0.14	0.87	0.65	1.16	−0.96	0.34	1.03	0.31
$\triangle x_{S-TP}$	−0.02	0.98	0.97	1.00	−2.53	**0.01**	0.02	0.88
$\triangle \dot{x}_{S-TP}$	−0.09	0.92	0.88	0.96	−4.02	**0.00**	0.23	0.63
$\ddot{x}_{TF}$	0.13	1.14	0.82	1.59	0.79	0.43	0.20	0.65
$\triangle x_{S-TF}$	−0.01	0.99	0.97	1.00	−1.99	**0.05**	0.02	0.88
$\triangle \dot{x}_{S-TF}$	0.05	1.05	1.01	1.09	2.69	**0.01**	2.05	0.15
$\triangle x_{TP-TF}$	0.02	1.02	1.01	1.04	3.44	**0.00**	0.01	0.92

impact on the LCD. The three factors that most significantly affect LCD are: the speed of the subject vehicle, the speed difference with the target-lane preceding vehicle and the speed difference with the target-lane following vehicle. Take the variable $\dot{x}_S$ as example, one unit increase in the speed of the subject vehicle would lead to the increase of the baseline hazard by a factor of $exp(0.11) = 1.12$. The baseline hazard will increase about 12%. The higher the hazard, the more likely the subject vehicle is about to complete the LC.

PH (proportional hazard) assumption assumes that all the individuals share the same hazard function, while have different scaling factor. The PH test result is also presented in Table 5. The p-values are higher than 0.05, which indicates that all covariates satisfied the PH assumption. For some $t = \bar{t}$, suppose there are two individuals i, j, their ratio of hazard rates with vectors of characteristics x_i and x_j is:

$$\frac{h(\bar{t}|x_i)}{h(\bar{t}|x_j)} = \frac{h_0(\bar{t})exp(x_i^T\beta)}{h_0(\bar{t})exp(x_j^T\beta)} = exp\left(\left(x_i^T - x_j^T\right)\beta\right) \tag{18}$$

Where the right part is constant for all t, which indicates that the effect of the covariates in the PH model are assumed to be fixed across the whole time.

The results of three AFT models (LoglogisticAFT, LognormalAFT, WeibullAFT) are presented in Table 6. The corresponding AIC values are 1524.39, 1528.37, and 1523.38 respectively, which are relatively close to each other. Take the coefficient $\dot{x}_S$ in WeibullAFT model for example, a unit increase of the $\dot{x}_S$ will result in 2% decrease of baseline survival time T_0 as shown in (17), while a unit increase of $\triangle \dot{x}_{S-TP}$ will lead to 2%

Table 6
The summary of three different AFT models.

Variable	LoglogisticAFT (AIC = 1524.39)				LognormalAFT (AIC = 1528.37)				WeibullAFT (AIC = 1523.38)			
	Coef	Exp (coef)	z	p value	Coef	Exp (coef)	z	p value	Coef	Exp (coef)	z	p value
$\dot{x}_S$	−0.02	0.98	−11.27	**0.00**	−0.02	0.98	−11.58	**0.00**	−0.02	0.98	−17.00	**0.00**
$\ddot{x}_S$	−0.03	0.97	−1.34	0.18	−0.02	0.98	−1.02	0.31	−0.04	0.96	−1.90	0.06
$\ddot{x}_{CP}$	0.01	1.01	0.35	0.73	0.02	1.02	0.94	0.35	0.04	1.04	1.93	0.05
$\triangle x_{S-CP}$	0.00	1.00	−2.01	**0.04**	0.00	1.00	−1.91	0.06	0.00	1.00	−1.55	0.12
$\triangle \dot{x}_{S-CP}$	0.01	1.01	1.55	0.12	0.01	1.01	1.26	0.21	0.01	1.01	1.93	0.05
$\ddot{x}_{CF}$	−0.01	0.99	−0.40	0.69	−0.01	0.99	−0.51	0.61	0.01	1.01	0.51	0.61
$\triangle x_{S-CF}$	0.00	1.00	−3.69	**0.00**	0.00	1.00	−3.20	**0.00**	0.00	1.00	−4.34	**0.00**
$\triangle \dot{x}_{S-CF}$	0.00	1.00	−1.28	0.20	0.00	1.00	−0.89	0.37	0.00	1.00	0.15	0.88
$\ddot{x}_{TP}$	0.00	1.00	0.00	1.00	0.00	1.00	0.18	0.86	0.03	1.03	1.38	0.17
$\triangle x_{S-TP}$	0.00	1.00	3.27	**0.00**	0.00	1.00	3.63	**0.00**	0.00	1.00	3.48	**0.00**
$\triangle \dot{x}_{S-TP}$	0.01	1.01	3.83	**0.00**	0.02	1.02	4.14	**0.00**	0.02	1.02	4.81	**0.00**
$\ddot{x}_{TF}$	0.00	1.00	−0.10	0.92	−0.01	0.99	−0.21	0.84	−0.02	0.98	−0.60	0.55
$\triangle x_{S-TF}$	0.00	1.00	2.88	**0.00**	0.00	1.00	3.24	**0.00**	0.00	1.00	2.98	**0.00**
$\triangle \dot{x}_{S-TF}$	0.00	1.00	−1.62	0.11	−0.01	0.99	−1.90	0.06	−0.01	0.99	−2.64	**0.01**
$\triangle x_{TP-TF}$	0.00	1.00	−4.00	**0.00**	−0.01	0.99	−4.45	**0.00**	0.00	1.00	−4.70	**0.00**

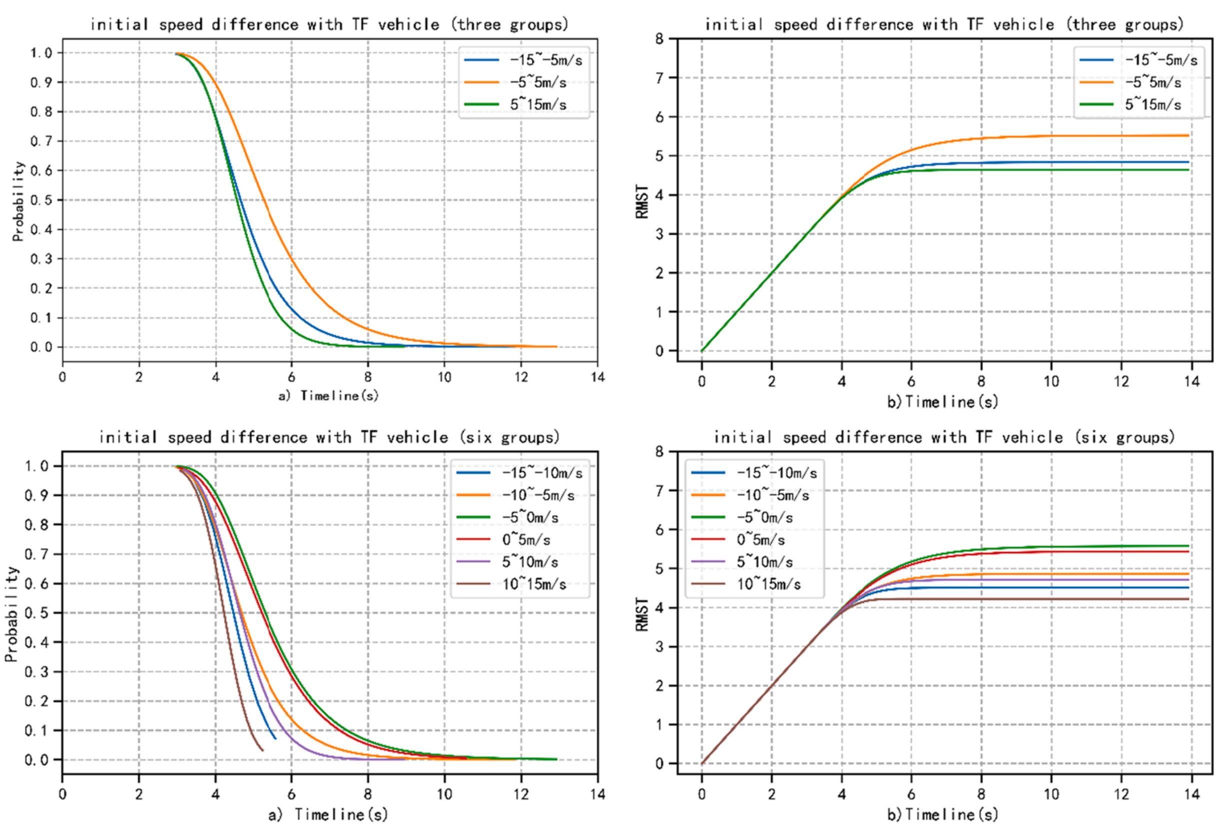


Fig. 13. Univariate analysis of the speed difference with TF vehicle

increase of LCD. So far, we have analyzed the influencing factors of LCD using the CPH model and AFT models. It could be found that there are six same variables $\dot{x}_S$, $\triangle x_{S-CF}$, $\triangle x_{S-TP}$, $\triangle \dot{x}_{S-TP}$, $\triangle x_{S-TF}$, $\triangle x_{TP-TF}$ that significant affect the LCD in four models, while the major difference is the variable $\triangle x_{S-CP}$ and the $\triangle \dot{x}_{S-TF}$. The next subsection will further investigate this difference.

4.3. Univariant analysis of LCD

The univariant analysis method is introduced in this subsection to further investigate the above difference. The Generalized Gamma distribution function and the metric of RMST (Restricted mean survival times) are employed. RMST is similar to ROC curve, which refers to the area under the survival curve in a certain period of time, that is the average survival time.

$$RMST(t) = \int_0^t S(t)dt \tag{19}$$

Comparison of survival difference via RMST integrates difference in survival probabilities from the beginning to the end of the follow-up time is a more meaningful metric. The larger the value of RMST of a group is, the longer the LCD is. We divided the data into different groups

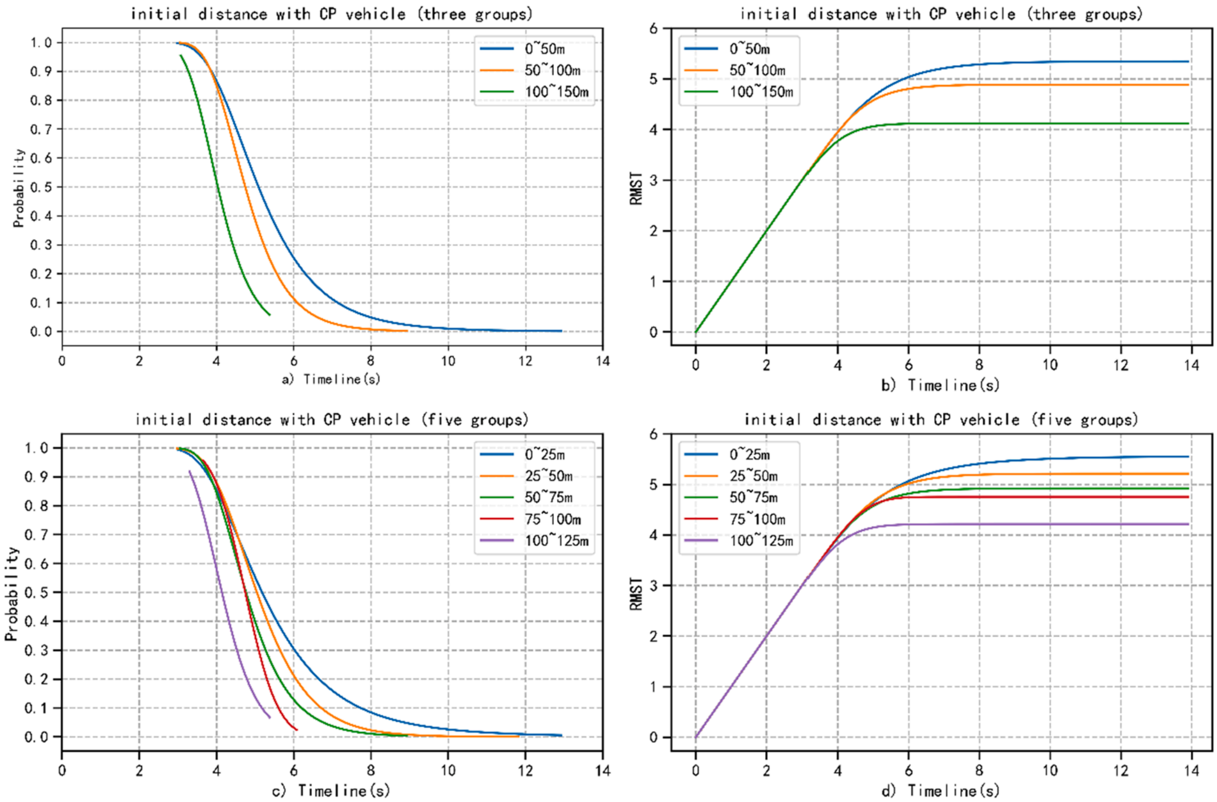


Fig. 14. Univariate analysis of the distance with CP vehicle

Table 7
MST and RMST value of LCD under different distance with CP vehicle.

Groups of $\triangle x_{S-CP}$		Confidence interval of MST		MST (s)	RMST	
		lower (s)	upper (s)		t = 4 s	t = 12 s
Group 1	0 ~ 25 m	4.97	5.43	5.18	3.95	5.54
Group 2	25 ~ 50 m	4.93	5.20	5.05	3.96	5.21
Group 3	50 ~ 75 m	4.57	5.01	4.76	3.95	4.92
Group 4	75 ~ 100 m	4.53	5.08	4.74	3.96	4.75
Group 5	100 ~ 125 m	3.82	4.60	4.13	3.20	4.22

so as to investigate the difference of LCD. Fig. 13 and Fig. 14 presents the univariate analysis of the variable $\triangle x_{S-CP}$ and $\triangle \dot{x}_{S-TF}$. It could be obviously found that with the gradual increase of $\triangle \dot{x}_{S-TF}$, the survival function and RMST function of LCD do not increase or decrease correspondingly.

While with the decrease of $\triangle x_{S-CP}$, the subject vehicle is more likely to have a longer LCD. Table 7 also presents the corresponding MST and RMST value of LCD (five groups) under different groups of $\triangle x_{S-CP}$. The MST difference between the Group 1 and Group 5 is about one second, and the RMST difference is about 1.32. Therefore, it is reasonable to consider that the influence of $\triangle x_{S-CP}$ on LCD is significant, while the $\triangle \dot{x}_{S-TF}$ is not. The results obtained by LoglogisticAFT model may be more reliable than other models. Thereafter, we could obtain the partial effects of other six covariates on LCD as shown in Fig. 15.

5. Discussion

5.1. Main findings and implications

A complete survival analysis of LCD has been conducted in this paper in order to bridge the existing research gaps. Both detailed comparative univariate and regression analysis of LCD have been conducted for the first time. One may argue such synthesis analysis seems to be trivial. However, what's really matters is the underlying consideration behind this research, which has been elaborated in the Introduction. The LC trajectories are extracted from the HighD dataset, and preliminary analysis of LCD is conducted. The HighD dataset indicates that the average LCD is about 5.2 s, and the standard deviation is about 1.32 s. And we found that the median value of LCD is lower than the average value, which indicates that half of the vehicles pull up the average LCD of all the vehicles. Then, the joint distribution between the LCD with $\triangle \dot{x}_{S-CP}$, $\triangle \dot{x}_{S-TP}$, $\triangle x_{TP-TF}$, $\dot{x}_S$ is further analyzed. Meanwhile, we also found that there is no significant difference in LCD between the vehicles moving from the fast lane to the slow lane and the vehicles moving from the slow lane to the fast lane. In order to further investigate the LCD, the survival analysis framework has been introduced and employed.

The characteristic of the whole survival function of the LCD has been researched at first. The whole survival function decreased rapidly in 3 s ~ 8 s, while decreased gradually in 8 s ~ 12 s. The corresponding MST metric is about 4.98 s, which indicates that each vehicle has a 50% chance of completing its LC maneuver when LCD equals to 4.98 s. At the same time, the at-risk table shows that nearly 15% of vehicles complete LC within 4 s, nearly 80% of vehicles complete LC within 6 s, 94% of vehicles complete LC within 8 s, 99% of vehicles complete LC within 10 s. This indicates that in the time span of 4 s to 6 s, most of the vehicles have completed LC, followed by the time span of 6 s to 8 s. Meanwhile, the comparative univariate analysis results indicate that Generalized Gamma distribution function performs better than other distributions, and has high degree of coincidence with the non-parametric method.

After that, the influencing factors of LCD have been explored through the comparative analysis of the CPH model and the AFT model. Through the univariate analysis method, we found that the LoglogisticAFT model exhibits more credible results than other regression models. Regression results demonstrate that there are seven variables that significantly affect the LCD, which are $\dot{x}_S$, $\triangle \dot{x}_{S-TP}$, $\triangle x_{TP-TF}$, $\triangle x_{S-TP}$, $\triangle x_{S-TF}$, $\triangle x_{S-CF}$ and $\triangle x_{S-CP}$. It is interesting to find that the distance with the

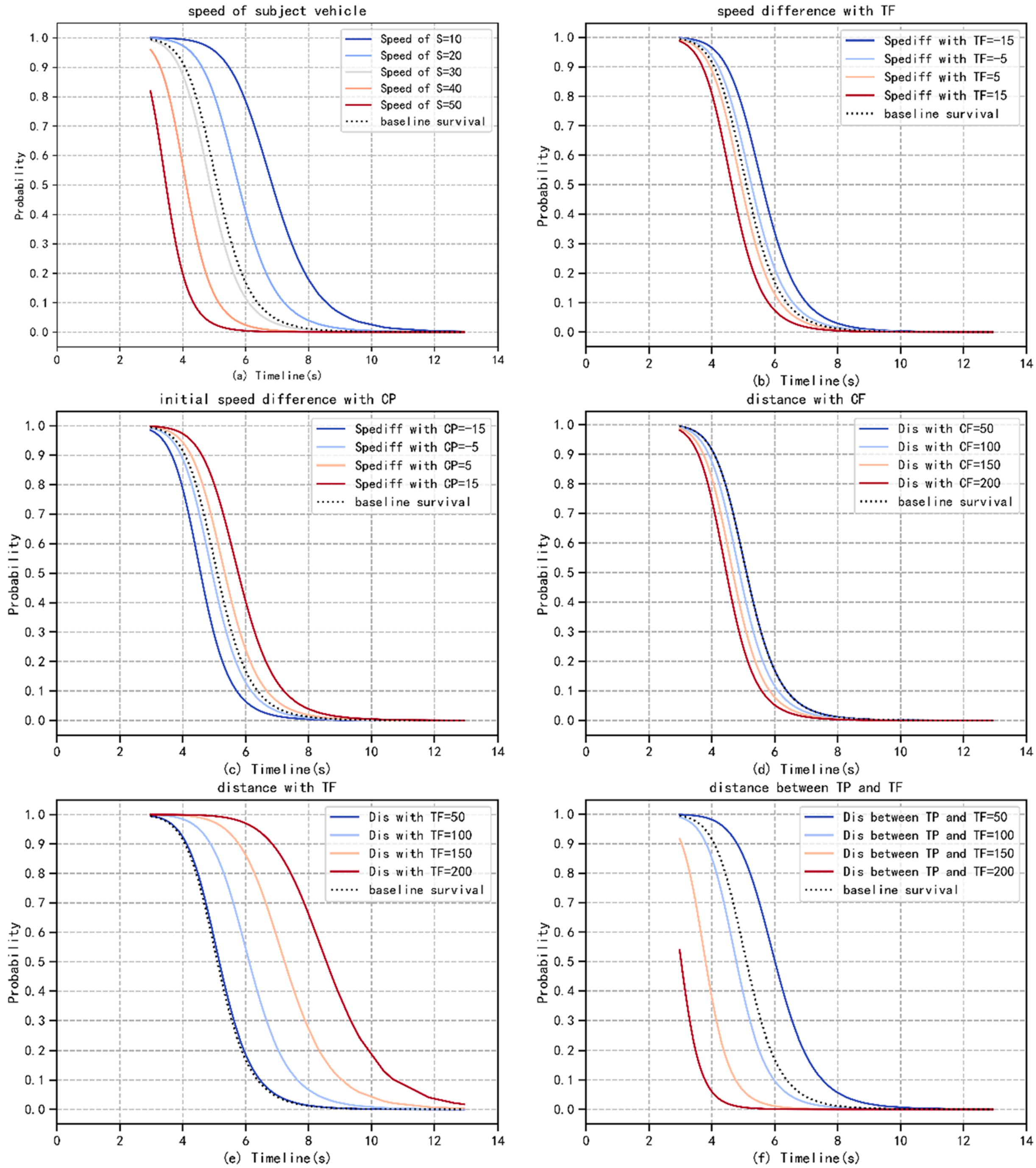


Fig. 15. Partial effects of covariates on LCD using LoglogisticAFT model

current-lane following vehicle would slightly affect the LCD. To some extent, this may indicate that the driver may pay attention to the distance with the current-following vehicle when the driver is about to execute the LC action. Subsequent research might be necessary to further extend this finding. The most important variable that affect the LCD is the speed of the subject vehicle, in turn is the speed difference with the target-preceding vehicle, is followed by the initial gap distance on the target lane.

5.2. Synthesis and future work

The purpose of this research is to choose the perspective of survival analysis to conduct a comprehensive analysis of LCD. The reason why we choose this perspective is due to its capacity in formulating the relationships among the influencing factors, survival time and the outcome of each observation sample. Through such comprehensive analysis, some novel findings have been discovered and summarized above. The findings in the comparative univariate and regression analysis may provide certain guidance for traffic modeling in the future. When analyzing other datasets, the Generalized Gamma distribution function and the LoglogisticAFT model could be considered first. The MST metric, RMST metric, and the influencing factors of LCD may assist us explore the differences in LC behavior between different regions or different time periods or different type of drivers. These differences might contribute to our research of driving behaviors, or might be directly/ indirectly applicable to the design of the ADAS (Advanced

Driving Assistance System). For instance, the approximate LCD for the vehicle to complete LC could be estimated in advance, and several important influencing factors could be adopted as the preliminary input variables for the LC trajectory prediction model.

Undoubtedly, many aspects of this paper need further research. First, given page limit, we only introduce several most commonly-used models in this paper, and we mainly focus on exploring the influencing factors of successful LC events. Future research may try to cover a wider range of survival models, and investigate the influencing factors of unsuccessful LC events. Second, subject to data restrictions, we only investigate the impacts of the interactions with surrounding vehicles, though other factors may also affect LCD [12–15]. Therefore, how to collect LC trajectory data containing more variable information may become one of the goals of future research. And we believe the survival analysis framework is also applicable to other datasets.

6. Conclusion

In this paper, we present a comprehensive analysis of LCD from the perspective of survival analysis. We extracted a total number of 560 LC trajectories from the HighD dataset, which is measured from an aerial perspective in Germany highways. Thereafter, we conduct univariate analysis on this dataset and examine the characteristic of the survival function. Furthermore, we perform survival regression analysis on this dataset, and investigate the influencing factors of LCD. In such kind of synthesis analysis, some novel findings have been made, and the modeling implications have also been discussed. We hope these findings could contribute to improving our further understanding of LCD and LC behaviors.

CRediT authorship contribution statement

Yang Li: Conceptualization, Data curation, Writing - original draft. **Linbo Li:** Methodology, Funding acquisition, Writing - original draft. **Daiheng Ni:** Investigation, Writing - review & editing. **Yue Zhang:** Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by the National Key R&D Program of China [grant numbers 2018YFE0102800].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.measurement.2021.109707.

References

[1] Z. Zheng, Recent Developments And Research Needs In Modeling Lane Changing, Transport. Res. Part B: Methodolog. 60 (2014) 16–32.
[2] E. Balal, R.L. Cheu, T. Sarkodie-Gyan, A binary decision model for discretionary lane changing move based on fuzzy inference system, Transport. Res. Part C: Emerg. Technol. 67 (2016) 47–61.
[3] M.-Y. Pang, B. Jia, D.-F. Xie, X.-G. Li, A probability lane-changing model considering memory effect and driver heterogeneity, Transportmetrica B: Trans. Dynam. 8 (1) (2020) 72–89.
[4] Y. Luo, Y. Xiang, K. Cao, K. Li, A dynamic automated lane change maneuver based on vehicle-to-vehicle communication, Transport. Res. Part C: Emerg. Technol. 62 (2016) 87–102.
[5] H. Bai, J. Shen, L. Wei, Z. Feng, Accelerated Lane-Changing Trajectory Planning of Automated Vehicles with Vehicle-to-Vehicle Collaboration, J. Adv. Transport. 2017 (2017) 8132769.
[6] D.a. Yang, S. Zheng, C. Wen, P.J. Jin, B. Ran, A Dynamic Lane-Changing Trajectory Planning Model For Automated Vehicles, Transport. Res. Part C: Emerg. Technol. 95 (2018) 228–247.
[7] M. Yue, X. Hou, J. Zhao, X. Wu, Robust Tube-Based Model Predictive Control for Lane Change Maneuver of Tractor-Trailer Vehicles Based on a Polynomial Trajectory, IEEE Trans. Syst., Man, Cybernet.: Syst. (2018) 1–9.
[8] J. Huang, h.-x. Ji, X.-Y. Peng, L. Hu, Driving Style Adaptive Lane-changing Trajectory Planning and Control, China J. Highway Trans. 32 (2019) 226–239.
[9] Yimin Chen, Chuan Hu, Junmin Wang, Motion Planning With Velocity Prediction and Composite Nonlinear Feedback Tracking Control for Lane-Change Strategy of Autonomous Vehicles, IEEE Trans. Intell. Veh. 5 (1) (2020) 63–74.
[10] L. Li, Y. Li, D. Ni, Y. Zhang, Dynamic trajectory planning for automated lane-changing, The 97th Annual Meeting of the Transportation Research Board, Washington, DC, 2021.
[11] Xiaohui Zhang, Jie Sun, Xiao Qi, Jian Sun, Simultaneous modeling of car-following and lane-changing behaviors using deep learning, Transport. Res. Part C: Emerg. Technol. 104 (2019) 287–304.
[12] T. Toledo, D. Zohar, Modeling Duration Of Lane Changes, Transp. Res. Rec. 2007 (1999) 71–78.
[13] Eleni I. Vlahogianni, Modeling Duration Of Overtaking In Two Lane Highways, Transport. Res. Part F: Traffic Psychol. Behav. 20 (2013) 135–146.
[14] Q. Wang, Z. Li, L. Li, Investigation Of Discretionary Lane-Change Characteristics Using Next-Generation Simulation Data Sets, J. Intell. Transport. Syst. 18 (2014) 246–253.
[15] Minming Yang, Xuesong Wang, Mohammed Quddus, Examining Lane Change Gap Acceptance, Duration And Impact Using Naturalistic Driving Data, Transport. Res. Part C: Emerg. Technol. 104 (2019) 317–331.
[16] E.C. Olsen, S.E. Lee, W.W. Wierwille, M.J. Goodman, Analysis Of Distribution, Frequency, And Duration Of Naturalistic Lane Changes, Proc. Hum. Factors Ergonom. Soc. Ann. Meeting 46 (2002) 1789–1793.
[17] S. Moridpour, M. Sarvi, G. Rose, Modeling The Lane-Changing Execution Of Multiclass Vehicles Under Heavy Traffic Conditions, Transp. Res. Rec. 2161 (2010) 11–19.
[18] K. Aghabayk, S. Moridpour, W. Young, M. Sarvi, Y.-B. Wang, Comparing Heavy Vehicle And Passenger Car Lane-Changing Maneuvers On Arterial Roads And Freeways, Transp. Res. Rec. 2260 (2011) 94–101.
[19] X. Cao, W. Young, M. Sarvi, Exploring Duration Of Lane Change Execution, Australasian Transport Research Forum, Australia, 2013.
[20] J. Wu, S. Zhang, A.K. Singh, S. Qin, Hazard-Based Model Of Mandatory Lane Change Duration, 17th COTA International Conference of Transportation, China, 2008.
[21] Liu Yang, Xiaomeng Li, Wei Guan, H. Michael Zhang, Lingling Fan, Effect Of Traffic Density On Drivers' Lane Change And Overtaking Maneuvers In Freeway Situation—A Driving Simulator-Based Study, Traffic Inj. Prev. 19 (6) (2018) 594–600.
[22] Saravanan Gurupackiam, Steven Lee Jones, Empirical Study Of Accepted Gap And Lane Change Duration Within Arterial Traffic Under Recurrent And Non-Recurrent Congestion, Int. J. Traffic Trans. Eng. 2 (4) (2012) 306–322.
[23] S. Washington, M.G. Karlaftis, F. Mannering, P. Anastasopoulos, Statistical And Econometric Methods For Transportation Data Analysis, CRC Press, 2020.
[24] Huizhao Tu, Hao Li, Hans van Lint, Henk van Zuylen, Modeling travel time reliability of freeways using risk assessment techniques, Transport. Res. Part A: Policy Pract. 46 (10) (2012) 1528–1540.
[25] J.W.C. van Lint, Henk J. van Zuylen, H. Tu, Travel time unreliability on freeways: Why measures based on variance tell only half the story, Transport. Res. Part A: Policy Pract. 42 (1) (2008) 258–277.
[26] E.I. Vlahogianni, Modeling Duration Of Overtaking In Two Lane Highways Transportation Research Part F: Traffic Psychology And Behaviour, 20 (2013) 135–146.
[27] Chandra R. Bhat, Teresa Frusti, Huimin Zhao, Stefan Schönfelder, Kay W. Axhausen, Intershopping duration: an analysis using multiweek data, Transport. Res. Part B: Methodol. 38 (1) (2004) 39–60.
[28] F.L. Mannering, M.M. Hamed, Occurence, frequency, and duration of commuters' work-to-home departure delay, Transport. Res. Part B: Methodol. 24 (1990) 99–109.
[29] D.A. Niemeier, J.G. Morita, Duration of trip-making activities by men and women, Transportation 23 (1996) 353–371.
[30] A. Tavassoli Hojati, L. Ferreira, S. Washington, P. Charles, Hazard based models for freeway traffic incident duration, Accid. Anal. Prev. 52 (2013) 171–181.
[31] D. Nam, F. Mannering, An exploratory hazard-based analysis of highway incident duration, Transport. Res. Part A: Policy Pract. 34 (2000) 85–102.
[32] R. Krajewski, J. Bock, L. Kloeker, L. Eckstein, The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems, 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 2018, pp. 2118–2125.
[33] D.G. Kleinbaum, M. Klein, Survival Analysis: A Self-Learning Text, Third Edition, Survival Analysis: A Self-Learning Text, Third Edition, Springer, New York, 2012, pp. 1–700.
[34] E.L. Kaplan, P. Meier, Nonparametric estimation from incomplete observations, J. Am. Stat. Assoc. 53 (1958) 457–481.
[35] V. Alexiadis, J. Colyar, J. Halkias, R. Hranac, G. McHale, The Next Generation Simulation Program, Institute of Transportation Engineers, ITE J. 74 (2004) 22.