Analytic Methods in Accident Research 46 (2025) 100388

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

# Autonomous vehicle lane-changing dynamics and impact on the immediate follower

Yasir Ali

*School of Architecture, Building, and Civil Engineering, Loughborough University, Leicestershire LE11 3TU, United Kingdom*

A R T I C L E I N F O

*Keywords:*
Autonomous vehicle
Lane-changing impact
Random parameters linear regression
Extreme value theory

A B S T R A C T

Understanding and modelling lane-changing behaviour are critical aspects of microscopic traffic flow modelling, safety analyses, and microsimulation due to their significant impact on traffic flow characteristics and safety. Among the three aspects of lane-changing behaviour—decision-making, execution, and impact—the lane-changing impact has been comparatively underexplored in the literature, which is disproportionate to its importance. A lack of proper understanding of lane-changing impact may lead to inaccurate planning and interpretation of mixed traffic stream comprising both autonomous and human-driven vehicles. Motivated by this research gap, the current study investigates the lane-changing impact of autonomous vehicles on the immediate follower using the publicly available Waymo Open Dataset. Human-driven vehicle lane-changing data are also extracted from the same database and used for comparison. Lane-changing impact on traffic flow efficiency and safety is examined through the speed reduction of the follower in the target lane and deceleration rate to avoid a collision for the same follower, respectively. A correlated random parameters linear regression model is employed to assess the speed reduction of the follower as a function of lane-change duration, lag gap, lane-changer speed, and a dummy variable indicating whether the lane-changer is an autonomous vehicle or a human-driven vehicle. The results reveal that lane changes executed by autonomous vehicles may cause greater or lesser speed reductions for the follower compared to those executed by human-driven vehicles, which could be attributed to the heterogeneous behaviour of followers perceiving and responding differently to autonomous vehicle lane-changes compared to human-driven ones. Further, the block maxima and peak over threshold models are developed to estimate crash risk for the follower in the target lane using a deceleration rate to avoid a collision conflict measure. The results suggest that the risk of a collision increases substantially when the lane-changer is an autonomous vehicle. This elevated risk may be associated with drivers' lack of trust in autonomous vehicles and traffic dynamics, reflecting self-inflicting hard deceleration to avoid potential collisions. Overall, this study highlights the heterogeneous impacts of lane-changing by autonomous vehicles on the immediate follower, emphasising the need for tailored models that accurately capture the dynamics of surrounding traffic behaviour. The findings will be helpful to road safety engineers and policymakers in planning mixed traffic with the safe integration of autonomous vehicles.

*E-mail address:* y.y.ali@lboro.ac.uk.

https://doi.org/10.1016/j.amar.2025.100388
Received 9 September 2024; Received in revised form 7 May 2025; Accepted 10 May 2025
Available online 12 May 2025

## 1. Introduction

Lane-changing is a complex manoeuvre, involving different objectives, lanes, and multiple decision-makers. Lane-changing increases a driver's mental workload, making it more error-prone and riskier. For example, about 622,222 lane change collisions occurred in the United States in 2020, accounting for 11.8 % of total collisions (NHTSA, 2021). These statistics highlight the safety associated with lane-changing and the need to properly understand lane-changing behaviour.

Lane-changing behaviour is often characterised by lane-changing decision-making, lane-changing execution, and lane-changing impact.[1] Whilst lane-changing decision-making and execution processes have been significantly studied in the literature, lane-changing impact has received comparatively less attention (Zheng, 2014). This disproportionate emphasis on lane-changing impact could be attributed to the insufficient understanding and evidence from real-world data—a gap that this study aims to address.

Notably, all past studies focus on lane-changing performed by human drivers and assessed its impact on traffic flow and safety (Xie et al., 2022; Yang et al., 2019). However, autonomous vehicles are increasingly becoming a reality now and significant research efforts are geared to better understand their integration into traffic stream and its impact on safety. Past studies along this line could be categorised into two groups: (a) driving simulator-based (Lin et al., 2025; Gouy et al., 2014), and (b) real-world (Ali et al., 2024; Wen et al., 2023). Most driving simulator studies are aimed at understanding the behavioural adaptation of human drivers in the presence of autonomous vehicles in different scenarios (or tasks), for example, a dedicated lane for connected and autonomous vehicles (Rad et al., 2021), overtaking autonomous truck platoon (Lin et al., 2025), car-following (De Zwart et al., 2023), among others. However, these studies did not analyse lane-changing scenarios, and their study design suffers from limitations like the fixed speed of autonomous vehicles during lane-changing, which may not be realistic. Concerning real-world autonomous vehicle data, two studies are prominent for lane-changing (Ali et al., 2024; Wen et al., 2023), but again these studies analysed lane-changing decision-making and execution, not lane-changing impact.

Some past studies (De Zwart et al., 2023; Wen et al., 2023) used trajectory data of autonomous vehicles and confirmed the differences in lane-changing behaviour of autonomous and human-driven vehicles. However, our understanding remains elusive of how autonomous vehicle lane-changes invoke behavioural adaptation to the immediate follower in the target lane (lane-changing impact), which will form the foundation of tailored algorithms/models for autonomous vehicles to minimise their negative impact on the traffic stream. Our literature review (see Table 1) also suggests that past studies in the lane-changing impact domain mostly use descriptive statistics or simple statistical models, raising the question of whether the lane-changing impact will be homogeneous or heterogeneous. Finally, to fully understand the safety of autonomous vehicles, past studies neglected the relationship among decision-making (gap selection), lane-changing execution, and lane-changing impact, which is crucial to properly understand lane-changing impact and relate it with other components of lane-changing behaviour.

Motivated by these research needs, the objective of this study is to investigate the impact of autonomous vehicle lane-changing on the immediate follower in the target lane (hereafter referred to as the "follower") and compare it with the impact of human-driven vehicle lane-changing. Specifically, this study aims to answer the following two research questions.

1. Which vehicle's lane-changing (autonomous versus human-driven) exerts more impact on the immediate follower? Whether such impact is homogeneous or heterogeneous across different human drivers (or followers)?
2. How are lane-changing execution and gap selection related to lane-changing impact?

By answering these questions, this study makes two contributions. First, as one of the foremost studies investigating the relatively underexplored topic of lane-changing impact—particularly in the context of autonomous vehicles—this study sheds light on the impact of autonomous vehicle lane-changing on traffic flow efficiency and safety. Specifically, for traffic flow efficiency, this study provides an in-depth understanding of follower's speed reduction in the target lane and compares it with speed reduction due to human-driven vehicle lane-changing. For safety, extreme value models are developed, characterising the crash risk for the follower with different lane-changers (autonomous vs. human-driven). This comparative analysis will aid in developing tailored models for quantifying autonomous vehicles' lane-changing impact, which can be embedded in microsimulations to comprehensively analyse lane-changing behaviour. Further, this study—for the first time—correlates decision-making (gap selection) and lane-changing execution (duration) with lane-changing impact, which, otherwise, are viewed as individual components of lane-changing behaviour. Second, through the application of advanced econometric models (correlated random parameter linear regression), this study highlights the varied effects of autonomous vehicle lane-changing on the follower, providing evidence of heterogeneous behaviour of the follower, which could inform tailored training programmes for driver education.

The remainder of the paper is structured as follows. The next section presents a literature review on lane-changing impact in general and in the context of autonomous vehicles in particular. The subsequent section describes the data collection process and modelling methodology, followed by a summary of the results. The discussion section elaborates on the research findings, and the last section concludes the study and suggests future research directions.

[1] Lane-changing impact is generally defined as the change in driving behaviour of the immediate follower in the target and original lanes caused by the lane-changing manoeuvre.

**Table 1**
Summary of representative studies on lane-changing impact and behavioural adaptation.

| Study | Vehicle type (AV or HDV) | Data used | Method | Heterogeneity | Relationship with LCD/LCE |
|---|---|---|---|---|---|
| Zhang and Zhao (2009) | HDV | Simulation | Finite state machine | × | × |
| Zheng et al. (2013) | HDV | NGSIM | Newell car-following model and linear regression | × | × |
| Gouy et al. (2014) | AV | Simulator | Descriptive statistics | × | × |
| Yang et al. (2019) | HDV | Shanghai naturalistic driving data | Descriptive statistics | × | × |
| Zhang (2019) | HDV | NGSIM | Descriptive and ANOVA | × | × |
| Ali et al. (2020) | HDV | NGSIM | Descriptive statistics | × | × |
| Rad et al. (2021) | CAV | Simulator | Linear mixed effects models | ✓ | × |
| Soni et al. (2022) | AV | Field test | Descriptive statistics | × | × |
| He et al. (2022) | HDV | Zen-Traffic | Descriptive statistics | × | × |
| Shin et al. (2022) | HDV | NGSIM | Linear mixed model | ✓ | × |
| Xie et al. (2022) | AV | Simulation | Curve fitting and descriptive statistics | × | × |
| De Zwart et al. (2023) | AV | Simulator | Descriptive statistics | × | × |
| Sultana and Hassan (2024) | CAV | Simulator | Generalised estimation equations | × | × |
| Lin et al. (2025) | AV | Simulator | Descriptive statistics | × | × |

**Abbreviations**
LCD: lane-changing decision-making; LCE: lane-changing execution; AV = autonomous vehicle; HDV = human-driven vehicle; CAV: connected and autonomous vehicle; "✓" Yes; "×" No; ANOVA: analysis of variance.

## 2. Background

This section briefly summarises representative studies related to the lane-changing impact of human-driven vehicles and the overall impact of autonomous vehicles. Providing a comprehensive review of lane-changing impact is beyond the scope of this study and can be found in Zheng (2014).

### 2.1. Review of lane-changing impact studies

A thorough literature review is conducted to ascertain the state-of-the-art related to lane-changing impact on the follower. It is worth noting that representative studies selected in this Table 1 are based on using real data for analysing lane-changing impact. Further, the review is neither comprehensive nor systematic, which is beyond the scope of the study, and readers are referred to existing reviews, e.g., Bevly et al. (2016) and Ali et al. (2025). Along this line, several studies are summarised in Table 1 and some key observations are as follows. Firstly, compared to two other lane-changing components (decision-making and execution), the lane-changing impact has been hardly studied. Notably, past studies on lane-changing impact are also in the context of human-driven vehicles. For instance, Yang et al. (2019) used naturalistic data to analyse the effects of lane-changing on the follower in the current lane and the target lane and reported that (a) about 20 % of followers responded to lane changes with an acceleration rate exceeding 10 % of the mean acceleration, (b) 90 % braked before time-to-collision reached 4.7 s, and (c) in over 70 % of lane changes, the minimum time-to-collision occurred between the initiation and cross-lane points.

Secondly, data used for human-driven vehicles is mostly NGSIM, whereas, for autonomous vehicles, a driving simulator is frequently used. However, utilising the data from large-scale field trials for autonomous vehicles (like Lyft and Waymo) appears to be overlooked. Thirdly, past studies have analysed lane-changing impact either using descriptive statistics or linear mixed models. Although linear mixed models can capture heterogeneity, understanding driver-level impact remained unexplored. Fourthly, some studies evaluated the lane-changing of autonomous vehicles using the Safety Pilot Deployment Program data and modelled cut-in scenarios. Results indicated that their proposed autonomous vehicle-based method can accelerate the evaluation process by at least 10 times (Zhao et al., 2015). Fifthly, none of the past studies linked lane-changing impact with two other components of lane-changing behaviour, that are, lane-changing decision-making (gaps) and lane-changing execution (duration). Understanding such linkage becomes paramount for evaluating the safety impacts of autonomous vehicles on the traffic stream.

Finally, some studies also used microsimulation tools to understand the impact of lane-changing on the traffic stream. For instance, Monteiro and Ioannou (2023) used the VISSIM microsimulation tool to understand the impact of lane-changing of different vehicle technologies like adaptive cruise control, autonomous vehicles, and connected and autonomous vehicles and found that connected and autonomous vehicles can achieve almost zero collision risk in congested and uncongested scenarios. Another study (Wang et al., 2022) developed an ego-efficient lane-changing strategy for connected and autonomous vehicles and found that the proposed strategy could benefit the entire traffic flow only if the market penetration rate of these vehicles is less than 50%. Different from microscopic evaluation, some studies also evaluated macroscopic impact through simulations. For example, Duell et al. (2016) analysed the impact of autonomous vehicles on traffic management, especially for dynamic lane reversal, and recommended using time-varying demand profiles when the dynamic lane reversal concept is explored.

### 2.2. Autonomous vehicles' impact on safety and behavioural adaptation

From a microscopic perspective and using real data, several studies developed models for quantifying safety associated with autonomous vehicles but in the context of car-following with a few exceptions. For example, Hu et al. (2023) found that autonomous vehicles are much safer than human-driven vehicles using traffic conflicts obtained from the Waymo dataset. Ali et al. (2024) evaluated the lane-changing execution of autonomous vehicles using the same data and found smoother manoeuvring of these vehicles compared to human-driven vehicles. Another study analysed the same data and reported that the crash risk of autonomous vehicles was found to be 50% less riskier than that of human-driven vehicles, with significant differences in the gap acceptance behaviour of autonomous and human-driven vehicles (Wen et al., 2023).

Existing literature extensively evaluates how the presence of autonomous vehicles in the traffic stream alters the driving behaviour of human-driven vehicles, and some of these studies are summarised in Table 1. These studies confirm a change in driving behaviour, which is contextual to driving conditions and tasks, leading to both safer and riskier behaviours (Lin et al., 2025). However, notably, all these studies[2] evaluated either car-following or overtaking.

Our review highlights some exigent gaps in lane-changing impact literature that are summarised as follows. Firstly, given the novelty of autonomous vehicles and the consequent scarcity of real-world data, the impact of autonomous vehicle lane-changing on the immediate follower remains unexplored. Secondly, some past studies (Hu et al., 2023; Li et al., 2023) suggest that autonomous vehicles will make traffic stream efficient and safer; however, much of these hypotheses are based on numerical simulations, ignoring human factors that are crucial in investigating the real impact of these vehicles (Sharma et al., 2018). Therefore, confirming past studies' conjecture and understanding the extent to which autonomous vehicles' impact varies from human-driven vehicles needs an investigation. Finally, given the uncertainty and varied surrounding traffic conditions, it is hypothesised that lane-changing impact may not be homogeneous. Addressing these research gaps requires an extensive investigation of the lane-changing impact of autonomous vehicles using real-data, which form the base of the current study.

## 3. Data and pre-processing

### 3.1. Dataset

For autonomous vehicles defined per the Society of Automotive Engineers Level 4, this study utilises the Waymo Open Dataset, accessible through the Waymo website (Sun et al., 2020). This dataset contains a three-dimensional object detection and tracking perception dataset and a motion dataset for motion/interaction prediction. This study uses the perception dataset of approximately 103,354 segments (Sun et al., 2020). Note that one segment (also referred to as a scenario) corresponds to one episode of driving, which includes 200 frames with a data resolution of 0.1 s. For each segment, trajectory data are extracted for lane-changing research, containing information about Waymo's autonomous vehicles and other vehicles detected by the Waymo vehicle, which are assumed to be all human-driven vehicles (Hu et al., 2022).

Waymo autonomous vehicles are fitted with high-resolution sensors that collect large-scale data in three U.S. cities, i.e., San Francisco, Phoenix, and Mountain View. This dataset contains rich information about road types (urban streets and freeways), weather (sunny and rainy), and time of day (dawn, day, dusk, and night). This study focusses on urban streets with sunny weather, which comprises 96.5% of the dataset, and data quality is superior to NGSIM (Hu et al., 2022). Note that the mean and standard deviation of segment-level traffic volume are 59.6 and 14.71, respectively.

In preliminary data processing, the original Waymo dataset was restructured to an accessible tabular format of trajectory data like NGSIM (see a detailed description of features of the dataset in Hu et al. (2022)). Briefly, the restructured dataset contained 25 attributes, including the segment environment information (time of day, weather, etc.), object features (object type, length, etc.), and object tracking trajectory (position, speed, heading, etc.). Qualitative verification was performed using camera videos and trajectory view animations that were generated from original videos (Hu et al., 2023) to ensure correct trajectories were obtained. Then, lane changes performed by Waymo autonomous vehicles were extracted by recording their IDs, their leaders in the current lane and the target lane, and followers in the current lane and the target lane. A typical example of a Waymo trajectory can be seen in Fig. 1. Similarly, lane changes performed by human-driven vehicles were also extracted for comparison and recorded in a similar format. Human-driven vehicle data were obtained through Waymo autonomous vehicles from the same location as that of autonomous vehicles. Human-driven vehicle data entail the same attributes as autonomous vehicle data including speed, acceleration, and position. Significant efforts have been made to ensure that only discretionary lane-changing manoeuvres were extracted, whereas lane changes performed close to intersections were excluded.[3] Note that this study obtained 147 and 151 discretionary lane-changing manoeuvres of autonomous vehicles and human-driven vehicles, respectively.

[2] Note that there are several other studies that evaluate change in driving behaviour of lane-changer during automation, which are not reviewed herein since the focus of the current study is on the follower, not on the lane-changer.

[3] Following Ali et al. (2021), discretionary lane-changing manoeuvres have been meticulously separated from mandatory ones; however, due to short time span of driving episodes (20 s each), deducing accurate information about mandatory lane-changing is very difficult as longer trajectories are required to infer about mandatory lane-changing intentions.

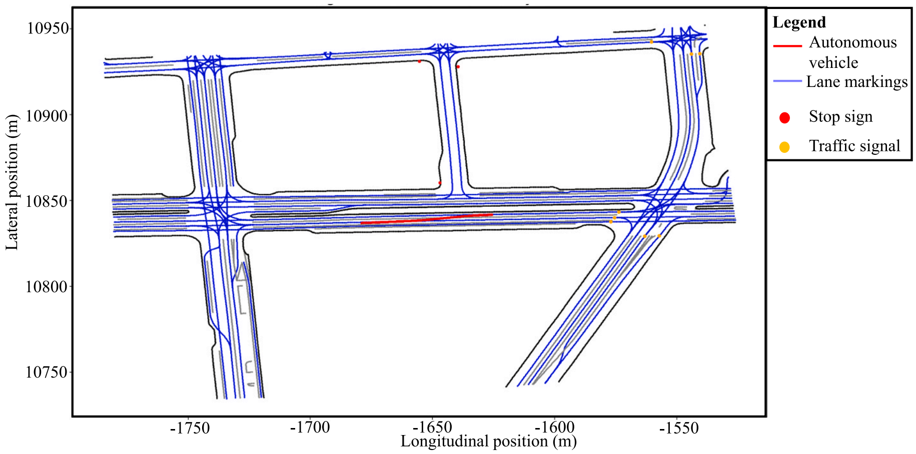



**Fig. 1.** A typical layout of an urban road segment with a lane-changing trajectory.

### *3.2. Data pre-processing*

One of the aims of this study is to understand the relationship between lane-changing execution (or duration) of lane-changers and its impact on the immediate follower. To this end, accurately identifying the start of a lane-changing manoeuvre and disentangling lane-changing decision-making from lane-changing execution becomes crucial. For this purpose, this study applies a Wavelet Transform method for identifying the starting point of lane-changing through lateral movement profiles. Mallat and Hwang (1992) and Zheng and Washington (2012) concluded that the wavelet modulus is effective in identifying subtle variations within a signal, which, in this case, pertains to a lateral movement profile. To obtain wavelet modulus, a lateral movement profile with respect to time is used as an input. The next step is to determine the type of wavelet. Ali et al. (2020) found the Mexican hat wavelet to be appropriate for lateral movement profiles, which is adopted herein. The subsequent step is to fine-tune the scale parameter of the wavelet, either by compressing or stretching to detect the start of a lane-changing manoeuvre. Lastly, the starting point is traced through long vertical maxima lines (a more detailed description can be found in two earlier studies: Zheng and Washington (2012) and Ali et al. (2020)). Using this procedure, maxima lines are drawn through each wavelet modulus (and connected by the red dotted line), the lane-changing starting point (i.e., Point A) can be located and a typical example is shown in Fig. 2. Identifying Points A and B and taking the difference between these two points allow for the measurement of the lane-changing duration for each manoeuvre.

From the trajectory data, a follower in the target lane is identified for each lane-changer and its basic driving behaviour variables are recorded, which are linked to the lane-changer's trajectory. Table 2 presents the list of these variables including lag gap, relative speed, among others (note that the last two rows indicate the dependent variables used in forthcoming modelling). The speed reduction of the follower is selected as a traffic flow efficiency measure because several studies have confirmed the relationship between speed and traffic flow efficiency (Zhang et al., 2020; Soriguera et al., 2017). For safety analysis, the deceleration rate to avoid a collision (DRAC) traffic conflict measure is selected in this study. As reported in a review of traffic conflict measures (Arun et al., 2021), DRAC is an appropriate measure to characterise the crash risk of the follower in the target lane. DRAC is calculated for the entire lane-changing execution period and its maximum value—reflecting the highest crash risk—is selected for the analysis.

## 4. Modelling methodologies

To compare the impact of lane-changing manoeuvres executed by autonomous vehicles with those performed by human drivers, this study develops two separate models: one for evaluating traffic flow efficiency and the other for assessing safety. The first model employs a correlated random parameters linear regression approach to analyse the speed reduction of the follower. The second model uses an extreme value theory approach to model the deceleration rate required to avoid a collision for the follower. The following subsections provide a detailed explanation of these modelling methodologies.

### *4.1. Correlated random parameters linear regression model*

To understand the impact of lane-changing on the speed reduction of the follower and simultaneously control for vehicle-specific factors, a linear regression model is developed. The linear regression model is selected for two reasons. Firstly, the speed reduction (response) variable follows a normal distribution (see Fig. 3 for an illustration). Further, an Anderson-Darling test is performed, and the null hypothesis for this test is that the speed reduction variable follows a normal distribution, and the test statistic confirms a failure to reject the null hypothesis at a 95% confidence level (*test statistic* $= 0.738$; $p$-value $= 0.06$). Secondly, several non-linear models were developed, assuming that speed reduction follows gamma or truncated normal/student-$t$ distributions. These models did not show a statistically superior fit to a linear regression model; hence a linear regression model is selected.

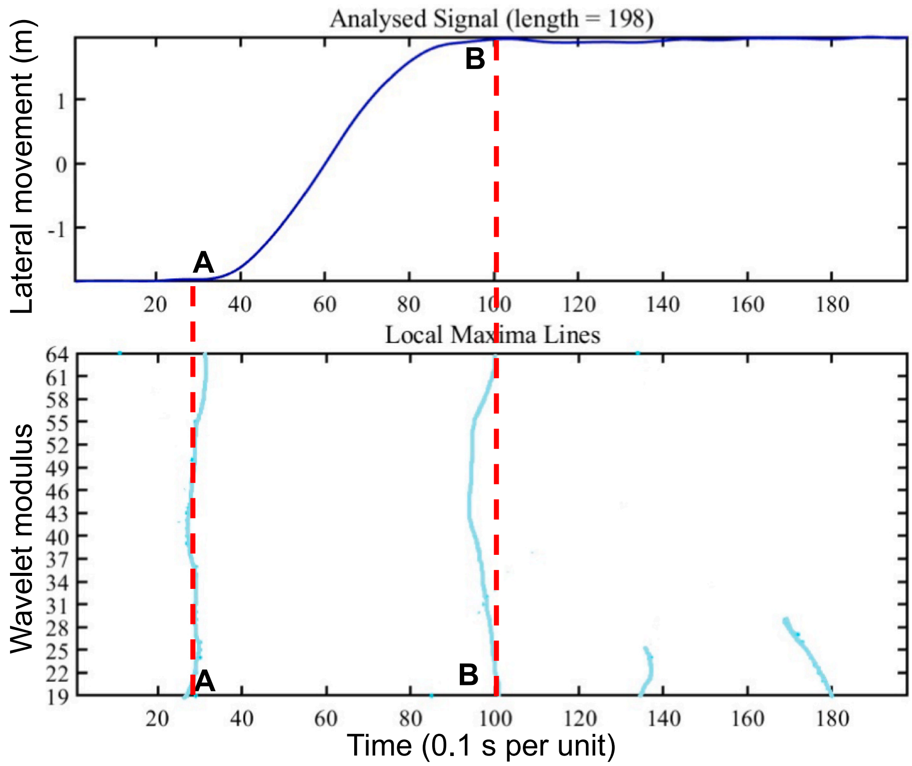



**Fig. 2.** A typical example of extracting lane-changing execution.

**Table 2**
Summary driving behaviour measures for analysis.

| Variable | Description |
|---|---|
| Lane-change duration | The time taken by a lane-changer to complete its physical lane-changing manoeuvre from the current lane to the target lane. |
| Lag gap | The bumper-to-bumper distance between the rear end of the lane-changer in the current lane to the front end of the follower in the target lane. |
| Speed of lane-changer (follower) | The instantaneous speed of a lane-changer (follower) during lane-changing. |
| Relative speed | The difference between the speed of the lane-changer in the current lane and the follower in the target lane. |
| Acceleration noise | The standard deviation of acceleration/deceleration of the follower. |
| Jerk | Rate of change of acceleration/deceleration of the follower. |
| Speed reduction | The reduction in follower's speed from the onset of lane-changing execution and when it finishes the lane-changing manoeuvre. |
| DRAC (deceleration rate to avoid a collision) | The rate at which a following vehicle must decelerate to avoid a collision with the leading vehicle is defined as the differential speed between a following vehicle and its leading vehicle divided by their closing time. |

Mathematically, a linear regression model can be defined as (Washington et al., 2020)

$$y_i = \alpha + \boldsymbol{\beta}_i \boldsymbol{X}_i + \varepsilon \tag{1}$$

where, $y_i$ is the dependent variable (i.e., the speed reduction of the follower), which is a function of the constant term, $\alpha$, and the coefficient $\boldsymbol{\beta}_i$ multiplied by the value of the independent variables, $\boldsymbol{X}_i$ (speed, lag gap, and lane-change duration) for vehicle $i$, plus an error term $\varepsilon$.

The formulation presented in Equation (1) assumes that the effect of independent variables on speed reduction is homogeneous, which may be unrealistic because of unobserved heterogeneity (that is, unobserved factors varying systematically across the observations). To capture unobserved heterogeneity, a random parameters approach is generally preferred (Mannering et al., 2016) and employed in this study, whereby $\boldsymbol{\beta}$ parameter is allowed to vary across the observations as (Washington et al., 2020)

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \Gamma \delta_i \tag{2}$$

where, $\boldsymbol{\beta}_i$ is the parameter for vehicle $i$, $\boldsymbol{\beta}$ is the mean parameter estimate across all observations, $\Gamma$ is the Cholesky matrix, whose elements are used for calculating the standard deviations of the random parameters, and $\delta_i$ is a randomly distributed term with mean

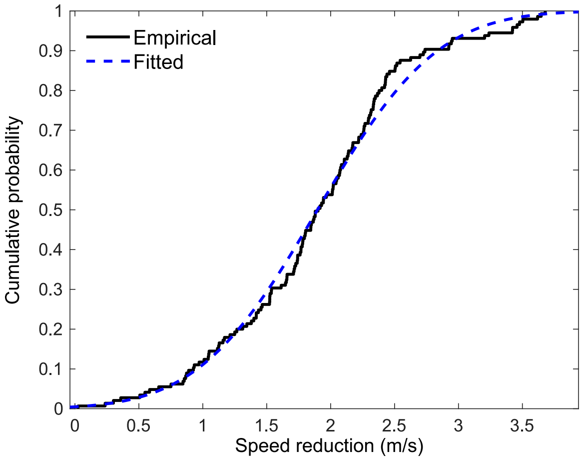



**Fig. 3.** A typical illustration of autonomous vehicle speed reduction with the fitted normal distribution.

and variance equal to zero and one, respectively.

To account for possible correlations between the random parameters, the unrestrictive form of the Cholesky matrix ($\Gamma$) allows non-zero off-diagonal elements, capturing the potential correlation during the estimation of the random parameters (Greene, 2012). More specifically, the variance–covariance matrix ($V$) of the random parameters based on the Cholesky decomposition can be derived as $V = \Gamma\Gamma'$. Note that the squared values of the standard deviations of the correlated random parameters are represented as the diagonal elements of the variance–covariance matrix. Considering the case of the uncorrelated random parameters, the off-diagonal and diagonal elements of the variance–covariance matrix are equal to zero and the standard deviations of the correlated random parameters, respectively (Greene, 2012). Contrastingly, the unrestricted form of the Cholesky matrix estimates the correlation between random parameters, providing the impact of the interactive effect of unobserved heterogeneity related to the explanatory variables with correlated random parameters.

The random parameters model is estimated using a simulated maximum likelihood estimation procedure, whereby Halton draws (Bhat, 2003) are used for complex numerical integrations. Following the relevant literature (Huo et al., 2020; Fountas et al., 2018), 1,000 Halton draws are used to obtain stable random parameters.

### 4.2. Extreme value theory models

Whilst the random parameters model provides an understanding of factors affecting speed reduction and indirectly relates to safety, extreme value models provide direct safety understanding by calculating crash risk. For this purpose, this study applies an extreme value theory approach that allows extrapolating crash risk from frequently observed events (conflicts) to rarely observed events (crashes) using crash-conflict relationships. One peculiar advantage of this approach is that it does not rely on crash data, which is difficult to obtain for autonomous vehicles due to their smaller penetration rate. Note that several studies have applied extreme value models to understand crash risk in different contexts, e.g., pedestrian safety (Hussain et al., 2024), the effect of a connected environment of lane-changing safety (Ali et al., 2022b), and using autonomous vehicle data for network safety (Singh et al., 2024). To realise the crash-conflict relationship, two sampling approaches exist, namely block maxima and peak over threshold, which are described below.

#### 4.2.1. Block maxima approach

The block maxima approach segments observations in a temporal domain by specifying predefined time-based blocks and identifying extremes within each block. Mathematically, let $x_1, x_2, x_3, \cdots x_n$ be a sequence of random and independent variables with a common distribution function and $M_n = max(x_1, x_2, x_3, \cdots x_n)$ yields the block maximum of $n$ blocks, where $x_i$ refers to a traffic conflict measure (deceleration rate required to avoid a collision) extracted from each block. Theoretically, the maximum of conflict values converges to a Generalised Extreme Value distribution when $n \to \infty$, which can be defined as

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

where, $-\infty < \mu < \infty$ corresponds to the location parameter, $\sigma > 0$ represents the scale parameter, and $-\infty < \xi < \infty$ denotes the shape parameter.

#### 4.2.2. Peak over threshold approach

The peak over threshold approach is an event-based approach that identifies extremes from a set of observations that surpass a predefined threshold. Given $x_1, x_2, x_3 \cdots x_n$ represents random observations that are independent and identically distributed, the cumulative distribution function of exceedances, $Y = X - u$, conditioned upon $X > u$, can be obtained as $F_u(y) = P(X \leq y + u | X > u)$. The distribution can be approximated to a Generalised Pareto distribution for a threshold $u$ with a sufficiently large value and can be obtained as

$$G(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, \ \xi \neq 0 \ and\, y > 0 \tag{4}$$

where, $y$ is the exceedance, $\sigma$ and $\xi$ are the scale and shape parameters of Generalised Pareto distribution, respectively.

The tail of extreme value distributions is of particular interest in estimating crash risk without relying on crash data, which is especially relevant in the context of autonomous vehicles. As such, the tail of the fitted extreme value distributions is used to obtain the follower's crash risk during lane-changing. For this purpose, the deceleration rate to avoid a crash (DRAC) traffic conflict measure is used for modelling and characterising follower-lane-changer interactions. Note that other conflict metrics like time-to-collision could also be selected; however, past studies (Arun et al., 2021) reported the constant speed assumption of time-to-collision to be problematic. Further, the choice of conflict metric will have a negligible effect on the comparison since our focus is on relative comparison. A large DRAC value represents a higher crash risk, whereby DRAC exceeding the maximum available deceleration rate (MADR) indicates a crash. As such, crash risk ($R$) can be obtained as

$$R = Pr(DRAC \geq MADR) \tag{5}$$

Unlike proximity-based or temporal traffic conflict measures (time-to-collision) where the boundary between crash and non-crash events is clear, the boundary condition in DRAC is neither fixed nor unified for different scenarios, such as braking capabilities of individual vehicles under different road, traffic, and environmental conditions. Therefore, this study finds MADR to follow a truncated normal distribution (similar to Zheng and Sayed (2019)): $MADR\ normal(7.96, 2.73^2)\mathrm{I}(4.21,\ 10.23)$.

To estimate crash risk using the fitted extreme value models, the probability that DRAC exceeds MADR is obtained through a simulation procedure (Zheng and Sayed, 2019) described as follows. Firstly, $m$ and $n$ samples of DRAC and MADR are drawn from the fitted extreme value distributions and truncated normal distribution, respectively. Secondly, these two samples are combined and ranked in increasing order. Thirdly, each element in the DRAC and MADR sets is replaced by its corresponding rank from the second step. Let $r_1 < r_2 < \cdots r_m$ represent the DRAC samples and let $r_i$ be selected. The probability that MADR is smaller than the selected DRAC value is calculated as the number of smaller values in the MADR set divided by $n$. If $r_i - 1$ values are smaller in the DRAC set, then the conditional probability that DRAC exceeds MADR for a given DRAC value can be computed as

$$Pr(DRAC > MADR) = \sum_{i=1}^{m} Pr(DRAC > MADR | DRAC = r_i) \tag{6}$$

$$Pr(DRAC = r_i) = \frac{1}{mn}(r_i - 1 \tag{7}$$

A crucial aspect of the peak over threshold approach is determining an appropriate threshold, which is obtained by identifying the intersection of the mean residual plot and the threshold stability plot (Coles, 2001). Specifically, a linear range of thresholds is first identified in the mean residual plot followed by a similar range in the threshold stability plots, where the parameters remain constant. The maximum value from the intersection of these two ranges is selected as the threshold.

Another important aspect of extreme value models is the non-stationarity of traffic conflicts, which must be addressed during model estimation. For this purpose, several covariates (or explanatory variables) are introduced in extreme value distribution parameters using an identity link function to capture time-varying non-stationarity and unobserved heterogeneity. For the block maxima model, location and scale parameters of Generalised Extreme Value (GEV) are parameterised, whereas the shape parameter is not parameterised because of difficulty during model estimation (Coles, 2001). Mathematically, GEV parameters can be expressed as

$$\begin{cases} \mu = \mu_0 + \mu_1 X \\ \sigma = \sigma_0 + \sigma_1 Y \\ \xi = \xi_0 \end{cases} \tag{8}$$

where, $\mu_0$, $\sigma_0$, and $\xi_0$ are intercept parameters for the location, scale, and shape, respectively. Similarly, $\mu_1$ and $\sigma_1$ are estimable parameters corresponding to explanatory variable vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively.

For the peak over threshold model, only the scale parameter of the Generalised Pareto distribution is parameterised using a log link function to ensure the positiveness of the scale parameter, which can be expressed as

$$ln(\sigma) = \sigma_0 + \sigma_1 \boldsymbol{Z} \tag{9}$$

where, $\boldsymbol{Z}$ denote a vector of explanatory variables used for the peak over threshold model.

Extreme value model parameters are estimated using the maximum likelihood procedure and model goodness-of-fitness is assessed using conventional metrics as explained in the next section.

## 5. Results

### 5.1. Descriptive analysis of lane-changing impact

As mentioned in the introduction, it remains unclear whether the impact of lane-changing differs between autonomous and human-driven vehicles. To answer this question, a descriptive analysis (Table 3) of various driving behaviour variables is conducted and statistically compared using the Wilcoxon rank-sum test (the normality assumption was assessed, and the results justified the use of non-parametric tests) and some key observations are as follows. Firstly, autonomous vehicles tend to change lanes in larger lag gaps compared to human-driven vehicles, suggesting large distance availability for the follower to accommodate a lane-changer. Secondly, the followers of autonomous vehicles drive faster than those of human-driven vehicles. Although the faster driving of the follower does not necessarily mean a high risk of collision, in the case of a lane-changing scenario, it represents a risky situation because the follower needs to brake hard to accommodate the lane-changer, which could be detrimental to the safety of the follower and for the entire traffic stream (Loulizi et al., 2019; Kim et al., 2016). Thirdly, the acceleration noise of followers for autonomous vehicle lane-changing is about 0.07 $m/s^2$ higher than human-driven vehicles, exhibiting significant changes in followers' acceleration behaviour when interacting with autonomous vehicles. Finally, followers exhibit significantly higher jerks when a lane-changer is an autonomous vehicle relative to a human-driven vehicle. This preliminary analysis suggests that the impact of lane-changing indeed varies when the lane-changer is an autonomous vehicle compared to a human-driven vehicle.

From Table 3, it is evident that autonomous vehicles take more time to complete their lane-changing manoeuvres compared to human-driven vehicles, which could be attributed to smoother movement from the current lane to the target lane. However, this longer duration may cause greater disruption to the follower compared to shorter lane-changing durations. To further explore this relationship, a correlated random parameters linear regression model is developed, as explained below.

### 5.2. Correlated random parameters linear regression model

Table 3 presents the summary of model estimation results for the correlated random parameters linear regression model fitted to the speed reduction data of the follower. Two random parameters are found to be significant in the model and the literature suggests capturing correlation among them (Fountas et al., 2018). Using the unrestrictive form of the Cholesky matrix, the correlation between random parameters is captured and this model is compared with an uncorrelated random parameters linear regression model. To justify the statistical superiority of a modelling approach, a likelihood ratio test is performed. The test statistics for the correlated random parameters model versus the uncorrelated random parameters model is $\chi^2 = -2 \times [-866.84 - 874.33] = 14.98$. With one degree of freedom, the test statistic suggests that the null hypothesis that the uncorrelated random parameters model (or the simpler model) could be preferred over the correlated random parameters model (or the complex model) can be rejected at a 95% confidence level. In other words, the test statistic does not support that both models have the same parameters. The Akaike Information Criterion (AIC) for the correlated random parameters is 1751.6, whereas the corresponding AIC for the uncorrelated random parameters model is 1764.6, indicating a better fit of the correlated random parameters model, and hence selected in this study. Further, the Nakagawa's Pseudo $R$-square[4] (Nakagawa and Schielzeth, 2013) for the correlated random parameters model is 0.261, which is about 1.62 times greater than a fixed effect linear regression model (0.161).

Another set of comparisons is performed to ascertain whether advanced approaches like heterogeneity in mean or variance or both could be used. Three models are developed, namely correlated random parameters with heterogeneity in means, random parameters with heterogeneity in variance, and random parameters with heterogeneity in means and variance. Several variables are tested for heterogeneity in mean and variance, but none of them is found to be significant and increases model goodness of fit. Further, a series of likelihood ratio tests are performed, confirming the statistical superiority of the correlated random parameters model without heterogeneity in mean/variance over competing models. For example, the test statistic for the correlated random parameters model without heterogeneity in mean/variance versus the random parameters model with heterogeneity in mean and variance is $\chi^2 = -2 \times [-866.84 - 864.94] = 3.80$. With two degrees of freedom, the test statistic suggests that the null hypothesis that the correlated random parameters model without heterogeneity in mean/variance (or the simpler model) could be preferred over the random parameters model with heterogeneity in mean and variance (or the complex model) can be rejected at a 95% confidence level. Similar results are found for other model comparisons, thereby justifying our selection of a correlated random parameters model without heterogeneity in mean or variance.

Table 4 presents the parsimonious model results, containing all statistically significant parameters at a 95% confidence level. The speed of a lane-changer and the dummy variable for the autonomous vehicle are random parameters. Following the existing literature, several popularly used probability density functions, such as normal, log-normal, Weibull, uniform, and triangular, were tested for the random parameters, and the normal distribution outperformed others in terms of statistical fit and interpretation, which corroborates the safety literature (Fountas et al., 2018; Mannering et al., 2016). The non-random parameters in the model are lane-change duration, lag gap, and initial speed of the follower. The linear regression function (Equation (1) can be rewritten as

[4] The Nakagawa's Pseudo $R$-square provides the marginal $R^2$ (proportion of variance explained by the fixed effects alone) and conditional $R^2$ (proportion of variance explained by both fixed and random effects).

**Table 3**
Statistical summary of different driving behaviour variables.

| Variable | Autonomous vehicle (standard deviation) | Human-driven vehicle (standard deviation) | Significance by a Wilcoxon test | Remark (5 % significance) |
|---|---|---|---|---|
| Lane duration (s) | 7.573 (2.839) | 5.908 (2.113) | $p$-value < 0.001 | Significant |
| Lag gap (m) | 15.154 (14.942) | 8.898 (6.053) | $p$-value = 0.539 | Insignificant |
| Speed of lane-changer (m/s) | 11.135 (0.755) | 13.535 (1.098) | $p$-value = 0.003 | Significant |
| Speed of follower (m/s) | 12.616 (1.792) | 10.962 (0.861) | $p$-value = 0.040 | Significant |
| Relative speed of lane-changer and follower (m/s) | 0.755 (0.462) | 1.361 (0.696) | $p$-value = 0.264 | Insignificant |
| Acceleration noise (m/s$^2$) | 1.728 (0.979) | 1.656 (0.960) | $p$-value < 0.001 | Significant |
| Jerk (m/s$^3$) | 1.262 (0.836) | 0.059 (0.665) | $p$-value < 0.001 | Significant |

**Table 4**
Estimation results of the correlated random parameters linear regression model for speed reduction.

| Parameter | Estimate | Std. Error | $z$-statistics | $p$-value | 95% confidence interval Lower bound | Upper bound |
|---|---|---|---|---|---|---|
| Constant | 0.927 | 0.088 | 10.5 | <0.001 | — | — |
| ***Non-random parameters*** | | | | | | |
| Lane-change duration | 0.311 | 0.029 | 10.610 | <0.001 | 0.253 | 0.368 |
| Lag gap | −0.005 | 0.002 | −2.600 | 0.009 | −0.009 | −0.001 |
| Initial speed of the follower | 0.624 | 0.024 | 26.00 | <0.001 | 0.577 | 0.672 |
| ***Random parameters*** | | | | | | |
| Speed of lane-changer | 0.073 | 0.010 | 6.940 | <0.001 | 0.052 | 0.093 |
| Autonomous vehicle | 3.991 | 0.408 | 9.790 | <0.001 | 4.790 | 3.192 |
| ***Diagonal elements in the Cholesky matrix*** | | | | | | |
| Speed of lane-changer | 0.040 | 0.005 | 7.420 | <0.001 | 0.030 | 0.051 |
| Autonomous vehicle | 5.970 | 0.199 | 29.940 | <0.001 | 5.579 | 6.361 |
| ***Below diagonal element in the Cholesky matrix*** | | | | | | |
| Speed of lane-changer: Autonomous vehicle | −3.955 | 0.336 | −11.790 | <0.001 | −4.613 | −3.298 |
| Variance parameter ($\sigma$) | 2.377 | 0.092 | 25.720 | <0.001 | 2.196 | 2.558 |
| Number of observations = 297; LL ($\beta$) = -866.84; LL (0) = -890.42; Chi-square = 47.16 ($p$-value < 0.001); AIC = 1751.6. | | | | | | |

$$\begin{aligned} \textit{Speed reduction} = {} & 0.927 + 0.311 \times \textit{lane change duration} - 0.005 \times \textit{lag gap} + 0.624 \times \textit{follower speed} + \beta_{\textit{Lane changer speed}} \\ & \times \textit{Lane changer speed} + \beta_{\textit{Autonomous vehicle}} \times \textit{Autonomous vehicle} \end{aligned} \tag{10}$$

where,

$$\begin{pmatrix} \beta_{\textit{Speed of lane changer}} \\ \beta_{\textit{Autonomous vehicle}} \end{pmatrix} = \begin{pmatrix} 0.073 \\ 3.991 \end{pmatrix} + \begin{pmatrix} 0.040 & 0 \\ -3.955 & 5.97 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} \tag{11}$$

is the specified correlation structure between random parameters with $\varphi_1$ and $\varphi_2$ representing the independent standard normally distributed random variables. Note that multicollinearity among model parameters was assessed and found to be smaller than 2, signifying no multicollinearity. Along this line, the correlation coefficient for the speed of the lane-changer and the initial speed of the follower is 0.09 and this correlation is found to be statistically insignificant. Further, several variables, in addition to those mentioned in Table 2, were also tested during the model development process like relative speed, relative acceleration, and time-to-collision. However, these variables were not retained in the parsimonious model as they neither showed statistical significance nor increased model goodness-of-fit. Along the same line, interaction effects were included in the model like lane-changer speed and dummy variable for autonomous vehicles but were not retained due to the aforementioned reason.

Table 4 also presents the diagonal and below diagonal elements of the Cholesky matrix for each random parameter, which can be used for calculating the standard deviation of the random parameters as the square root of the variance (elements on the diagonal of the variance–covariance matrix, calculated as $\Gamma\Gamma'$). The standard deviations of the speed of lane-changer and autonomous vehicle parameters are calculated as $\sqrt{0.0016} = 0.04$ and $\sqrt{51.29} = 7.162$, respectively.

*Lane-change duration* is positively associated with the speed reduction of the follower, indicating that with an increase in lane-change execution time, the follower's speed reduction also increases. Specifically, each additional second of lane-change duration tends to increase the follower's speed reduction by 0.3 m/s. Similarly, the *initial speed of the follower*—measured as the instantaneous speed at the onset of lane-change execution—is also significant and positively associated with the speed reduction of the follower. This relationship suggests that a follower, moving at a faster speed, is likely to experience a greater speed reduction during lane-changing, with a 0.9 m/s speed reduction for every 1 m/s increase in the follower's initial speed. Contrastingly, the *lag gap* is significant and negatively related to speed reduction, indicating that when the follower maintains a large gap to the lane-changer, their speed reduction decreases. Specifically, with every one metre increase in the lag gap, speed reduction decreases by 0.005 m/s.

The *lane-changer speed* parameter is found to be random in the model, whereby its mean and standard deviation are significant in the model. The mean of the lane-changer speed parameter is positive, indicating that when the lane-changer executes the manoeuvre at a higher speed, the follower's speed reduction increases. This higher speed reduction could be follower's risk compensation for faster lane-changers, who are likely to complete their lane-change manoeuvres quickly, thereby inducing greater speed reduction. Notably, heterogeneity is found in the effect of lane-changer speed (see Fig. 4 (a)) as a small proportion of coefficients are found to be negative (4%), suggesting that in some cases, a higher lane-changer speed may actually reduce the follower's speed reduction, which could be safer for the follower.

The model estimation results indicate that the mean and standard deviation of the *dummy variable for autonomous vehicles* are significant, suggesting significant heterogeneity in speed reduction caused by vehicle type. Fig. 4 (b) illustrates this heterogeneity, showing that speed reduction increases for most autonomous vehicle lane changes (72%), but not in all cases. The positive sign implies that the follower's speed reduction tends to increase when the lane-changer is an autonomous vehicle as compared to a human-driven vehicle, which may be attributed to a negative attitude towards and response to autonomous vehicles. This behaviour could also be due to the aggressive nature of some followers who attempt to discourage lane-changing manoeuvres, as these manoeuvres significantly affect their motion. However, the random parameter also indicates that speed reduction may decrease when the lane-changer is an autonomous vehicle relative to a human-driven vehicle, which could be associated with courteous behaviour by followers—a tendency often observed during the lane-changing process.

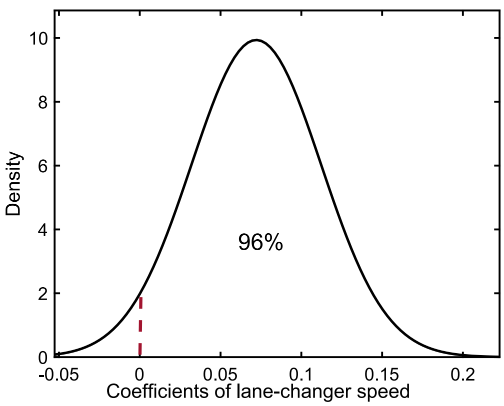
Table 4 presents the diagonal and below-diagonal elements of the Cholesky matrix, which are used to calculate the variance–covariance of the correlated random parameters, thereby assisting in determining the correlation coefficient for these parameters. Using the post-hoc estimation technique presented by Fountas et al. (2018), this study finds that the speed of the lane-changer and the dummy variable for autonomous vehicles are statistically correlated at a 5% significance level ($t$-stat $= 4.03$; $p$-value $< 0.001$), with a covariance of $-0.159$ and a correlation coefficient of $-0.55$. In this study, the negative correlation between the speed of the lane-changer and the dummy variable for autonomous vehicles suggests a heterogeneous effect of unobserved characteristics on the speed reduction of the follower associated with the speed of autonomous vehicles. In other words, an increase in the speed of an autonomous vehicle lane-changing may either increase or decrease the follower's speed reduction due to unobserved heterogeneity associated with these two parameters. This finding implies that followers may experience varying levels of speed reduction when the lane-changer is an autonomous vehicle and increases its speed compared to a human-driven vehicle. This variation could be related to the follower's risk perception as some followers may perceive a faster speed of autonomous vehicle lane-changing as riskier, leading to a greater speed reduction to avoid a rear-end collision. Conversely, some followers might interpret the faster lane-changing speed of autonomous vehicles as a swift manoeuvre into the target lane, thereby requiring less speed reduction.

### 5.3. *Extreme value theory models*

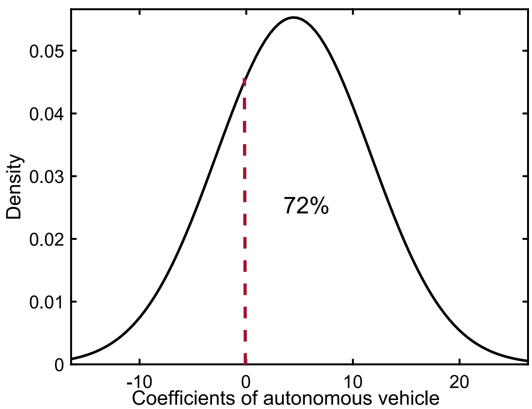
Extreme value (block maxima and peak over threshold) models are developed to model the impact of lane-changing using the deceleration rate to avoid a collision (DRAC) conflict measure. For the block maxima model, the block size is selected as an episode of 20 s—similar to a previous autonomous vehicle study (Singh et al., 2024) and a driving simulator study (Ali et al., 2022a).

The goodness-of-fit of the best-fitted models is assessed using Q-Q plots of the empirical and modelled DRAC, as shown in Fig. 5. By visually inspecting these plots, it can be inferred that models have fitted the data well as most of the points lie on the 45° line of equality.

For the peak over threshold model, determining the appropriate threshold is crucial. Following the procedure outlined in the previous section, the mean residual plot (Fig. 6 (a)) shows a linear portion between 5.75 and 6, whilst the threshold stability plots (Fig. 6 (b) and (c)) indicate stable portions between 5 and 6. The maximum intersection value of the mean residual and threshold stability plots, which is 6, is selected. Additionally, Fig. 6 (d) presents the Akaike Information Criterion (AIC) plot, representing AIC values for models estimated at different thresholds. The plot indicates the minimum value at 6, aligning with the intersection of the mean residual and threshold stability plots. Consequently, this threshold is used for peak over threshold modelling.

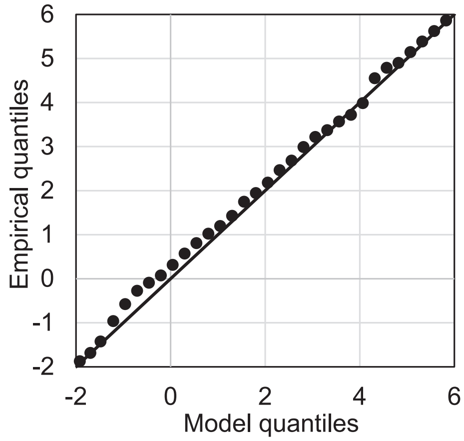



(a) Speed of lane-changer

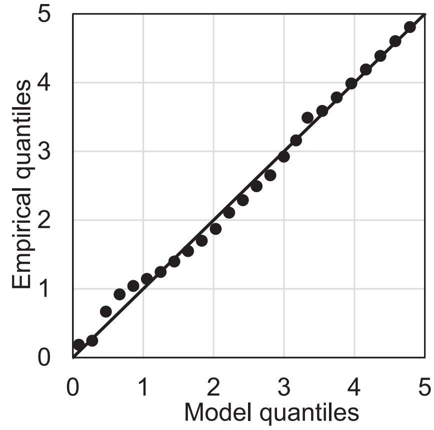



(b) Dummy variable for autonomous vehicles

**Fig. 4.** Distributional effect of the random parameters.

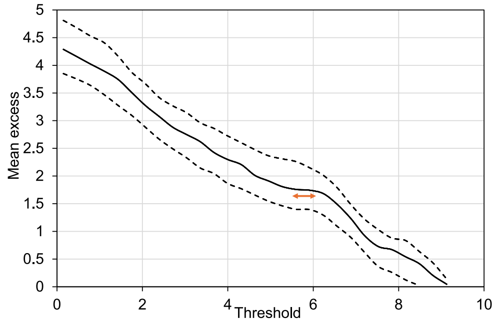



(a) Block maxima model

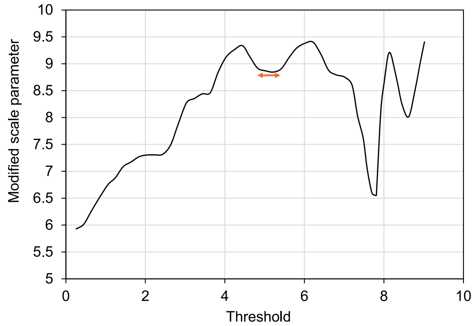



(b) Peak over threshold

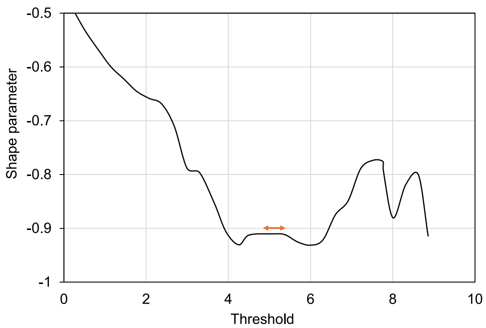
**Fig. 5.** Extreme value model diagnostics.



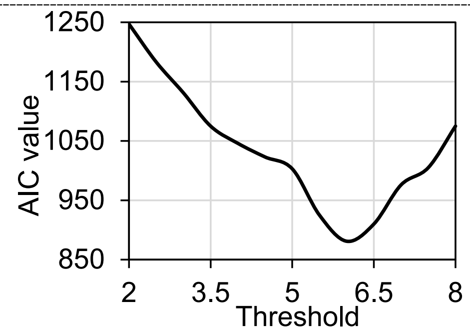

(a) Mean residual life plot




(b) Modified scale parameter plot




(c) Shape parameter plot




(d) Goodness of fit

**Fig. 6.** Schematics for threshold selection plots.

Table 5 presents the parsimonious extreme value model results. Several explanatory variables were tested in the models (see a list in Table 2), but the best-fitted model was selected with all parameters being significant at a 95% confidence level and providing intuitive interpretation. Further, separate models were developed for human-driven vehicles and autonomous vehicles, but these models did not show a statistically superior fit compared to the combined model for human-driven vehicles and autonomous vehicles.

For the block maxima model, the autonomous vehicle dummy variable is incorporated in the location parameter, whereas the lag gap and speed of the lane-changer are introduced in the scale parameter. For the peak over threshold model, all three explanatory variables are included in the scale parameter. Several other combinations of parameterisation (location only or scale only) were also tested, but they did not outperform the models presented in Table 5. Given the sign convention of model parameters is the same in the block maxima and peak over threshold models, model parameters are only explained for the block maxima model and a similar

**Table 5**
Summary of extreme value model estimation results for deceleration rate to avoid a collision.

| Model | Parameter | Description | Notation | Estimate | Std Error | AIC (BIC) |
|---|---|---|---|---|---|---|
| Block maxima | Location | Intercept | $\mu_0$ | 9.0678 | 0.5295 | 1755 (1777) |
| | | Autonomous vehicle (AV) lane-changer dummy | $\mu_{AV}$ | 7.5248 | 0.5381 | |
| | Scale | Intercept | $\sigma_0$ | 1.9305 | 0.0717 | |
| | | Lag gap | $\sigma_{laggap}$ | −0.0214 | 0.0021 | |
| | | Speed of lane-changer (LC) | $\sigma_{LCspeed}$ | 0.0010 | 0.0006 | |
| | Shape | Intercept | $\xi_0$ | 0.1173 | 0.0428 | |
| Peak over threshold | Scale | Intercept | $\mu_0$ | 2.2625 | 0.1009 | 881 (891) |
| | | Autonomous vehicle (AV) lane-changer dummy | $\mu_{AV}$ | 0.8779 | 0.0840 | |
| | | Lag gap | $\sigma_{laggap}$ | −0.0005 | 0.0001 | |
| | | Speed of lane-changer (LC) | $\sigma_{LCspeed}$ | 0.0192 | 0.0002 | |
| | Shape | Intercept | | −0.4797 | 0.0751 | |
| | | Number of exceedances = 112 | | | | |

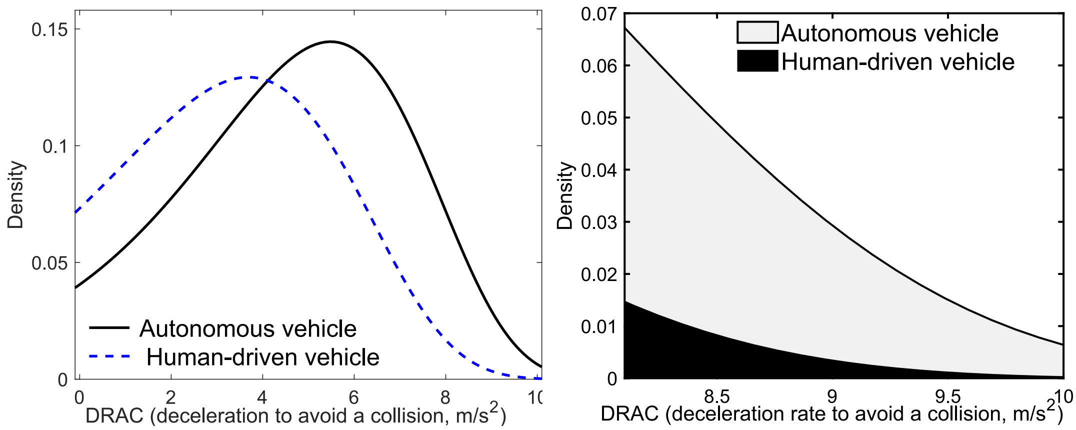
interpretation can be made for the peak over threshold model. Similarly, only generalised extreme value distributions are generated for understanding crash risk, whereby generalised Pareto is not shown herein because of the aforementioned reason.

The dummy variable of an autonomous vehicle in the location parameter of Generalised Extreme Value (GEV) distribution is positive, suggesting that the mean of GEV distribution shifts to the right, thereby increasing followers' crash risk (as the rightmost DRAC values are riskier). Similarly, the speed of the lane-changer parameter is found to be significant in the scale parameter of the GEV distribution, suggesting more extreme values, wider spread, and heavier tail of the GEV distribution, leading to increased follower's crash risk. Contrastingly, the negative relationship of the lag gap leads to compressed extreme values with a shorter tail, thereby reducing the follower's crash risk. Note that the peak over threshold model results can similarly be interpreted.

Further, using the fitted model, Generalised Extreme Value distributions are generated separately for human-driven and autonomous vehicles, as shown in Fig. 7 (a). Following the crash risk analysis procedure outlined in the previous section, the overlapping area of the deceleration rate to avoid a collision and the maximum available deceleration rate distributions is found to be over 8 $m/s^2$, as illustrated in Fig. 7 (b). The area under the curve is computed, allowing for a comparative analysis of crash risk between the two vehicle types, with a larger value indicating a higher crash risk. The areas under the curve for human-driven and autonomous vehicles are 0.0098 and 0.0633, respectively, indicating a higher crash risk when the lane-changer is an autonomous vehicle compared to a human-driven vehicle. This finding reaffirms the results obtained from the random parameters model, suggesting that the lane-changing behaviour of autonomous vehicles impacts followers more significantly than that of human-driven vehicles.

## 6. Discussion

This study investigates the impact of lane-changing on the immediate follower in the target lane—a topic that Zheng (2014) notes has received disproportionate attention in the lane-changing literature. With autonomous vehicles being just on the horizon, understanding the lane-changing impacts of these vehicles becomes increasingly urgent. A critical question in this regard is how the impact of lane-changing varies when it is performed by an autonomous vehicle compared to a human-driven vehicle. This study aims to answer this question, along with other questions raised in the introduction, thereby addressing the significant gap in understanding the




(a) Model fitted distributions (b) Area under the curve

**Fig. 7.** Fitted Generalised Extreme Value distributions and area under the curves for different driving.

lane-changing impacts of both autonomous and human-driven vehicles.

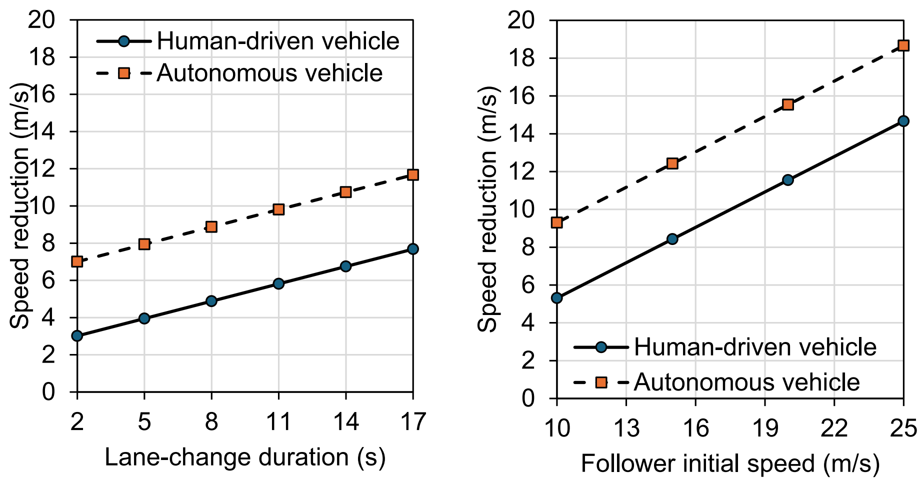
Using real-world Waymo trajectory data from Level 4 autonomous vehicles, this study finds that the impact of lane-changing by autonomous vehicles on the immediate follower is significantly higher than that caused by human-driven vehicles. Whilst this finding may seem counterintuitive—given that several studies argue that autonomous vehicles are safer (Wang et al., 2020; Hu et al., 2023)—past studies' findings should be interpreted with two caveats. Firstly, the focus of the majority of past studies remains on autonomous vehicles per se, rather than the safety of surrounding traffic. Secondly, the limited evidence derived from real data indicates a higher safety margin for car-following scenarios (Hu et al., 2023; Li et al., 2023), whereas the literature is devoid of any concrete evidence of the lane-changing impact of autonomous vehicles. Li et al. (2023) investigated car-following decision predictions of autonomous vehicles and their surrounding vehicles using a mature baseline model and a new conditional variational autoencoder framework. Using the Lyft Level 5 dataset, their study demonstrates the need to consider surrounding vehicle interactions and driving behaviour to make autonomous vehicles and surrounding traffic safer.

A natural question arises: why is the impact of autonomous vehicle lane-changing higher than that of human-driven vehicles? This question can be partly answered by examining the lane-changing execution behaviour of autonomous vehicles. A recent study found that autonomous vehicles take longer to complete their lane-changing manoeuvres compared to human-driven vehicles (Ali et al., 2024), suggesting safer and smoother manoeuvring. However, this safer manoeuvring of autonomous vehicles comes at the cost of higher deceleration for the immediate follower. In general, vehicles that take longer to change lanes significantly disrupt the traffic flow in both the current and target lanes (Zheng, 2014), which could be one reason for autonomous vehicles exerting a greater impact on the immediate follower.

Leveraging lane-change duration relationship with speed reduction from the random parameters duration model, this study further analyses how lane-changing impact varies for two vehicle types with varying lane-change duration. Specifically, using Equation (10) with mean values for all explanatory variables and keeping the dummy variable to the reference category (autonomous vehicle = 1), speed reduction values are obtained at varying lane-change duration (between minimum and maximum observed in the Waymo dataset) for autonomous and human-driven vehicles, which are shown in Fig. 8 (a). For example, at 8 s lane-change duration, the follower reduces its speed by about 4.88 m/s when the lane-changer is a human-driven vehicle, whereas the corresponding speed reduction for an autonomous vehicle is higher, that is 8.87 m/s. The maximum lane-change duration, that is 17 s, reduces the follower's speed by 7.68 m/s and 11.67 m/s for human-driven and autonomous vehicles, which is close to the maximum posted speed limit on sub-urban roads.

Another important parameter that governs the speed reduction behaviour of the follower is their initial speed. Fig. 8 (b) demonstrates the relationship between a follower's initial speed and its speed reduction when lane-changing is performed by human-driven and autonomous vehicles. In general, a higher initial speed of the follower is likely to lead to hard deceleration in response to lane-changing, which can significantly impact traffic flow characteristics and can increase crash risk (Loulizi et al., 2019; Kim et al., 2016). Kim et al. (2016) found a direct correlation between high crash rates and hard deceleration and reported that highway segments with deceleration greater than 4 $m/s^2$ are crash-prone. Several studies have also confirmed the effect of such hard deceleration in creating traffic disturbances and breakdowns (see Zheng et al. (2011) for an example).

The random parameters model confirms the presence of heterogeneity in autonomous vehicle lane-changing impact on speed reduction, suggesting that the speed reduction of the follower may increase or decrease when the lane-changer is an autonomous vehicle relative to a human-driven vehicle. A higher speed reduction in the case of autonomous vehicles can be associated with follower's lack of trust in these vehicles and their perception of making inaccurate decisions. De Miguel et al. (2019) found varying levels




(a) Speed reduction versus lane-change duration

(b) Speed reduction versus the follower's initial speed

**Fig. 8.** Variation in speed reduction with lane-change duration and the follower's initial speed.

of trust, uncertainty, and fear among the general public for autonomous vehicles. These factors could be associated with followers exhibiting hard decelerations to prioritise their safety. Several research studies have demonstrated that the presence of autonomous vehicles in the traffic stream alters the decision-making of other road users due to their different composition and unfamiliarity (Lin et al., 2025). For instance, Lin et al. (2025) observed extreme behaviours of human-driven vehicles in the presence of an automated truck platoon. Similarly, De Zwart et al. (2023) found increased penetration rate of autonomous vehicles leads to shorter time headways, smaller following distances, and reduced reaction times with faster acceleration at traffic lights.

Besides trust, the elevated risk may also be attributed to traffic dynamics like maintaining a smaller gap between the lane-changer and the follower and the speed of the lane-changer. For instance, drivers with a smaller gap to the lane-changer are likely to brake hard to yield to the lane-changer, resulting in higher crash risk. Ahmed et al. (1996) reported the gap between lane-changer and follower to be an influencing factor in the lane-changing decision-making process and directly affecting both lane-changer and follower. The speed of the lane-changer has also been found to increase the follower's crash risk, which is discussed in Section 5.2.

Further, a smaller speed reduction of the follower for autonomous vehicle lane-changing relative to human-driven vehicle lane-changing could be explained by autonomous vehicles selecting a large lag gap where they exert a smaller impact on the followers and making it safer for the follower.

Meanwhile, one may question from the follower's perspective whether it matters if the lane-changer is an autonomous vehicle or a human-driven vehicle. As autonomous vehicles are distinctly different from human-driven vehicles in terms of their size and appearance, drivers tend to engage in risk compensation behaviour through greater speed reduction. Several studies have investigated the effect of change in vehicle shape/size on lane-changing and car-following behaviour. For instance, Moridpour et al. (2010) reported significant differences in lane-changing characteristics between passenger cars and heavy vehicles, and these differences are more pronounced on arterials compared to freeways (Aghabayk et al., 2011). Further, the literature suggests that passenger vehicles following heavy vehicles are likely to amplify traffic disturbances, albeit with lower probability and magnitude compared to the dampening effect (Chen et al., 2016). Additionally, heavy vehicles tend to discourage lane changes, especially for vehicles following behind them (Chen et al., 2016).

From the safety perspective, crash risk analysis is performed using the developed extreme value (block maxima and peak over threshold) models. It is worth noting that this study intentionally did not compare these modelling approaches (as opposed to vast extreme value literature in traffic safety) for two reasons. Firstly, to justify which modelling approach is better (although abundant evidence in the literature, for example, Ankunda et al. (2024) and Hussain et al. (2022)), ground truth is required to perform this comparison. The ground truth serves as a global goodness-of-fit measure, whereas local goodness-of-fit measures like AIC do not hold as the sample size is different in these approaches. Given the ground truth is unavailable for crash risk associated with autonomous vehicles, justifying one approach over another is difficult. Secondly, the focus of this study is not on selecting the best extreme value modelling approach per se, but rather on using it to obtain relative crash risk, informing about crash risk for different vehicle types. Therefore, it does not really matter which modelling approach is used for relative crash risk analysis as long as both approaches yield a consistent outcome.

Although crash frequency from the extreme value model is not estimated, which is a common way of assessing model global goodness-of-fit, a cross-comparison of crash risk is performed. Ali et al. (2022a) used driving simulator data and found lane-changing crash risk for human-driven vehicles using the block maxima approach between 0.11 and 0.697. In another study (Wen et al., 2023), the block maxima approach yielded lane-changing crash risks of autonomous vehicles and human-driven vehicles using the Waymo data as 0.01 and 0.02, respectively. Our study found similar and comparable crash risk estimates with the relevant literature, thereby suggesting the appropriateness of the applied methods.

Fig. 9 presents the relative crash risk analysis results, whereby crash risk for the follower is calculated using Equation (7) for the block maxima and peak over threshold models. Results indicate that irrespective of modelling type, the crash risk for the follower is

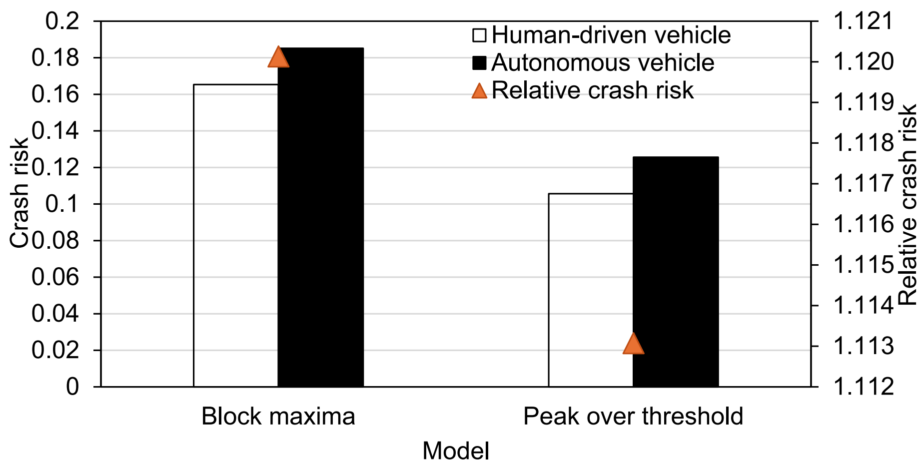



**Fig. 9.** Crash risk analysis for the follower for different vehicle types.

higher when the lane-changer is an autonomous vehicle compared to a human-driven vehicle. For instance, the crash risk using the block maxima model for human-driven vehicles is 0.165, whereas the corresponding crash risk for autonomous vehicles is 0.185. Further, relative crash risk is computed, reflecting the increase or decrease in crash risk for human-driven vehicle lane changes compared to autonomous vehicle lane changes—this concept is similar to the odds ratio that is frequently used in treatment effect analysis. A relative crash risk of greater (less) than one indicates a deleterious (desired) safety impact on the follower with an autonomous vehicle compared to a human-driven vehicle. Irrespective of the modelling type, the relative crash risk remains above one (see triangles in Fig. 9), indicating the followers are at a higher risk when the autonomous vehicle changes lanes.

The findings of this study are valuable for developing new, tailored models that capture the lane-changing impact of autonomous vehicles by considering follower behaviour, which has been recognised as a factor influencing model performance (Li et al., 2023). Given the significant impact demonstrated in this study, there is a need to develop impact-aware lane-change decision models capable of reproducing lane-changing-related traffic phenomena (anticipation, relaxation, and capacity drop). Further, the identified relationship between lane-change duration and its impact on the follower strengthens the argument for considering lane-change execution in microsimulation tools, where it is often assumed to be instantaneous. These findings also suggest the potential for developing a unified model for lane-changing behaviour, encapsulating the three intertwined aspects of decision-making, execution, and impact. Such a unified model is likely to be more robust and capable of reproducing macroscopic traffic phenomena more accurately.

From a policy perspective, given the continuous evolution of autonomous vehicles, it is crucial to properly understand their coexistence with human-driven vehicles. These vehicles will not be phased out overnight, and they will continue to share road space in a complex mixed-traffic environment. In general, human drivers are already uncertain and heterogeneous in their behaviours, and their response to autonomous vehicles heavily depends on their trust in technology, user acceptance, and sensation-seeking tendencies, further complicating their decision-making processes (Zhang et al., 2019). To mitigate aggressive responses to autonomous vehicles, policymakers and stakeholders could initiate public awareness campaigns and focus group programmes to educate human drivers about the capabilities and limitations of autonomous vehicles. Such initiatives will help build driver confidence, enabling them to respond to autonomous vehicles as they respond to human-driven vehicles.

## 7. Conclusions

This study examined the lane-changing impact of Level 4 autonomous vehicles on the immediate follower in the target lane using the publicly available Waymo dataset. The lane-changing impact was compared for the entire execution period, for which a wavelet transform-based local maxima lines method was applied to identify the starting and end points of a lane-changing manoeuvre. The Waymo data were also used to extract human-driven vehicle lane-changing data. Descriptive analyses were performed on different driving behaviour variables and two models were developed for understanding the impact on traffic flow efficiency and safety, namely, a linear regression model and extreme value theory models, respectively. As recently demonstrated (Ali et al., 2024) that autonomous vehicles also exhibit heterogeneous impact, a random parameters approach was adopted to capture unobserved heterogeneity associated with lane-changing impact.

Preliminary analyses indicated that the immediate follower in the target lane, when responding to an autonomous vehicle lane-changing compared to human-driven ones, exhibited higher acceleration noise and jerky driving, indicating aggressive braking. To establish the relationship among the determinants of lane-changing impact across different vehicle types, a correlated random parameters linear regression model was fitted to the speed reduction of the follower data, whereas extreme value models were developed and fitted to the deceleration rate to avoid collision (DRAC) data. The random parameters model found two parameters to be significant, with consistent signs and intuitive interpretation. The random parameter for the dummy variable representing autonomous vehicle lane-changing demonstrated heterogeneous effects on the follower's speed reduction relative to human-driven vehicles, indicating that not all autonomous vehicle lane changes impact the immediate follower similarly—this finding answered the first question. In other words, the follower's speed reduction in response to an autonomous vehicle lane change is heterogeneous and different from human-driven vehicles. Although autonomous vehicles are expected to reduce driver heterogeneity, their presence as lane-changers prompted varied speed reductions by the follower. The model also suggested that followers exhibited smaller speed reductions when responding to lane changes performed by human-driven vehicles, possibly due to drivers' familiarity with human lane-change decisions, which they routinely encounter and respond to. In contrast, autonomous vehicles are still relatively uncommon in the traffic stream and may not be perceived as safe. Further, the extreme value models indicated that crash risk for the follower of an autonomous vehicle is higher than for a follower of a human-driven vehicle, possibly due to the follower's aggressive response to the autonomous vehicle's lane-changing and lower trust in the technology. This finding also implicitly highlights the follower's self-inflicted hard braking, which is detrimental not only to their safety but also to other vehicles.

This study discussed the importance of lane-changing impact and embedding it into microscopic models to accurately reproduce macroscopic traffic phenomena like oscillations and capacity drops. Along this line, an important aspect is the relationship between lane-changing execution (and accepted gap size) and its impact on the immediate follower. This study confirms the significant impact of both these parameters on lane-changing impact (characterised by speed reduction), whereby longer lane-change durations and smaller lag gaps increase speed reduction—this finding answered the second question—which is likely to lead to a larger capacity drop and can impose a stronger disturbance to traffic flow. This impact was not fully explored at a macroscopic level in this study and is left for future research.

This study has the potential for several extensions. Firstly, it only focussed on assessing the discretionary lane-changing impact on the immediate follower using descriptive analysis and models. Two follow-up studies can be (a) comparing this impact with mandatory lane-changing impact, which is assumed to be stronger (Zheng, 2014) and (b) assessing the driving behaviour using car-following models as performed in Zheng et al. (2013). Secondly, due to data limitations, this study restricted its analysis to only one follower in the target lane. However, past studies have demonstrated that lane-changing impact can be experienced by multiple vehicles in both (current and target) lanes (Ali et al., 2020; Yang et al., 2019). Therefore, to fully evaluate autonomous vehicles' lane-changing impact, a large dataset with continuous data of multiple followers in both lanes should be used and analysed. Similarly, the current study only analysed driver's speed reduction; however, drivers may also decide to change lanes in response to autonomous vehicle lane-changing, which is a common strategy (Ali et al., 2019). Given our study is constrained by data where such information is unavailable due to the sensing capabilities of autonomous vehicles, future studies could model the behaviour of followers who opt to change lanes instead of braking when encountering autonomous vehicle lane-changing.

Thirdly, although this study captured unobserved heterogeneity through a random parameters approach, testing other advanced approaches like latent class models, correlated random parameters with heterogeneity in means and random parameters with heterogeneity in both means and variances with greater number of observations merits an investigation to determine which approach can better reveal heterogeneity and trace its sources. Fourthly, extreme value models were only utilised to calculate crash risk, whereas conventionally extreme value models are validated against observed crash frequency, which was not performed due to data unavailability. Future studies could validate extreme value model findings using the methodology proposed by Singh et al. (2024). Finally, this study used data from only one vehicle manufacturer with one Society of Automotive Engineers (SAE) Level 4, but several open-source datasets are available with different SAE levels. It would be interesting to investigate whether the impact observed with Waymo vehicles is comparable to that of other vehicles and levels, such as Argo, Lyft or Level 2 Mercedes or Tesla.

## CRediT authorship contribution statement

**Yasir Ali:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

## Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this paper, the author used ChatGPT 4 to conduct the proofreading and correct language errors wherever necessary. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

## 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

The author would like to thank Dr Danjue Chen for providing access to the processed Waymo dataset. A special mention to Mr Faizan Nazir for helping with extreme value models. The author also acknowledges the constructive comments of three anonymous reviewers whose comments greatly enhanced the quality of the paper.

## Data availability

Data will be made available on request.

## References

Aghabayk, K., Moridpour, S., Young, W., Sarvi, M., Wang, Y.-B., 2011. Comparing heavy vehicle and passenger car lane-changing maneuvers on arterial roads and freeways. *Transportation Research Record* 2260, 94–101.

Ahmed, K., Ben-Akiva, M., Koutsopoulos, H., Mishalani, R., 1996. Models of freeway lane changing and gap acceptance behavior. *Transportation and Traffic Theory* 501–515.

Ali, Y., Haque, M., Zheng, Z., 2022a. 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., Haque, M., Zheng, Z., 2022b. 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., Sharma, A., Chen, D., 2024. Investigating autonomous vehicle discretionary lane-changing execution behaviour: similarities, differences, and insights from Waymo dataset. *Analytic Methods in Accident Research* 42, 100332.

Ali, Y., Sharma, A., Zheng, Z., 2025. Empirical research on car-following and lane-changing: recent developments, emerging vehicle technologies' impact, and future research needs. *Transportation Research Interdisciplinary Perspectives* 31, 101368.

Ali, Y., Zheng, Z., Haque, M., Wang, M., 2019. A game theory-based approach for modelling mandatory lane-changing behaviour in a connected environment. *Transportation Research Part C* 106, 220–242.

Ali, Y., Zheng, Z., Haque, M., Yildirimoglu, M., Washington, S., 2020. Detecting, analysing, and modelling failed lane-changing attempts in traditional and connected environments. *Analytic Methods in Accident Research* 28, 100138.

Ali, Y., Zheng, Z., Haque, M., Yildirimoglu, M., Washington, S., 2021. CLACD: a complete LAne-Changing decision modeling framework for the connected and traditional environments. *Transportation Research Part C* 128, 103162.

Ankunda, A., Ali, Y., Mohanty, M., 2024. Pedestrian crash risk analysis using extreme value models: new insights and evidence. *Accident Analysis and Prevention* 203, 107633.
Arun, A., Haque, M., Washington, S., Sayed, T., Mannering, F., 2021. A systematic review of traffic conflict-based safety measures with a focus on application context. *Analytic Methods in Accident Research* 32, 100185.
Bevly, D., Cao, X., Gordon, M., Ozbilgin, G., Kari, D., Nelson, B., Woodruff, J., Barth, M., Murray, C., Kurt, A., 2016. Lane change and merge maneuvers for connected and automated vehicles: a survey. *IEEE Transactions on Intelligent Vehicles* 1 (1), 105–120.
Bhat, C., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B* 37 (9), 837–855.
Chen, D., Ahn, S., Bang, S., Noyce, D., 2016. Car-following and lane-changing behavior involving heavy vehicles. *Transportation Research Record* 2561, 89–97.
Coles, S., 2001. *An introduction to statistical modeling of extreme values*. Springer.
De Miguel, M., Fuchshuber, D., Hussein, A. and Olaverri-Monreal, C. Perceived pedestrian safety: Public interaction with driverless vehicles. IEEE intelligent vehicles symposium (IV), 2019. IEEE, 90-95.
De Zwart, R., Kamphuis, K., Cleij, D., 2023. Driver behavioural adaptations to simulated automated vehicles, potential implications for traffic microsimulation. *Transportation Research Part F* 92, 255–265.
Duell, M., Levin, M.W., Boyles, S., Waller, S., 2016. Impact of autonomous vehicles on traffic management: case of dynamic lane reversal. *Transportation Research Record* 2567 (1), 87–94.
Fountas, G., Sarwar, M., Anastasopoulos, P., Blatt, A., Majka, K., 2018. Analysis of stationary and dynamic factors affecting highway accident occurrence: a dynamic correlated grouped random parameters binary logit approach. *Accident Analysis and Prevention* 113, 330–340.
Gouy, M., Wiedemann, K., Stevens, A., Brunett, G., Reed, N., 2014. Driving next to automated vehicle platoons: how do short time headways influence non-platoon drivers' longitudinal control? *Transportation Research Part F* 27, 264–273.
Greene, W., 2012. LIMDEP Version 10/NLOGIT version 5. *Econometric Modeling Guide*.
He, J., Qu, J., Zhang, J., He, Z., 2022. The impact of a single discretionary lane change on surrounding traffic: an analytic investigation. *IEEE Transactions on Intelligent Transportation Systems* 24 (1), 554–563.
Hu, X., Zheng, Z., Chen, D., Sun, J., 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.
Hu, X., Zheng, Z., Chen, D., Zhang, X., Sun, J., 2022. Processing, assessing, and enhancing the Waymo autonomous vehicle open dataset for driving behavior research. *Transportation Research Part C* 134, 103490.
Huo, X., Leng, J., Hou, Q., Yang, H., 2020. A correlated random parameters model with heterogeneity in means to account for unobserved heterogeneity in crash frequency analysis. *Transportation Research Record* 2674 (7), 312–322.
Hussain, F., Ali, Y., Li, Y., Haque, M., 2024. Revisiting the hybrid approach of anomaly detection and extreme value theory for estimating pedestrian crashes using traffic conflicts obtained from artificial intelligence-based video analytics. *Accident Analysis and Prevention* 199, 107517.
Hussain, F., Li, Y., Arun, A., Haque, M., 2022. A hybrid modelling framework of machine learning and extreme value theory for crash risk estimation using traffic conflicts. *Analytic Methods in Accident Research* 36, 100248.
Kim, S., Song, T.-J., Rouphail, N., Aghdashi, S., Amaro, A., Gonçalves, G., 2016. Exploring the association of rear-end crash propensity and micro-scale driver behavior. *Safety Science* 89, 45–54.
Li, T., Han, X., Ma, J., Ramos, M., Lee, C., 2023. Operational safety of automated and human driving in mixed traffic environments: a perspective of car-following behavior. *Proceedings of the Institution of Mechanical Engineers, Part O* 237 (2), 355–366.
Lin, Z., Chen, F., Öztürk, İ., Merat, N., 2025. Overtaking automated truck platoons: effects of platoon organisations and traffic situation on driving behaviours of nearby manual vehicle drivers. *Transportation Research Part F* 109, 1208–1228.
Loulizi, A., Bichiou, Y., Rakha, H., 2019. Steady-state car-following time gaps: an empirical study using naturalistic driving data. *Journal of Advanced Transportation* 2019 (1), 7659496.
Mallat, S., Hwang, W., 1992. Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory* 38 (2), 617–643.
Mannering, F., Shankar, V., Bhat, C., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research* 11, 1–16.
Monteiro, F., Ioannou, P., 2023. Safe autonomous lane changes and impact on traffic flow in a connected vehicle environment. *Transportation Research Part C* 151, 104138.
Moridpour, S., Rose, G., Sarvi, M., 2010. Effect of surrounding traffic characteristics on lane changing behavior. *Journal of Transportation Engineering* 136 (11), 973–985.
Nakagawa, S., Schielzeth, H., 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4 (2), 133–142.
NHTSA. 2021. *Crashes, by First Harmful Event, Manner of Collision, and Crash Severity, 2020* [Online]. Washington, D.C. Available: https://cdan.nhtsa.gov/SASStoredProcess/guest [Accessed 12 December 2022].
Rad, S., Farah, H., Taale, H., Van Arem, B., Hoogendoorn, S., 2021. The impact of a dedicated lane for connected and automated vehicles on the behaviour of drivers of manual vehicles. *Transportation Research Part F* 82, 141–153.
Sharma, A., Ali, Y., Saifuzzaman, M., Zheng, Z. and Haque, M. Human factors in modelling mixed traffic of traditional, connected, and automated vehicles. Advances in Human Factors in Simulation and Modeling: Proceedings of the AHFE 2017 International Conference on Human Factors in Simulation and Modeling, July 17–21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA 8, 2018. Springer, 262-273.
Shin, Y.W., Kim, D.-K., Kim, E.-J., 2022. Impact of driver behavior and vehicle type on safety of vehicle platoon under lane change situation. *Transportation Research Record*, 03611981221127274.
Singh, S., Ali, Y., Haque, M., 2024. A Bayesian extreme value theory modelling framework to assess corridor-wide pedestrian safety using autonomous vehicle sensor data. *Accident Analysis and Prevention* 195, 107416.
Soni, S., Reddy, N., Tsapi, A., Van Arem, B., Farah, H., 2022. Behavioral adaptations of human drivers interacting with automated vehicles. *Transportation Research Part F* 86, 48–64.
Soriguera, F., Martínez, I., Sala, M., Menéndez, M., 2017. Effects of low speed limits on freeway traffic flow. *Transportation Research Part C* 77, 257–274.
Sultana, T., Hassan, H., 2024. Does recognizability of connected and automated vehicles (CAVs) platoons affect drivers' behavior and safety? *Transportation Research Part F* 103, 368–386.
Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y. and Caine, B. Scalability in perception for autonomous driving: Waymo open dataset. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020. 2446-2454.
Wang, J., Zhang, L., Huang, Y., Zhao, J., 2020. Safety of autonomous vehicles. *Journal of Advanced Transportation* 2020 (1), 8867757.
Wang, Y., Wang, L., Guo, J., Papamichail, I., Papageorgiou, M., Wang, F.-Y., Bertini, R., Hua, W., Yang, Q., 2022. Ego-efficient lane changes of connected and automated vehicles with impacts on traffic flow. *Transportation Research Part C* 138, 103478.
Washington, S., Karlaftis, M., Mannering, F., Anastasopoulos, P., 2020. *Statistical and econometric methods for transportation data analysis*. Chapman and Hall/CRC.
Wen, X., Huang, C., Jian, S., He, D., 2023. Analysis of discretionary lane-changing behaviours of autonomous vehicles based on real-world data. *Transportmetrica A* 1–24.
Xie, H., Ren, Q., Lei, Z., 2022. Influence of lane-changing behavior on traffic flow velocity in mixed traffic environment. *Journal of Advanced Transportation* 2022 (1), 8150617.
Yang, M., Wang, X., Quddus, M., 2019. Examining lane change gap acceptance, duration and impact using naturalistic driving data. *Transportation Research Part C* 104, 317–331.

Zhang, F. and Zhao, Q.-X. 2009. Effects of Lane Changing in Heterogeneous Traffic Flow. *ICCTP 2009: Critical Issues in Transportation Systems Planning, Development, and Management.*

Zhang, G. 2019. *Impact of the number of lane-changing times on throughput by fundamental diagram based on NGSIM data.* Master of Science, University of South Florida.

Zhang, T., Tao, D., Qu, X., Zhang, X., Lin, R., Zhang, W., 2019. The roles of initial trust and perceived risk in public's acceptance of automated vehicles. *Transportation Research Part C* 98, 207–220.

Zhang, X., Xu, J., Liang, Q., Ma, F., 2020. Modeling impacts of speed reduction on traffic efficiency on expressway uphill sections. *Sustainability* 12 (2), 587.

Zhao, D., Peng, H., Lam, H., Bao, S., Nobukawa, K., Leblanc, D. and Pan, C. Accelerated evaluation of automated vehicles in lane change scenarios. Dynamic Systems and Control Conference, 2015. American Society of Mechanical Engineers.

Zheng, L., Sayed, T., 2019. Comparison of traffic conflict indicators for crash estimation using peak over threshold approach. *Transportation Research Record* 2673 (5), 493–502.

Zheng, Z., 2014. Recent developments and research needs in modeling lane changing. *Transportation Research Part B* 60, 16–32.

Zheng, Z., Ahn, S., Chen, D., Laval, J., 2011. Freeway traffic oscillations: microscopic analysis of formations and propagations using wavelet transform. *Procedia-Social and Behavioral Sciences* 17, 702–716.

Zheng, Z., Ahn, S., Chen, D., Laval, J., 2013. The effects of lane-changing on the immediate follower: anticipation, relaxation, and change in driver characteristics. *Transportation Research Part C* 26, 367–379.

Zheng, Z., Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C* 25, 18–33.