Analysis of discretionary lane-changing behaviors of autonomous vehicles based on real-world data

Xiao (Luke) Wen[b], Chunxi Huang[d], Sisi Jian[c*], Dengbo He[a, c, e*]

[a] Intelligent Transportation Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China

[b] Intelligent Transportation, Division of Emerging Interdisciplinary Areas (EMIA) under Interdisciplinary Programs Office (IPO), The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR

[c] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR

[d] Robotics and Autonomous Systems, Interdisciplinary Programs Office, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR

[e] HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen, China

[*]Corresponding author. Email: cesjian@ust.hk; dengbohe@hkust-gz.edu.cn

Word count: 9,719 words

Abstract

With the deployment of autonomous vehicles, a transition period where autonomous vehicles share the roads with human-driven vehicles is inevitable where the discretionary lane-changing behaviors of autonomous vehicles can be safety-critical. This study aims to quantify the impact of discretionary lane-changing behaviors on following vehicles in the target lane using a real-world dataset. This study uses the Waymo Open Dataset to identify the differences between the discretionary lane-changing maneuvers of autonomous vehicles and human-driven vehicles and compare their impacts on the driving volatility metrics. Then, the block maxima model is applied to estimate the crash risks. Finally, the multivariate adaptive regression splines model is adopted to model gap acceptance behaviors of autonomous vehicles and human-driven vehicles. Results show that compared to human-driven vehicle discretionary lane-changing, autonomous vehicle discretionary lane-changing leads to lower speeds and yaw rate volatility and smaller acceleration rates of the following vehicles. Further, the block maxima model reveals that the crash risk in the autonomous vehicle discretionary lane-changing events is half of that in the human-driven vehicle discretionary lane-changing events. In addition, autonomous vehicles and human-driven vehicles show different lead gap acceptance behaviors, according to the results of multivariate adaptive regression splines. The findings highlight the benefits of mixing autonomous vehicles in traffic flow and guide the improvement of autonomous vehicle controllers.

Keywords: Autonomous vehicles; Lane-changing; Gap acceptance; Extreme value theory; Crash risk

1. Introduction

With the gradual deployment of autonomous vehicles, human-driven vehicles are expected to share the roads with autonomous vehicles shortly, leading to a transition period with mixed traffic, in which human drivers may exhibit different behaviors as compared to when the traffic is composed of only human-driven vehicles (Mahdinia et al., 2021; Wen et al., 2022b; Zhao et al., 2020). Hence, understanding human-driven vehicles' behavioral changes in mixed traffic is the foundation of the analysis of autonomous vehicle impacts on traffic safety, traffic efficiency, energy consumption and exhaust emissions (Hu et al., 2022). Further, modeling the interactions between autonomous vehicles and human-driven vehicles can provide insights into the improvements of autonomous vehicle control algorithms and guide appropriate public policies toward the acceptance of autonomous vehicles (Di and Shi, 2021).

As limited by the low market penetration rates of autonomous vehicles at the current stage, empirical data on autonomous vehicles and surrounding traffic is scarce. When investigating the impacts of autonomous vehicles on the surrounding traffic, previous studies mostly adopted two approaches, i.e., traffic/numerical simulations (e.g., Dixit et al., 2019) and field experiments (e.g., Mahdinia et al., 2021). However, traffic/numerical simulations may simplify and even omit important features of mixed traffic flow, resulting in questionable effects of autonomous vehicles. Field experiments are usually conducted with limited sample sizes (i.e., the number of driving events) and cannot replicate the driving scenarios with large speed fluctuations and complex interactions between road agents. The limitations of these two approaches may induce biased results. With the development of autonomous vehicle technologies, more and more autonomous vehicles are being tested or implemented on public roads in recent years and some tech firms (e.g., Waymo) have released large-scale real-world datasets collected by their autonomous vehicle fleets (Wen et al., 2023). These datasets contain fine-grained field observations of autonomous vehicle movements and behaviors of road agents surrounding the autonomous vehicles on public roads and thus can provide the transportation research community with new opportunities to analyze the impacts of autonomous vehicles on mixed traffic flow, as well as human-driven vehicles' behavioral adaptations when interacting with autonomous vehicles in the real world.

Previous literature on autonomous vehicle-human-driven vehicle interactions mainly focused on the car-following scenario, in which human-driven vehicles drive behind autonomous vehicles in the same lane (Mahdinia et al., 2021; Rahmati et al., 2019; Wen et al., 2022b; Zhao et al., 2020). In contrast, the lane-changing scenario, which is often related to rear-end and sideswipe crashes (Ali et al., 2022a), is rarely studied. The lane-changing scenario is correlated with both longitudinal and lateral movements of involving road agents. In the lane-changing scenario, the lead vehicle changes the current lane into an adjacent lane, which may cause the following vehicle in the target lane to decelerate or stop, leading to the formation of stop-and-go oscillations and bottlenecks in traffic flow (Jiang et al., 2021; Jiang et al., 2022). Based on the intention of drivers, Yang and Koutsopoulos (1996) categorized lane-changing scenarios into two types, i.e., mandatory lane-changing and discretionary lane-changing. The former is a required task and must be conducted to reach a specific destination while the latter is voluntary and usually carried out to improve the current driving conditions. Therefore, the latter is more difficult to predict and more complex and dangerous than the former (Ali et al., 2022b; Toledo et al., 2005). Hence, our study focuses on the fundamental mechanisms of autonomous vehicle-human-driven vehicle interactions in the discretionary lane-changing scenario, i.e., how the discretionary lane-changing behaviors of autonomous vehicles affect the behaviors of surrounding human-driven vehicles, especially the following vehicles in the target lane.

In the current study, we take the first attempt to study the impacts of autonomous vehicles' discretionary lane-changing maneuvers on mixed traffic in terms of driving volatility, crash risks and gap acceptance and compare them to those of human-driven vehicles' discretionary lane-changing maneuvers. Trajectories of autonomous vehicles' and surrounding human-driven vehicles' are extracted and processed from the real-world autonomous driving dataset -- Waymo Open Dataset (Ettinger et al., 2021). We summarize our contribution as follows. First, rather than traffic microsimulation and field experiments, it uses the real-world dataset which provides more insights into complicated driving conditions in reality. Second, an in-depth analysis is conducted to explore the traffic and safety effects of autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing by quantifying driving volatility and crash risks. Third, gap acceptance behaviors of autonomous vehicles and human-driven vehicles are modeled and compared to understand the discretionary lane-changing decision-making process.

The paper is organized in the following manner. The next section reviews relevant studies on the analysis of human-driven vehicle lane-changing maneuvers and autonomous vehicle-human-driven vehicle interactions. Sections 3 and 4 present the data sources used in this study and methodologies for discretionary lane-changing behavior measurements and comparisons, respectively. Section 5 discusses the comparison results of discretionary lane-changing characteristics including driving volatility, crash risks and gap acceptance, with Section 6 providing conclusions and research recommendations.

2. Literature review

This section mainly reviews studies related to two research topics: (1) analysis of human-driven vehicles' lane-changing behaviors based on naturalistic driving datasets; and (2) interactions between autonomous vehicles and human-driven vehicles.

2.1 Analysis of human-driven vehicle lane-changing behaviors

There are three characteristics related to the lane-changing behaviors of human-driven vehicles, including lane-changing duration, impacts on the following vehicle in the target lane and gap acceptance. This part covers representative studies corresponding to these characteristics.

lane-changing duration defines the time span of the lane-changing execution, which starts when the lane-changing vehicle in the current lane initiates its movement toward the target lane, and ends when the lane-changing vehicle stabilizes in the target lane (Wang et al., 2019; Yang et al., 2019). Naturalistic driving studies show that the duration of lane-changing can be best fitted through the lognormal distribution (Das et al. 2020 and Yang et al. 2019). For example, Wang et al. (2019) analyzed the real-world driving data collected from the Shanghai Naturalistic Driving Study (SH-NDS) and concluded that the lognormal distribution was the best fit for the lane-changing duration data. Further, the duration of lane-changing may be affected by the acceleration behavior of the lane-changing vehicle and the response of surrounding vehicles (Toledo and Zohar, 2007). Toledo and Zohar (2007) revealed that the inappropriate settings of lane-changing duration in microscopic simulations might negatively affect the realism of the microsimulation. In another study, Li et al. (2023) analyzed discretionary lane-changing duration using accelerated failure time (AFT) models based on vehicle types and discretionary lane-changing direction which also consider the heterogeneity of human drivers.

The lane-changing maneuvers of human-driven vehicles can affect the driving behaviors of following vehicles in the target lane. For instance, Sultan et al. (2002) found that a sudden lane-changing could lead to the abrupt acceleration of the following vehicle, which might lead to excessive exhaust emissions and fuel consumption, as well as stop-and-go oscillations, impairing traffic efficiency and safety (Jiang et al., 2021; Jiang et al., 2022). Wang et al. (2019) revealed that most braking behaviors of following vehicles occurred when lane-changing events were at the initial phase, i.e., before the lane-changing vehicle entered the target lane. Researchers further found that the effects of lane-changing events on the following vehicles' behaviors depend on the road attributes. For example, Yang et al. (2019) concluded that the effects of lane-changing events on the following vehicle speed depend on the road type. Mauch and Cassidy (2002) found that traffic oscillations were more likely to form near the facilities with more lane-changing events.

Another critical element of the lane-changing decision-making process is gap acceptance. Before performing lane-changing behaviors, drivers will evaluate whether the longitudinal gaps between them and the vehicles in the target lane are acceptable. The gaps are of two types including the lead gap representing the longitudinal distance between the lead vehicle in the target lane and the lane-changing vehicle, and the lag gap representing the longitudinal distance between the following vehicle in the target lane and the lane-changing vehicle (Toledo et al., 2003). In this paper, gaps are calculated in terms of time rather than distance suggesting that gaps are a function of the longitudinal distance and vehicle speed. This is because the available gaps of lane-changing vehicles are correlated with the current speed which may vary, which makes time gaps more generalizable (Bham, 2009). In previous research, commonly-used methods to model lane-changing vehicle gap acceptance include rule-based models (Jin et al., 2019), gametheoretic models (Ji and Levinson, 2020), linear regression models (Wang et al., 2019; Yang et al., 2019) and multivariate adaptive regression splines model (Das et al., 2020) and the gap acceptance in lane-changing events is significantly affected by several variables, such as relative position and relative speed (Bham, 2009; Toledo et al., 2003).

2.2 Autonomous vehicle-human-driven vehicle interactions

Current studies investigating autonomous vehicle-human-driven vehicle interactions are mainly based on two different views: (1) some studies adopted the conventional models to depict human-driven vehicles' behaviors assuming that they will drive the same way even if they can distinguish autonomous vehicles, e.g., traffic/numerical simulations; and (2) the others assumed that people's behaviors will change significantly in response to the existence of autonomous vehicles, e. g., field experiments (Di and Shi, 2021).

For example, Papadoulis et al. (2019), Sinha et al. (2020) and Zheng et al. (2020) considered the scenario where human-driven vehicles were following autonomous vehicles by developing traffic/numerical simulation platforms. They found that autonomous vehicles had significant efficiency and safety advantages compared to the scenario where human-driven vehicles were following human-driven vehicles, e.g., in human-driven vehicle-following-autonomous vehicle, the speed standard deviation (Std) of human-driven vehicles would be decreased with the increment of autonomous vehicle market penetration rates. In their studies, human-driven vehicles' car-following behaviors were depicted by traditional models, such as Wiedemann 74 and Wiedemann 99 models. Based on field experiments, Rahmati et al. (2019), Zhao et al. (2020) and Mahdinia et al. (2021) found that human-driven vehicles may exhibit different behaviors when following autonomous vehicles as compared to when following human-driven vehicles. In their studies, several human drivers were recruited to drive behind autonomous vehicles which were fulfilling either the human-driven vehicle's or autonomous vehicle's speed files. Similarly, traffic, safety and environmental benefits for human-driven vehicle-following-autonomous vehicle were identified and quantified. For example, Mahdinia et al. (2021) observed 20.9% larger values for minimum time-to-collision that indicate much safer car-following behaviors and lower rear-end crash risks in human-driven vehicle-following-autonomous vehicle. Besides car-following behaviors, the lane-changing behaviors of autonomous vehicles have also been found to impact the driving behaviors of following vehicles. For example, using a series of field experiments, Wang et al. (2021b) revealed that the lane-changing of autonomous vehicles induced more comfortable and safer responses of the following human-driven vehicles in the target lane as compared to the lane-changing of human-driven vehicles, leading to smaller acceleration, speed Std, and yaw rates of the following human-driven vehicles. According to Dong et al. (2021), as the penetration rates of cooperative adaptive cruise control (CACC) vehicles increased, there would be considerable benefits for road capacity and traffic safety at an off-ramp bottleneck.

However, for traffic/numerical simulations, complex traffic flow and heterogeneous driving behaviors are simplified, leading to a biased estimation of the effects of autonomous vehicles. Field experiments are usually conducted in a dedicated testbed hiring a limited number of drivers that cannot mimic the complicated mixed traffic environment. To this end, analyses of human-driven vehicles' behavioral adaptations using the realistic dataset are essential to quantify the impacts of autonomous vehicles on the safety and efficiency of mixed traffic.

3. Data description

3.1 Waymo motion dataset

Discretionary lane-changing events analyzed in this study are extracted from the Waymo Open Dataset. Waymo is a leading autonomous vehicle tech firm and has been conducting road tests using SAE Level 4 autonomous vehicles for more than 32 million km (kilometers) in the U.S. Waymo cars collect high-resolution data on autonomous vehicles' movements and environments surrounding autonomous vehicles at 10-Hz frequency. As shown in Figure 1, Waymo cars have distinguished exteriors (i.e., protruding cameras and frames) and Waymo stickers as well as the LiDAR sensors on the roof, all of which make their appearance distinguishable from normal human-driven vehicles and thus allow surrounding human-driven vehicles to recognize them. The Waymo Open Dataset is constituted of two datasets: the perception and motion datasets. Note that only the motion dataset is used in this study since the perception dataset includes very few lane-changing events (Hu et al., 2022).

Figure 1. Exterior appearance of Waymo cars

28,358 clips of 20-second scenes representing approximately 157.5 hours of driving data are retrieved from the motion dataset. Each scene in the motion dataset contains high-quality 3D ground truth bounding boxes and the speed vectors for each road user (e.g., vehicles, pedestrians, and cyclists). A high-resolution map for each scene is attached as a set of polylines and polygons sampled at 0.5 meters (Ettinger et al., 2021). The motion dataset contains high-quality and continuous records of road agents' type, size (e.g., length, width and height), position and movements (e.g., speed profile and yaw angle).

3.2 Lane-changing event extraction

Yang and Koutsopoulos (1996) indicated that lane-changing motivations were categorized into either mandatory or discretionary. Mandatory lane-changing (Mlane-changing) has three primary motivations: the vehicle has to change the lane to make a turn when approaching an intersection, the vehicle is entering or exiting the traffic facility with limited access, and the vehicle is avoiding obstacles. The major motivation for discretionary lane-changing (discretionary lane-changing) is to improve the driving condition, e.g., changing to the fast lane and avoiding the slow lead vehicle.

All the 20-second clips are manually reviewed by the research team to detect discretionary lane-changing events. Note that since the sample size of discretionary lane-changing events on highways is limited (40 events for autonomous vehicle discretionary lane-changing), only discretionary lane-changing events that occur on surface roads are used in this study. Taking into consideration the sample size and sensor detection range, following the previous studies (Das et al., 2020; Toledo and Zohar, 2007; Wang et al., 2019; Yang et al., 2019), the criteria for extracting discretionary lane-changing events are defined as follows: (1) the lane-changing vehicle should move from the current lane to the neighboring lane. lane-changing vehicles that cross more than one lane are considered as multiple lane-changing events; (2) the longitudinal distance from the lane-changing vehicle to the following vehicle should be less than $75m$ to guarantee that the lane-changing maneuver has direct effects on the following vehicle; and (3) the speed of both the following vehicle and the lane-changing vehicle should always be more than $1m/s$. This rule ensures that the two vehicles are moving.

For each detected discretionary lane-changing event, lane-changing vehicle and following vehicle trajectories are derived and confirmed whether they comply with the above criteria. After screening, 180 autonomous vehicle discretionary lane-changing and 178 human-driven vehicle discretionary lane-changing events have been extracted from the dataset. For each discretionary lane-changing event, we employ the second-order Savitzky–Golay filter to filter the speed and acceleration data to remove measurement noises.

Following the guidelines in Ali et al. (2022a), we use the lane lateral shift profile of the lane-changing vehicle to determine the start and end moments of a discretionary lane-changing. The lane lateral shift profile describes the lateral position offset corresponding to the closest lane center. Figure 2 shows a sample of the discretionary lane-changing maneuver (from left to right) extracted from the autonomous vehicle discretionary lane-changing dataset, which illustrates three key points, including the start point, cross-lane point and end point. The start moment of the monotonical decrease of the lane lateral shift marks the start point of a discretionary lane-changing maneuver (red line in Figure 2). When the center of the lane-changing vehicle crosses the lane boundary (blue line in Figure 2), the sign of the lane lateral shift value will change since the lane-changing vehicle becomes closer to the center of the target lane. The end of a discretionary lane-changing maneuver is defined as the first peak after the cross-lane point (green line in Figure 2). The time difference between the start and the end points is defined as the duration of discretionary lane-changing which is approximately $6.6s$.

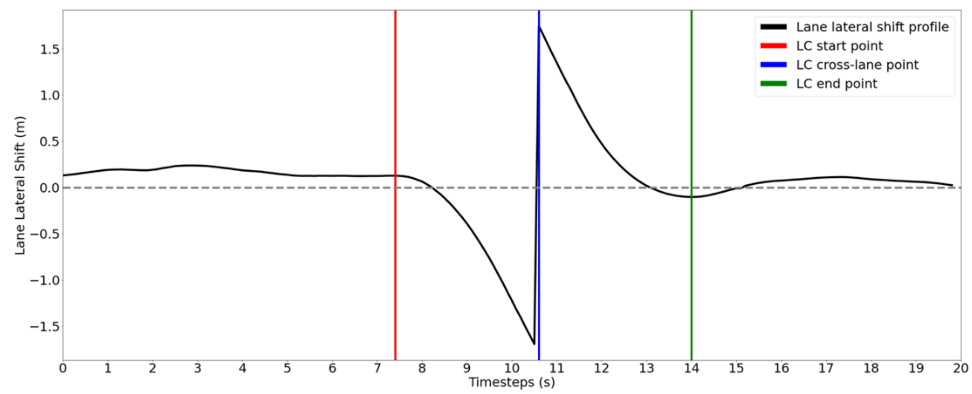


Figure 2. Sample of discretionary lane-changing duration using lane lateral shift profile

4. Methodology

This section will describe the procedures for the extraction of variables and methodologies for the discretionary lane-changing characteristic analysis. The methodology framework consists of four components. First, discretionary lane-changing duration, which measures the period of the discretionary lane-changing execution phase, is modeled. The second and third components are used to measure the effects of discretionary lane-changing behaviors on the following vehicle. Specifically, the driving volatility is introduced to capture variations in instantaneous driving decisions of the following vehicle in the target lane during the lane-changing maneuver. Then, the extreme value theory is used to estimate the crash risks in discretionary lane-changing events based on observed traffic conflicts between lane-changing vehicles and following vehicles. Finally, a non-parametric machine learning model -- multivariate adaptive regression splines is adopted to model the gap acceptance behaviors.

4.1 discretionary lane-changing duration

discretionary lane-changing duration, as one of the most important parameters in discretionary lane-changing maneuvers, has significant effects on the surrounding vehicles in congested traffic flow (Wang et al., 2021b). In this paper, discretionary lane-changing duration data is estimated at first because subsequent analysis will be conducted using the data collected within the discretionary lane-changing duration. Various candidate distributions have been used to model autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing duration data, such as exponential, gamma, normal, lognormal and logistic. Akaike Information Criterion (AIC) is chosen to measure the goodness-of-fit for each type of distribution where the distribution with the lowest AIC value will be selected.

4.2 discretionary lane-changing effects measurements

4.2.1 Driving volatility

The driving volatility measures are adopted to quantify the deviation of driving behaviors through the extraction of useful information from longitudinal and lateral vehicle control. In previous studies, several volatility functions have been developed to assess the variation in vehicle speed, acceleration and yaw rate (e.g., Arvin et al., 2019). It has been identified that higher driving volatility is correlated with higher driver instability, which is associated with higher crash risks, more energy consumption, and increased exhaust emissions (Wen et al., 2022b). Three groups of volatility measures are defined and calculated for the selected discretionary lane-changing events: speed-based volatility, acceleration-based volatility, and yaw-rate-based volatility. The mathematical equations of driving volatility functions are shown in Eqs. (1) and (2).

Standard deviation (Std): Std is one of the most commonly-used variation measures, which can be calculated as follows:

$$Std = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{1}$$

Mean Absolute Deviation (D_{mean}): D_{mean} represents the average distance between the observations and the mean value and can be computed as follows:

$$D_{mean} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n} \quad (2)$$

where x_i is the ith observation, $\bar{x}$ is the mean value of observations and n is the sample size. Both Std and D_{mean} can be applied to speed, acceleration and yaw rate.

Note that the driving volatility measure is computed for each following vehicle involved with the selected discretionary lane-changing events. Specifically, only observations that are collected within the discretionary lane-changing duration are used for the calculation. The outcomes of Eqs. (1) and (2) are aggregated values that represent each following vehicle's driving volatility during the discretionary lane-changing maneuvers.

4.2.2 Extreme value theory

Instead of using the crash data, we analyze traffic conflicts based on extreme value theory for three reasons: (1) the crash data related to autonomous vehicles and human-driven vehicles is quite limited so far; (2) the use of crash data is a reactive approach meaning that the crash has to occur first leading to the ethical dilemma of observing crashes to prevent crashes; and (3) the crash data is in an aggregated manner which makes the model incapable of providing insights into detailed driving behaviors (Farah and Azevedo, 2017; Wang et al., 2018; Zheng et al., 2014). To develop a crash–conflict relationship using the extreme value theory, it is a prerequisite to ensure that extreme events are sufficiently smooth to enable the extrapolation from observable events to unseen events. Thus, the approach to sampling extreme events is of great importance. Typically, there are two types of sampling approaches: (1) block maxima approach using the generalized extreme value (GEV) distribution; and (2) peak over threshold (POT) approach using the generalized Pareto distribution (GPD). Previous studies have shown that for POT, choosing the threshold for extreme events is subjective and the serial-dependency issue cannot be well handled (Li et al., 2018; Zheng et al., 2014; Zheng et al., 2018). On the contrary, when using the block maxima, the serial-dependency across the observations can be accounted for during the parameter estimation procedure automatically (more details can be found in Coles (2001)). As such, this study opts for the block maxima in favor of the POT.

In the block maxima approach, observations are aggregated into fixed intervals over time, and the maxima in each interval are treated as extremes. Suppose that there is a set of independently and identically distributed random observations $\{X_1, X_2, \ldots, X_n\}$ which follow an unknown distribution function $F(x) = Pr(X_i \leq x)$, and let maximum $M_n = max(X_1, X_2, \ldots, X_n)$. When n is approaching to the infinity ($n \rightarrow \infty$), M_n will converge to a GEV distribution as shown in Eq. (3):

$$G(x) = \exp\{-[1 + \epsilon(\frac{x-\mu}{\sigma})]^{\frac{-1}{\epsilon}}\} \tag{3}$$

where μ is the location parameter, σ is the scale parameter, and ϵ is the shape parameter, and $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \epsilon < \infty$.

The tail behavior of an extreme value distribution should be focused on since the extreme value theory enables the extrapolation of observable traffic conflicts to traffic crashes that are unobservable in a short time span. To measure crash risks in discretionary lane-changing maneuvers, the gap time is adopted to measure the risk of a discretionary lane-changing event (Gettman and Head, 2003). Gap time is defined as "*the time between the entries into the conflict spot of two vehicles*" (Wang et al., 2021a). Gap time is negatively proportional to crash risks where smaller gap time values indicate higher crash risks. For each discretionary lane-changing event, only the minimum gap time is retained, reflecting the degree of danger of a discretionary lane-changing event. When a $GT \leq 0$, there will be trajectory overlaps between the lane-changing vehicle and the following vehicle, indicating the occurrence of traffic crashes. As suggested by previous studies (e.g., Zheng et al., 2014), the negated values of gap times are used to fit the GEV distribution, and a crash can be identified if negated $GT \geq 0$. The crash risk is calculated based on the tail region of the GEV distribution as follows:

$$R = \Pr(Z \geq 0) = 1 - G(0) \tag{4}$$

where R is the crash risk and also the probability of negated $GT \geq 0$, Z represents the maximum negated gap time, and $G(\cdot)$ represents the fitted GEV distribution.

When employing an extreme value theory approach, three key considerations must be properly handled, i.e., sample size, serial dependency, and non-stationarity. For the sample size issue, the minimum sample size suggested in previous literature is 30 (Zheng et al., 2014). In our study, both autonomous vehicle discretionary lane-changing ($N = 180$) and human-driven vehicle discretionary lane-changing ($N = 178$) datasets comply with this requirement. The serial dependency issue occurs when the key assumption of extreme value theory that the extreme events are independently and identically distributed is violated (e.g., a lane-changing maneuver may be dependent on a previous lane-changing maneuver). As mentioned before, the serial dependency issue can be automatically addressed using the block maxima approach. As for non-stationarity, since certain time-varying factors may affect discretionary lane-changing maneuvers and cause the heterogeneity issue, covariates are included in the extreme value theory model to mitigate non-stationarity to retrieve a set of identically distributed observations. The literature implies that those covariates should be included in the location parameter of the GEV distribution using the identity link function (Songchitruksa and Tarko, 2006). Mathematically, the location parameter is written as:

$$\mu_i = \mu_0 + \mu_1\gamma_1 \tag{5}$$

where μ_i is the location parameter for the ith block, μ_0 is the intercept term, μ_1 and γ_1 means the vectors of estimated coefficients and covariates.

4.3 Multivariate adaptive regression splines

Multivariate adaptive regression splines is a multivariate piecewise regression model (Friedman, 1991), that has been implemented to analyze lane-changing gap acceptance behaviors (e.g., Das et al., 2020; Ghasemzadeh and Ahmed, 2018). The multivariate adaptive regression splines model has some key advantages over linear regression: (1) it takes into consideration both the nonlinear impacts of individual variables and the interaction impacts among variables; (2) the results of multivariate adaptive regression splines are presented as a set of basis functions (BFs) which can mitigate the black-box issue of traditional machine learning methods; and (3) multivariate adaptive regression splines is capable of handling multicollinearity between variables (Wen et al., 2022a). Therefore, providing higher predictive accuracy and more interpretability for the naturalistic driving data using multivariate adaptive regression splines is beneficial for understanding how different variables affect lane-changing gap acceptance.

The multivariate adaptive regression splines model classifies the space of variables into multiple regions separated by knots, and then fits a spline function between these knots smoothly. The spline function consists of a series of BFs, each of which is either a main function or an interaction term between variables. The general form of the multivariate adaptive regression splines model is given in Eq. (6).

$$\hat{y} = \alpha_0 + \sum_{m=1}^{M} \alpha_m \beta_m \tag{6}$$

where $\hat{y}$ defines the predicted response variable (which is the lead or lag gap in our study), α_0 means the constant BF coefficient, M represents the total number of BFs, α_m is the coefficient of the mth BF, and β_m corresponds to the mth BF.

In this study, the multivariate adaptive regression splines model is developed to identify the relationship between independent variables and accepted gaps when the discretionary lane-changing maneuver begins. The dependent variable is the logarithm of the accepted gap at the start point of discretionary lane-changing since the logarithmic transformation ensures that the predicted gaps are always non-negative. The potential independent variables that may affect the gap acceptance behavior are identified through a thorough literature review (Balal et al., 2014; Das et al., 2020; Wang et al., 2019; Yang et al., 2019). Then the independent variables are selected based on the literature review results and data availability. The detailed variable selection results will be presented in the following.

5. Results

First, the duration of autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing events is quantified. Second, the impacts of discretionary lane-changing on the following vehicles, e.g., driving volatility and crash risks, are computed and compared between autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing. Third, gap acceptance in autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing are analyzed and modeled. Note that a $p-value$ of 0.05 is adopted as the threshold to judge the statistical significance.

5.1 Duration

As Table 1 shows, the mean values of autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing duration are nearly the same, indicating that these two discretionary lane-changing modes have similar time efficiency. This conclusion is confirmed by the results of the Mann-Whitney U test ($p-value=.5$), suggesting that the differences in duration are not statistically significant. This might be because autonomous vehicles are programmed by autonomous vehicle algorithm developers to be socially compliant, understood and accepted by surrounding human drivers. Then the lognormal distribution is found to fit the duration data the best, which is in line with previous literature (e.g., Toledo and Zohar, 2007; Venthuruthiyil and Chunchu, 2022; Wang et al., 2019). The lognormal distribution parameters of autonomous vehicle discretionary lane-changing are: $mean = 1.847$ and $variance = 0.416$. For human-driven vehicle discretionary lane-changing, these parameters are: $mean = 1.864$ and $variance = 0.486$.

Table 1. Summary statistics of discretionary lane-changing duration

discretionary lane-changing direction	**discretionary lane-changing mode**									
	autonomous vehicle discretionary lane-changing					human-driven vehicle discretionary lane-changing				
	Size	Max (s)	Min (s)	Mean (s)	Std (s)	Size	Max (s)	Min (s)	Mean (s)	Std (s)
To the left	95	10.5	3.8	6.482	1.110	97	11.4	3.4	6.546	1.545
To the right	85	9.7	4.1	6.380	1.110	81	12.5	4	6.739	1.669
Total	180	10.5	3.8	6.434	1.108	178	12.5	3.4	6.634	1.600

Note: *Std*: standard deviation

Table 2. Comparison of driving volatility of following vehicles between autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing

Metrics	**lane-changing mode**								**Difference (%)**
	autonomous vehicle discretionary lane-changing ($n = 180$)				human-driven vehicle discretionary lane-changing ($n = 178$)				
	Max	Min	Mean	Std	Max	Min	Mean	Std	
				Speed volatility					
$Std(m/s)$	2.115	0.057	0.854	0.488	4.792	0.067	1.049	0.733	**-18.59%**
$D_{mean}(m/s)$	1.865	0.043	0.739	0.439	4.396	0.056	0.913	0.656	**-19.06%**
				Acceleration volatility					
$Std(m/s^2)$	0.807	0.015	0.376	0.185	1.305	0.045	0.389	0.250	**-3.34%**
$D_{mean}(m/s^2)$	0.709	0.013	0.327	0.164	1.218	0.032	0.338	0.226	**-3.25%**
				Yaw rate volatility					
$Std(degree/s)$	3.549	0.455	1.001	0.519	6.035	0.366	1.159	0.800	**-13.63%**
$D_{mean}(\text{degree}/s)$	2.477	0.348	0.758	0.373	4.665	0.299	0.894	0.607	**-15.21%**

Note: *Std*: standard deviation; D_{mean}: mean absolute deviation.

5.2 Impacts on the following vehicle

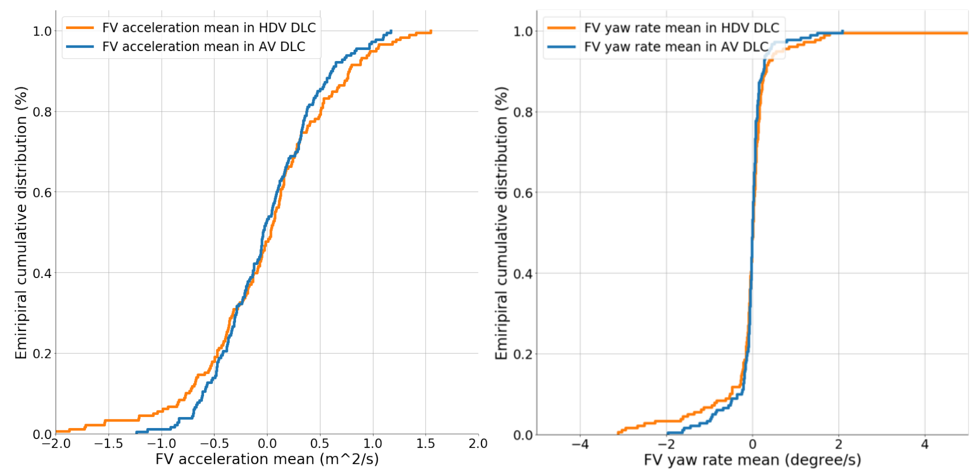
5.2.1 Driving volatility analysis

Table 2 shows the summary statistics of driving volatility of following vehicles in different discretionary lane-changing modes. It is noteworthy to mention that only the trajectory data collected within the discretionary lane-changing period is used in the computation process. The presented values are the aggregation and average of driving volatility of following vehicles. The column named "Difference (%)" represents the mean value changes in driving volatility of autonomous vehicle discretionary lane-changing concerning human-driven vehicle discretionary lane-changing. The positive (negative) values of "Difference (%)" represent the increase (decrease) in driving volatility of autonomous vehicle discretionary lane-changing relative to human-driven vehicle discretionary lane-changing.

The results shown in Table 2 indicate that the following vehicles are inclined to show lower speed volatility in autonomous vehicle discretionary lane-changing compared to human-driven vehicle discretionary lane-changing. The percentage changes of standard deviation (Std) and mean absolute deviation (D_{mean}) of speed for autonomous vehicle discretionary lane-changing are -18.59% and -19.06%, respectively. Both driving volatility measures are found to be significantly different between the two discretionary lane-changing modes ($Std: p-value = .046$; $D_{mean}: p-value = .04$) by the Mann-Whitney U test. It may be explained by the precise motion control module of autonomous vehicles which enables them to handle complex driving scenarios. The higher speed volatility for human-driven vehicle discretionary lane-changing is expected due to the stochastic behaviors of human-driven vehicles. The comparisons of speed volatility suggest that the penetration of autonomous vehicles in mixed traffic can potentially improve the driving smoothness of following vehicles.

As shown in Table 2, 3.34% and 3.25% reductions are found in the Std and D_{mean} of acceleration. The mean values of acceleration of following vehicles are displayed in Figure 3(a). Although the Mann-Whitney U test shows no significant differences in acceleration volatility ($Std\colon p - value = .5$; $D_{mean}\colon p - value = .5$), one can observe in Figure 3(a) that following vehicles in human-driven vehicle discretionary lane-changing are more likely to perform harsh acceleration and deceleration., indicating that autonomous vehicle discretionary lane-changing induces lower acceleration rates of following vehicles, and leads to better driving comfort.

The yaw rate describes the angular speed of the forward direction of the vehicle, which plays a crucial role in vehicle lateral dynamics (Aripin et al., 2014). It can be used to detect evasive actions of following vehicles where high yaw rates are significantly correlated with swerving maneuvers of following vehicles during the discretionary lane-changing event (Guo et al., 2018). From Table 2, one can observe that following vehicles in autonomous vehicle discretionary lane-changing events have smaller Std and D_{mean} of yaw rates than those in human-driven vehicle discretionary lane-changing events. The differences are statistically significant based on the Mann-Whitney U test ($Std\colon p - value = .04$; $D_{mean}\colon p - value = .02$). The empirical cumulative distributions of the mean values of the yaw rate are depicted in Figure 3(b). To summarize, following vehicles in autonomous vehicle discretionary lane-changing have smaller and more stable yaw rates and therefore more lateral stability compared to following vehicles in human-driven vehicle discretionary lane-changing.



(a) (b)

Figure 3. Empirical cumulative distributions of (a) acceleration mean and (b) yaw rate mean

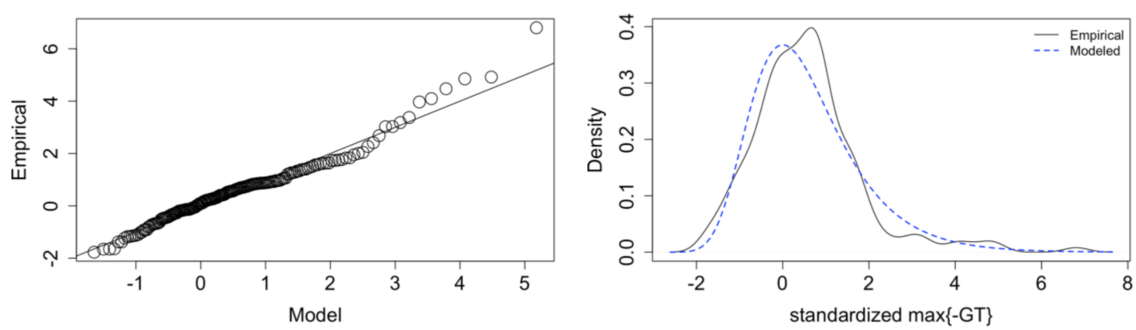
5.2.2 Crash risk analysis

Note that since the research scope of this study is to understand the effects of discretionary lane-changing behaviors of autonomous vehicles on following vehicles, only traffic conflicts between lane-changing vehicles and following vehicles are analyzed. As suggested by Farah and Azevedo (2017), each block represents a discretionary lane-changing event where the duration of the block is the same as the duration of the corresponding discretionary lane-changing event. For each block, the minimum value of gap time is chosen and used to develop the block maxima model. Former works have concluded that only the gap time value lower than $3s$ should be treated as an extreme event (Saul et al., 2021). Therefore, the gap time values above $3s$ are filtered, resulting in 177 maxima for autonomous vehicle discretionary lane-changing and 173 maxima for human-driven vehicle discretionary lane-changing.

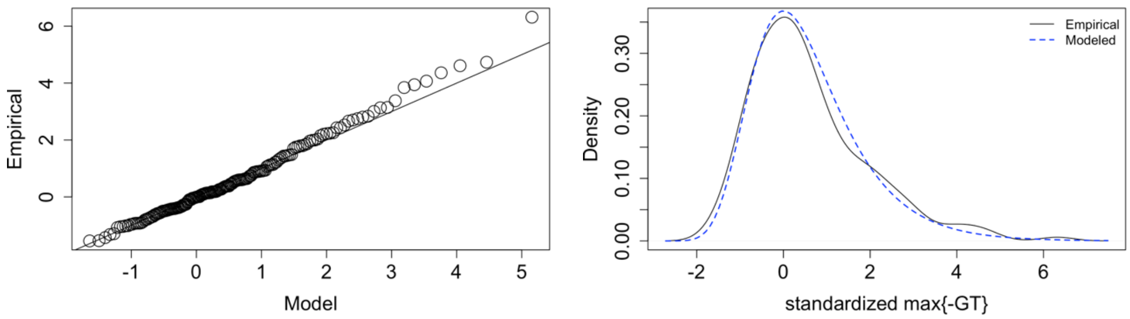
Table 3 presents the stationary and selected non-stationary block maxima models which are built using the maximum likelihood estimation (MLE) method. Two covariates are included in each non-stationary model: $lag\_spacing$ representing the distance (in meters) between the lane-changing vehicle and the following vehicle at the start point of discretionary lane-changing and $relspd\_mean\_lcv\_fv$ representing the average relative speed between the lane-changing vehicle and the following vehicle during the discretionary lane-changing period. It can be observed that incorporating the covariates into the location parameter can greatly reduce the negative log-likelihood and thus improve the model fit. Figure 4 shows the simulated quantile-quantile (Q-Q) plot and the probability density function of the empirical and modeled standardized maximum negated gap time derived from the non-stationary block maxima models. For both discretionary lane-changing modes, a visual inspection shows a good fit as both the empirical and the modeled GEV curves are inclined to overlap each other. Further, a Kolmogorov–Smirnov (K-S) test is implemented, of which the null hypothesis is that the sample is drawn from the fitted GEV distribution. In both conditions, $p - values$ are significantly greater than 0.05 ($autonomous\ vehicle\ discretionary\ lane - changing: p - value = .9$; $human - driven\ vehicle\ discretionary\ lane - changing: p - value = .99$), meaning that the null hypothesis cannot be rejected.

The interpretation of the estimation results of non-stationary block maxima models in Table 3 is straightforward. First, the negative sign of $\mu_{lag\_spacing}$ indicates that as the longitudinal spacing between the lane-changing vehicle and the following vehicle increases, the negated gap time will decrease, and the value of gap time will increase, which agrees with previous literature such as Ali et al. (2022a). This is reasonable because if the following vehicle is far away from the lane-changing vehicle, the lane-changing vehicle can have abundant space to perform the discretionary lane-changing behavior. Second, when the average relative speed between the lane-changing vehicle and the following vehicle ($\mu_{relspd\_mean\_lcv\_fv}$) is larger, the negated gap time decreases, and the gap time increases. The explanation is also intuitive: when the lane-changing vehicle is significantly faster than the following vehicle, the lane-changing vehicle takes less time to change to the target lane, getting further away from the following vehicle, which results in higher gap time values.

The crash risk representing the probability of a collision during an autonomous vehicle discretionary lane-changing or human-driven vehicle discretionary lane-changing is calculated using Eq. (4). According to Zheng et al. (2014), the confidence intervals of crash risks are generated based on simulation where estimated parameters are assumed to follow the normal distribution. After 10^6 simulation runs, the empirical distributions of estimation are attained and the lower and upper confidence interval bounds are calculated from the quantiles of the distributions. Based on the non-stationary block maxima model, the crash risk is computed as 0.010 with a 95% confidence interval (0.001,0.039) for autonomous vehicle discretionary lane-changing and 0.020 with a 95% confidence interval (0.003,0.078) for human-driven vehicle discretionary lane-changing. In summary, the discretionary lane-changing maneuvers of autonomous vehicles have been found to improve traffic safety significantly compared to those of human-driven vehicles, with a 2 times reduction in crash risk. This finding indicates the efficacy and potential of autonomous vehicles in eliminating discretionary lane-changing crash risks.



(a) autonomous vehicle discretionary lane-changing



(b) human-driven vehicle discretionary lane-changing

Figure 4. QQ-plot (left) and probability density function (right) for the non-stationary block maxima model

Table 3. Block maxima model estimation results

discretionary lane-changing mode	Model type	nllh	Location$(\mu)(standard\ error, SE)$			Scale$(\sigma)(SE)$	Shape$(\epsilon)(SE)$	AIC	BIC
			μ_0	$\mu_{lag\_spacing}$	$\mu_{relspd\_mean\_lcv\_fv}$				
autonomous vehicle	Stationary	203.655	-1.391(NA)	--	--	1.612(NA)	-1.185(NA)	413.310	422.839
discretionary lane-changing	Non-stationary	79.033	-0.242(0.069)	-0.043(0.003)	-0.338(0.027)	0.395(0.022)	-0.334(0.037)	168.065	183.946
human-driven vehicle	Stationary	143.023	-0.981(0.058)	--	--	0.690(0.052)	-0.679(0.075)	292.047	301.507
discretionary lane-changing	Non-stationary	125.490	-0.739(0.052)	-0.009(0)	-0.126(0.009)	0.637(0.045)	-0.707(0.053)	260.979	276.746

Note: nllh: negative log-likelihood; NA: indicates that the standard error of the corresponding parameter does not exist.

5.3 Gap acceptance

Through the analysis of driving volatility and crash risks, it can be found that autonomous vehicle discretionary lane-changing has significantly different effects on the following vehicle in comparison to human-driven vehicle discretionary lane-changing. Thus, by analyzing gap acceptance behaviors using the multivariate adaptive regression splines model, this part is aimed at exploring whether the discretionary lane-changing decision-making mechanisms of autonomous vehicles and human-driven vehicles are different.

Table 4. Summary statistics of gaps in discretionary lane-changing

Type	discretionary lane-changing mode									
	autonomous vehicle discretionary lane-changing					human-driven vehicle discretionary lane-changing				
	Size	Max	Min	Mean	Std	Size	Max	Min	Mean	Std
$Lead\ gap(s)$	153	5.975	0.236	1.703	0.981	154	6.175	0.227	1.506	1.202
$Lag\ gap(s)$	180	4.663	0.372	1.576	0.854	178	6.101	0.212	1.559	1.070

5.3.1 Gap acceptance characteristics

Before performing discretionary lane-changing, a driver usually estimates the gaps between the lane-changing vehicle and both the lead vehicle and the following vehicle, i.e., the lead gap and lag gap. As mentioned previously, the lead gap and lag gap are defined based on time, which refers to the time taken to traverse the longitudinal distance between the lane-changing vehicle and lead vehicle (for the lead gap) and between the lane-changing vehicle and the following vehicle (for the lag gap) when a discretionary lane-changing event starts. Table 4 shows the descriptive statistics of the lead and lag gap for different discretionary lane-changing modes. The Mann-Whitney U test is implemented to detect if there are any statistically significant differences in the lead or lag gap between autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing. The results reveal that the lead gap for autonomous vehicle discretionary lane-changing is significantly larger than that for human-driven vehicle discretionary lane-changing at the 95% confidence level ($p - value = .001$), while the lag gap comparison is not statistically significant ($p - value = .3$). It is suggested that compared to human-driven vehicles, autonomous vehicles need significantly larger lead gaps to initiate the discretionary lane-changing maneuvers, meaning that autonomous vehicles adopt a more conservative driving style.

Table 5. Summary statistics of independent variables in multivariate adaptive regression splines

Variables	Description	autonomous vehicle discretionary lane-changing		human-driven vehicle discretionary lane-changing	
		Mean	Std	Mean	Std
$spd\_start\_lv(m/s)$	The instantaneous speed of lead vehicle when the discretionary lane-changing starts	14.558	4.572	12.731	4.148
$spd\_start\_lcv(m/s)$	The instantaneous speed of lane-changing vehicle when the discretionary lane-changing starts	13.338	4.585	11.778	4.905
$spd\_start\_fv(m/s)$	The instantaneous speed of following vehicle when the discretionary lane-changing starts	13.319	4.957	11.699	4.563
$acc\_start\_lv(m/s^2)$	The instantaneous acceleration of lead vehicle when the discretionary lane-changing starts	0.235	0.498	0.308	0.755
$acc\_start\_lcv(m/s^2)$	The instantaneous acceleration of lane-changing vehicle when the discretionary lane-changing starts	0.210	0.645	0.266	0.806
$acc\_start\_fv(m/s^2)$	The instantaneous acceleration of following vehicle when the discretionary lane-changing starts	0.413	0.603	0.406	0.685
$relspd\_lcv\_lv(m/s)$	The instantaneous relative speed between lane-changing vehicle and lead vehicle when the discretionary lane-changing starts	-1.219	2.030	-0.952	2.643
$relspd\_lcv\_fv(m/s)$	The instantaneous relative speed between lane-changing vehicle and following vehicle when the discretionary lane-changing starts	0.019	2.003	0.079	2.554

Table 6. Results of the multivariate adaptive regression splines model for lead gap in autonomous vehicle discretionary lane-changing

BF	Basis function	Coefficient
Intercept	Intercept	-1.241
BF1	$Max(spd\_start\_lcv - 8.927,0)$	0.167
BF2	$Max(spd\_start\_lv - 21.483,0) * spd\_start\_fv$	-0.030
BF3	$Max(spd\_start\_lcv - 17.602,0) * spd\_start\_fv$	0.018
BF4	$Max(0.650 - relspd\_lcv\_lv, 0)$	0.133
BF5	$Max(10.879 - spd\_start\_fv, 0)$	2.389
BF6	$Max(10.330 - spd\_start\_fv, 0)$	-1.749
BF7	$Max(11.855 - spd\_start\_fv, 0)$	-2.064
BF8	$Max(spd\_start\_fv - 12.397,0)$	-0.152
BF9	$Max(12.397 - spd\_start\_fv, 0)$	1.521
BF10	$Max(20.468 - spd\_start\_lv, 0) * Max(spd\_start\_fv - 10.879,0)$	0.038
BF11	$acc\_start\_lcv * BF8$	0.100
MAE	0.302	
RMSE	0.386	

5.3.2 Gap acceptance modeling results

Note that only the lead gap acceptance behaviors for autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing are found to be significantly different. Given that the objective of this study is to compare autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing maneuvers, only the lead gap acceptance models are built. Table 5 presents the summary statistics of independent variables included in multivariate adaptive regression splines to model lead gap acceptance in autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing scenarios. Table 6 and Table 7 present the results of the multivariate adaptive regression splines models for lead gap acceptance in autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing, respectively. As indicated by the MAE (mean absolute error) and RMSE (root mean square error), both multivariate adaptive regression splines models fit the lead gap data well.

Table 7. Results of the multivariate adaptive regression splines model for lead gap in human-driven vehicle discretionary lane-changing

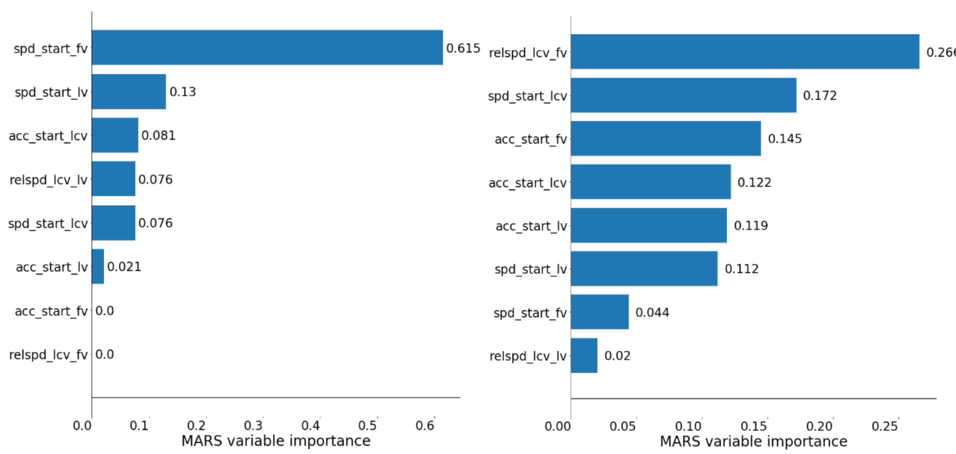
BF	Basis function	Coefficient
Intercept	Intercept	-1.049
BF1	$Max(relspd\_lcv\_fv + 2.565,0)$	0.188
BF2	$Max(10.689 - spd\_start\_lcv, 0)$	0.148
BF3	$Max(acc\_start\_fv - 0.898,0)$	2.188
BF4	$Max(0.898 - acc\_start\_fv, 0)$	0.621
BF5	acc_start_lv	-0.245
BF6	$spd\_start\_lv * acc\_start\_lcv$	0.030
BF7	$acc\_start\_fv * BF4$	0.289
BF8	$Max(10.156 - spd\_start\_fv, 0) * BF3$	0.371
BF9	$Max(18.323 - spd\_start\_lv, 0) * BF3$	-0.336
MAE	0.344	
RMSE	0.448	

To extract the total effect of a variable on lead gap acceptance, one should consider all main effects and interaction terms involving this variable of interest. For the variable spd_start_lcv, its total effect in autonomous vehicle discretionary lane-changing is written in Eq. (7) (see BF1 and BF3 in Table 6), which is referred to as "Impact Impression".

$$0.167 * BF1 + 0.018 * BF3$$
$$= 0.167 * Max(spd\_start\_lcv - 8.927,0) + 0.018 * Max(spd\_start\_lcv - 17.602,0) * spd\_start\_fv \quad (7)$$

"Rate of Impression Expression" is defined as the first derivative of Impact Impression. In the case of spd_start_lcv, Rate of Impression Expression can be determined in three different intervals. When spd_start_lcv is less than $8.927 m/s$, spd_start_lcv has no impact on lead gap acceptance. If spd_start_lcv is between $8.927m/s$ and $17.602m/s$, Rate of Impression Expression will be 0.167 suggesting that the accepted lead gap for autonomous vehicles is linearly correlated with spd_start_lcv. As spd_start_lcv becomes greater than $17.602m/s$, Rate of Impression Expression will be $0.167 + 0.018 * spd\_start\_fv$ which indicates that spd_start_fv has a positive effect on lead gap acceptance for autonomous vehicles driving at high speed. These findings demonstrate the nonlinear and complex impacts of spd_start_lcv on lead gap acceptance for autonomous vehicle discretionary lane-changing. When it comes to the human-driven vehicle discretionary lane-changing scenario, the effects of spd_start_lcv are relatively simple (see BF2 in Table 7). If spd_start_lcv is less than 10.689, Rate of Impression Expression is -0.148, indicating that the higher the speed of the human-driven vehicle is, the less the lead gap the human-driven vehicle needs.

Figure 5 shows the variable importance of multivariate adaptive regression splines for autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing, respectively. One can observe that spd_start_fv and spd_start_lv contribute the most to the multivariate adaptive regression splines model outcomes for autonomous vehicle discretionary lane-changing while these two are among the least important variables for human-driven vehicle discretionary lane-changing. Interestingly, $relspd\_lcv\_fv$ turns out to have the most significant impact on human-driven vehicle lead gap acceptance but it has an importance value of zero for autonomous vehicle discretionary lane-changing. The differences in variable importance between autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing might be attributed to the fact that autonomous vehicles' discretionary lane-changing decision-making process has been programmed by autonomous vehicle manufacturers. For example, the autonomous vehicle might be programmed to conduct a discretionary lane-changing maneuver based on the time-based lag gap and the threshold for the time-based lag gap may vary with the speed of the following vehicle for autonomous vehicle discretionary lane-changing, which may explain why *spd_start_fv* is the most important for autonomous vehicle discretionary lane-changing. However, this needs to be further validated. While human-driven vehicles may not be good at judging the time-based lag gap and thus prefer to make discretionary lane-changing decisions mainly based on *relsped_lcv_fv*. Overall, different variable importance sequences indicate that autonomous vehicles and human-driven vehicles adopt different decision-making logics when conducting discretionary lane-changing maneuvers.



(a) autonomous vehicle discretionary lane-changing

(b) human-driven vehicle discretionary lane-changing

Figure 5. Variable importance in multivariate adaptive regression splines models

6. Discussions and conclusions

We take the first attempt to investigate the characteristics of autonomous vehicle discretionary lane-changing and its effects on following vehicles in the target lane using the real-world autonomous driving dataset. 180 autonomous vehicle discretionary lane-changing and 178 human-driven vehicle discretionary lane-changing events are extracted from the Waymo Open

Dataset. Driving volatility is computed and compared to understand the traffic effects of autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing events. To quantify the safety impacts of autonomous vehicle discretionary lane-changing and compare them to human-driven vehicle discretionary lane-changing, using the gap time as the safety indicator, the block maxima approach is employed to calculate the crash risks from the observed traffic conflicts. Additionally, the gap acceptance behaviors of autonomous vehicles and human-driven vehicles are modeled and compared using multivariate adaptive regression splines.

The findings reveal that autonomous vehicles and human-driven vehicles take similar time to conduct the discretionary lane-changing maneuvers, indicating similar time efficiency of autonomous vehicles and human-driven vehicles. When it comes to driving volatility, if an autonomous vehicle changes to the target lane, the following vehicle in the target lane shows significantly lower speed and yaw rate volatility, indicating more longitudinal and lateral stability. Although the acceleration volatility reductions in autonomous vehicle discretionary lane-changing are not statistically significant, autonomous vehicle discretionary lane-changing leads to smaller acceleration rates of following vehicles, potentially improving the comfort level of the drivers and passengers in following vehicles. In summary, the insertion of autonomous vehicles into the traffic stream will have benefits in various aspects. The reasons may be that autonomous vehicles with better speed management and route planning tend to have fewer speed and yaw angle fluctuations, and are more conservative in discretionary lane-changing strategy, and thus impose less interference on the following vehicles. It is reasonable to foresee that these benefits will become even more significant as the market penetration rate of autonomous vehicles promotes in the future.

Using the developed block maxima models, we identify that when autonomous vehicles are performing discretionary lane-changing behaviors, crash risks are significantly lower compared to that of human-driven vehicle discretionary lane-changing. This may be attributed to autonomous vehicles' better ability to sense their surrounding environments and thus plan their routes, which can minimize the uncertainty during the discretionary lane-changing decision-making process. For human-driven vehicle discretionary lane-changing, higher crash risk is observed since most of the crashes happen due to the human driver's risky behaviors caused by uncertainty (Ali et al., 2022b). This finding is consistent with field experiment-based studies by Wang et al. (2021b), in which the discretionary lane-changing behaviors of autonomous vehicles were found to improve traffic safety.

We then model and compare the gap acceptance behaviors of autonomous vehicles and human-driven vehicles to understand the discretionary lane-changing decision-making process. The results show that autonomous vehicles adopt significantly larger lead gaps while there are no significant differences in accepted lag gaps. This can be explained as that autonomous vehicles are programmed to drive conservatively to avoid potential collisions with lead vehicles in the target lane. The multivariate adaptive regression splines models have been developed to explore the relationship between lead gap acceptance and a variety of variables in autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing scenarios. The modeling results indicate the nonlinear impacts of spd_start_lcv on the accepted lead gaps for autonomous vehicle discretionary lane-changing. The variable importance of multivariate adaptive regression splines models is also significantly different between autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing suggesting different decision-making processes. For example, autonomous vehicles allocate more weights to the speed of the following vehicle when deciding the start of discretionary lane-changing, while the weights of different measures are more evenly distributed for human-driven vehicles. Namely, autonomous vehicles pay more attention to whether their discretionary lane-changing maneuvers affect the speed of the following vehicles. These different decision-making processes may explain the differences in speed, acceleration and yaw rate volatility of following vehicles in autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing events. However, further research is still needed to reveal the underlying relationship between the weights of measures and the impacts on the behaviors of following vehicles.

According to the volatility-related indices and block maxima results, autonomous vehicles exhibit better performance than human-driven vehicles in terms of discretionary lane-changing behaviors on public roads. These results, along with the multivariate adaptive regression splines modeling outcomes, have practical implications for companies developing autonomous vehicle technology. Specifically, this study shows how discretionary lane-changing maneuvers of autonomous vehicles affect surrounding human drivers. This information can be used to evaluate and enhance autonomous vehicle controllers by considering the responses of human drivers in real-world scenarios. This is particularly useful as autonomous vehicles are currently being tested on public roads. One example is that there is no significant difference in the acceleration volatility of the following vehicle between autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing. Additionally, it is found that in 42.78% of autonomous vehicle discretionary lane-changing events, absolute values of the speed percentage change of following vehicles within the discretionary lane-changing period are greater than 20%, indicating that following vehicles are negatively impacted. Therefore, developers of the autonomous vehicle algorithms could improve control modules to ensure that following vehicles take lower acceleration rates and smoother speeds.

One question is: given that Waymo might be selective when producing this dataset, how representative are those 20-second segments used in this study? Based on our previous studies (e.g., Wen et al., 2022b) and other papers (e.g., Hu et al., 2023) using real-world autonomous driving datasets, one can find that at the early adoption stage, autonomous vehicles behave in a conservative way to ensure their safety. For example, Hu et al. (2023) found that the autonomous vehicle-following-human-driven vehicle has the largest jam spacing and critical spacing, indicating that the Waymo autonomous vehicle is more conservative than the human-driven vehicle. This conclusion is consistent with the findings in this paper, e.g., the accepted lead gap for autonomous vehicle discretionary lane-changing is significantly larger than that for human-driven vehicle discretionary lane-changing. Since the findings of this paper are supported by other studies, it is reasonable to believe that these short clips in the Waymo Open Dataset are somewhat representative. However, this conclusion needs to be verified using other autonomous driving datasets, such as Lyft and nuScenes.

On the one hand, the sample size used in our study is relatively small, which may affect the analysis results adversely. On the other hand, autonomous vehicle is relatively new to most surrounding human drivers. It is questionable whether the benefits of autonomous vehicles will continue to exist when most human drivers become familiar with autonomous vehicles. Thus, future research is needed to collect more real-world autonomous vehicle discretionary lane-changing and human-driven vehicle discretionary lane-changing data and confirm whether the analysis results remain the same or vary. Second, it should be noted that there is only one autonomous vehicle in each short clip. With the increasing prevalence of autonomous vehicles, the interactions between autonomous vehicles would also be an interesting topic, especially before the autonomous vehicles are fully connected. Future research can explore this topic using traffic simulation or by analyzing future mixed traffic data. Lastly, this study only focuses on accepted discretionary lane-changing behaviors due to data scarcity. Earlier studies have shown that failed lane-changing attempts and mandatory lane-changing maneuvers are fundamentally different from accepted discretionary lane-changing behaviors (Ali et al., 2020; Ali et al., 2022a). It would be interesting to examine the failed lane-changing attempts and mandatory lane-changing behaviors of autonomous vehicles and human-driven vehicles in various traffic settings.

Acknowledgement

This research is supported by the National Natural Science Foundation of China (No. 52202425), and Guangzhou Municipal Science and Technology Project (No. 2023A03J0011).

References

Ali, Y., M. M. Haque, and Z. Zheng. 2022a. "An Extreme Value Theory approach to estimate crash risk during mandatory lane-changing in a connected environment." Analytic methods in accident research 33: 100193.

Ali, Y., M. M. Haque, and Z. Zheng. 2022b. "Assessing a Connected Environment's Safety Impact During Mandatory Lane-Changing: A Block Maxima Approach." IEEE Transactions on Intelligent Transportation Systems 24(6): 6639-6649.

Ali, Y., Z. Zheng, M. M. Haque, M. Yildirimoglu, and S. Washington. 2020. "Detecting, analysing, and modelling failed lane-changing attempts in traditional and connected environments." Analytic methods in accident research 28: 100138.

Aripin, M. K., Y. M. Sam, K. A. Danapalasingam, K. Peng, N. Hamzah, and M. F. Ismail. 2014. "A review of active yaw control system for vehicle handling and stability enhancement." International journal of vehicular technology 2014.

Arvin, R., M. Kamrani, and A. J. Khattak. 2019. "How instantaneous driving behavior contributes to crashes at intersections: extracting useful information from connected vehicle message data." Accident Analysis & Prevention 127: 118-133.

Balal, E., R. L. Cheu, T. Gyan-Sarkodie, and J. Miramontes. 2014. "Analysis of discretionary lane changing parameters on freeways." International Journal of Transportation Science and Technology 3(3): 277-296.

Bham, G. H., 2009. "Estimating driver mandatory lane change behavior on a multilane freeway." Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C.

Coles, S. 2001. An Introduction to Statistical Modeling of Extreme Values. London, UK: Springer-Verlag.

Das, A., M. N. Khan, and M. M. Ahmed. 2020. "Nonparametric multivariate adaptive regression splines models for investigating lane-changing gap acceptance behavior utilizing strategic highway research program 2 naturalistic driving data." Transportation research record 2674(5): 223-238.

Di, X., and R. Shi. 2021. "A survey on autonomous vehicle control in the era of mixed-autonomy: From physics-based to AI-guided driving policy learning." Transportation research part C: emerging technologies 125: 103008.

Dixit, V., Z. Xiong, S. Jian, and N. Saxena. 2019. "Risk of automated driving: Implications on safety acceptability and productivity." Accident Analysis & Prevention 125: 257-266.

Dong, C., H. Wang, Y. Li, X. Shi, D. Ni, and W. Wang. 2021. "Application of machine learning algorithms in lane-changing model for intelligent vehicles exiting to off-ramp." Transportmetrica A: transport science 17(1): 124-150.

Ettinger, S., S. Cheng, B. Caine, C. Liu, H. Zhao, S. Pradhan, Y. Chai et al. 2021. "Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset." Proceedings of the IEEE/CVF International Conference on Computer Vision, 9710-9719.

Farah, H., and C. L. Azevedo. 2017. "Safety analysis of passing maneuvers using extreme value theory." IATSS research 41(1): 12-21.

Friedman, J. H. 1991. "Multivariate adaptive regression splines." The annals of statistics 19(1): 1-67.

Gettman, D., and L. Head. 2003. "Surrogate safety measures from traffic simulation models." Transportation Research Record 1840(1): 104-115.

Ghasemzadeh, A., and M. M. Ahmed. 2018. "Utilizing naturalistic driving data for in-depth analysis of driver lane-keeping behavior in rain: Non-parametric MARS and parametric logistic regression modeling approaches." Transportation research part C: emerging technologies 90: 379-392.

Guo, Y., T. Sayed, and M. H. Zaki. 2018. "Exploring evasive action–based indicators for PTW conflicts in shared traffic facility environments." Journal of Transportation Engineering, Part A: Systems 144(11): 04018065.

Hu, X., Z. Zheng, D. Chen, X. Zhang, and J. Sun. 2022. "Processing, assessing, and enhancing the Waymo autonomous vehicle open dataset for driving behavior research." Transportation Research Part C: Emerging Technologies 134: 103490.

Hu, X., Z. Zheng, D. Chen, and J. Sun. 2023. "Autonomous vehicle's impact on traffic: empirical evidence from Waymo Open Dataset and implications from modelling." IEEE Transactions on Intelligent Transportation Systems 24(6): 6711–6724.

Ji, A., and D. Levinson. 2020. "A review of game theory models of lane changing." Transportmetrica A: transport science 16(3): 1628-1647.

Jiang, L., Y. Xie, N. G. Evans, X. Wen, T. Li, and D. Chen. 2022. "Reinforcement Learning based cooperative longitudinal control for reducing traffic oscillations and improving platoon stability." Transportation Research Part C: Emerging Technologies 141: 103744.

Jiang, L., Y. Xie, X. Wen, D. Chen, T. Li, and N. G. Evans. 2021. "Dampen the stop-and-go traffic with connected and automated vehicles–a deep reinforcement learning approach." Proceedings of 7th International Conference on Models and Technologies for Intelligent Transportation Systems, 1-6.

Jin, C. J., V. L. Knoop, D. Li, L. Y. Meng, and H. Wang. 2019. "Discretionary lane-changing behavior: empirical validation for one realistic rule-based model." Transportmetrica A: transport science 15 (2): 244-262.

Li, G., Z. Yang, Y. Pan, and J. Ma. 2023. "Analysing and modelling of discretionary lane change duration considering driver heterogeneity." Transportmetrica B: Transport Dynamics 11(1): 343-360.

Mahdinia, I., A. Mohammadnazar, R. Arvin, and A. J. Khattak. 2021. "Integration of automated vehicles in mixed traffic: Evaluating changes in performance of following human-driven vehicles." Accident Analysis & Prevention 152: 106006.

Mauch, M., and M. J. Cassidy. 2002. "Freeway Traffic Oscillations: Observations and Predictions." Transportation and Traffic Theory in the 21st Century: Proceedings of the

15th International Symposium on Transportation and Traffic Theory, Adelaide, Australia, July 16–18, 2002, 653–673. Emerald Group Publishing Limited.

Papadoulis, A., M. Quddus, and M. Imprialou. 2019. "Evaluating the safety impact of connected and autonomous vehicles on motorways." Accident Analysis & Prevention 124: 12-22.

Rahmati, Y., M. K. Hosseini, A. Talebpour, B. Swain, and C. Nelson. 2019. "Influence of autonomous vehicles on car-following behavior of human drivers." Transportation research record 2673(12): 367-379.

Saul, H., M. Junghans, M. Dotzauer, and K. Gimm. 2021. "Online risk estimation of critical and non-critical interactions between right-turning motorists and crossing cyclists by a decision tree." Accident Analysis & Prevention 163: 106449.

Sinha, A., S. Chand, K. P. Wijayaratna, N. Virdi, and V. Dixit. 2020. "Comprehensive safety assessment in mixed fleets with connected and automated vehicles: A crash severity and rate evaluation of conventional vehicles." Accident Analysis & Prevention 142: 105567.

Songchitruksa, P., and A. P. Tarko. 2006. "The extreme value theory approach to safety estimation." Accident Analysis & Prevention 38(4): 811-822.

Sultan, B., M. Brackstone, B. Waterson, and E. R. Boer. 2002. "Modeling the dynamic cut-in situation." Transportation research record 1803(1): 45-51.

Toledo, T., H. N. Koutsopoulos, and M. E. Ben-Akiva. 2003. "Modeling integrated lane-changing behavior." Transportation Research Record 1857(1): 30-38.

Toledo, T., and D. Zohar. 2007. "Modeling duration of lane changes." Transportation Research Record 1999(1): 71-78.

Toledo, T., C. F. Choudhury, and M. E. Ben-Akiva. 2005. "Lane-changing model with explicit target lane choice." Transportation Research Record 1934(1): 157-165.

Venthuruthiyil, S. P., and M. Chunchu. 2022. "Interrupted and uninterrupted lane changes: a microscopic outlook of lane-changing dynamics." Transportmetrica A: transport science 18(3): 1679-1698.

Wang, C., Y. Xie, H. Huang, and P. Liu. 2021a. "A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling." Accident Analysis & Prevention 157: 106157.

Wang, C., C. Xu, J. Xia, Z. Qian, and L. Lu. 2018. "A combined use of microscopic traffic simulation and extreme value methods for traffic safety evaluation." Transportation Research Part C: Emerging Technologies 90: 281-291.

Wang, X., M. Yang, and D. Hurwitz. 2019. "Analysis of cut-in behavior based on naturalistic driving data." Accident Analysis & Prevention 124: 127-137.

Wang, Z., X. Zhao, Z. Xu, X. Li, and X. Qu. 2021b. "Modeling and field experiments on autonomous vehicle lane changing with surrounding human-driven vehicles." Computer-Aided Civil and Infrastructure Engineering 36(7): 877-889.

Wen, X., S. Jian, and D. He. 2023. Modeling the Effects of Autonomous Vehicles on Human Driver Car-Following Behaviors using Inverse Reinforcement Learning. IEEE Transactions on Intelligent Transportation Systems.

Wen, X., Y. Xie, L. Jiang, Y. Li, and T. Ge. 2022a. "On the interpretability of machine learning methods in crash frequency modeling and crash modification factor development." Accident Analysis & Prevention 168: 106617.

Wen, X., Z. Cui, and S. Jian. 2022b. "Characterizing car-following behaviors of human drivers when following automated vehicles using the real-world dataset." Accident Analysis & Prevention 172: 106689.

Yang, M., X. Wang, and M. Quddus. 2019. "Examining lane change gap acceptance, duration and impact using naturalistic driving data." Transportation research part C: emerging technologies 104: 317-331.

Yang, Q., and H. N. Koutsopoulos. 1996. "A microscopic traffic simulator for evaluation of dynamic traffic management systems." Transportation Research Part C: Emerging Technologies 4(3): 113-129.

Zhao, X., Z. Wang, Z. Xu, Y. Wang, X. Li, and X. Qu. 2020. "Field experiments on longitudinal characteristics of human driver behavior following an autonomous vehicle." Transportation research part C: emerging technologies 114: 205-224.

Zheng, F., C. Liu, X. Liu, S. E. Jabari, and L. Lu. 2020. "Analyzing the impact of automated vehicles on uncertainty and stability of the mixed traffic flow." Transportation research part C: emerging technologies 112: 203-219.

Zheng, L., K. Ismail, and X. Meng. 2014. "Freeway safety estimation using extreme value theory approaches: A comparative study." Accident Analysis & Prevention 62: 32-41.

Zheng, L., T. Sayed, and A. Tageldin. 2018. "Before-after safety analysis using extreme value theory: a case of left-turn bay extension." Accident Analysis & Prevention 121: 258-267.