

Evaluating the Safety and Efficiency Impacts of Forced Lane Change with Negative Gaps Based on Empirical Vehicle Trajectories

Kequan Chen

Graduate Research Assistant
Department of Transportation, Southeast University
Southeast University Road #2, Nanjing, China, 211189
Email: chenkequan@seu.edu.cn

Zhibin Li

Ph.D., Professor
Department of Transportation, Southeast University
Southeast University Road #2, Nanjing, China, 211189
Email: lizhibin@seu.edu.cn

Pan Liu (Corresponding Author)

Ph.D., Professor
Department of Transportation, Southeast University
Southeast University Road #2, Nanjing, China, 211189
Email: liupan@seu.edu.cn

Victor L. Knoop

Ph.D., Associate Professor
Civil Engineering of Technology, Delft University of Technology
Stevinweg 1, Delft, 2628CN
Email: v.l.knoop@tudelft.nl

Yu Han

Ph.D., Associate Professor
Department of Transportation, Southeast University
Southeast University Road #2, Nanjing, China, 211189
Email: Yuhan@seu.edu.cn

Yiru Jiao

Graduate Research Assistant
Civil Engineering of Technology, Delft University of Technology
Stevinweg 1, Delft, 2628CN
Email: y.jiao-1@tudelft.nl

Abstract:

A lane-changing (LC) maneuver may cause the follower in the target lane (new follower) to decelerate and give up space, potentially affecting crash risk and traffic flow efficiency. In congested flow, a more aggressive LC maneuver occurs where the lane changer is partially next to the new follower and creates negative gaps, namely negative gap forced LC (NGFLC). Although NGFLC forms the foundation of sideswipe crashes, little has been done to address its impacts and the contributing factors. To tackle this issue, a total of 15,810 LC trajectory samples are extracted from three drone videos at different locations. These samples are categorized into NGFLC and normal LC groups for comparative analysis. Five commonly used conflict indicators are extended into two-dimensional to evaluate the crash risk of LC maneuver. The change of time gaps during LC maneuver are examined to quantify the impact of LC on traffic flow efficiency. We find that NGFLCs significantly increase crash risk, reflected by the number of hazardous LC events and potential crash areas compared to normal LC. Additionally, results reveal that both the lane changer and the new follower tend to maintain a larger time gap after NGFLCs. Factors including time headway, relative speed, and historical gaps in the target lane significantly affect NGFLC incidence. Once the movement of the leader in the original lane is taken into account, the prediction accuracy improves from 81% to 91%. The transferability tests indicate that the findings about the negative impact of NGFLC and the accuracy of its prediction model are consistent across different locations. These findings hold implications for driving assistance systems to better predict and mitigate NGFLCs.

Keywords: Forced lane-changing behavior; lane-changing impact; traffic safety; flow efficiency; trajectory data analysis

1. Introduction

A lane-changing (LC) maneuver typically involves interactions between the lane changer and three surrounding vehicles: the leading vehicle in the original lane (referred to as the original leader), the following vehicle in the target lane (new follower), and the leading vehicle in the target lane (new leader). Numerous studies have emphasized the adverse effects of LC maneuvers on traffic safety and traffic flow efficiency (Yun et al., 2017; Hess et al., 2020; Chen et al., 2021, Reinolsmann et al., 2021). For example, Pei et al. (2010) report that over 30% of crashes are linked to improper LC maneuvers. Zheng et al. (2011a) found that LC maneuvers could lead to speed reductions in the target lane, subsequently causing stop-go waves and traffic flow disturbances. Laval and Daganzo (2006) noted that slow-moving LC vehicles with limited acceleration are a major cause of capacity drops, resulting in decreased traffic flow rates. Given these negative impacts induced by LC maneuvers, investigating LC maneuvers is of interest in this study, particularly focusing on those that are risky and inefficient.

Previous research has identified several improper LC types, such as forced LC (Yang et al., 2016), unsuccessful LC (Ali et al., 2020a), and cut-in LC (Wang et al., 2019). Each LC type has demonstrated its profound impact on traffic flow. However, these studies typically assume that the lag gap in the target lane always exceeds a critical positive value during the LC maneuver, implying sufficient space for the lane changer. Indeed, this assumption of a critical positive gap does not always hold true in real-world conditions, particularly in congested traffic flow. During the study of the massive amount of empirical LC maneuvers from expressways in China, numerous negative gap forced LC (NGFLC) maneuvers were observed. The NGFLC is defined as a unique type of aggressive LC during which negative lag gaps exist at some moments. **Figure 1** visually depicts an NGFLC case. Specifically, from time T to $T+3$ in this example, the new follower refuses to yield and attempts to close the gap, creating a negative gap with the lane changer. Subsequently, the new follower decelerates from $T+4$ to $T+5$, enlarging the lag gap to avoid a sideswipe crash and allowing the lane changer to complete the LC maneuver.

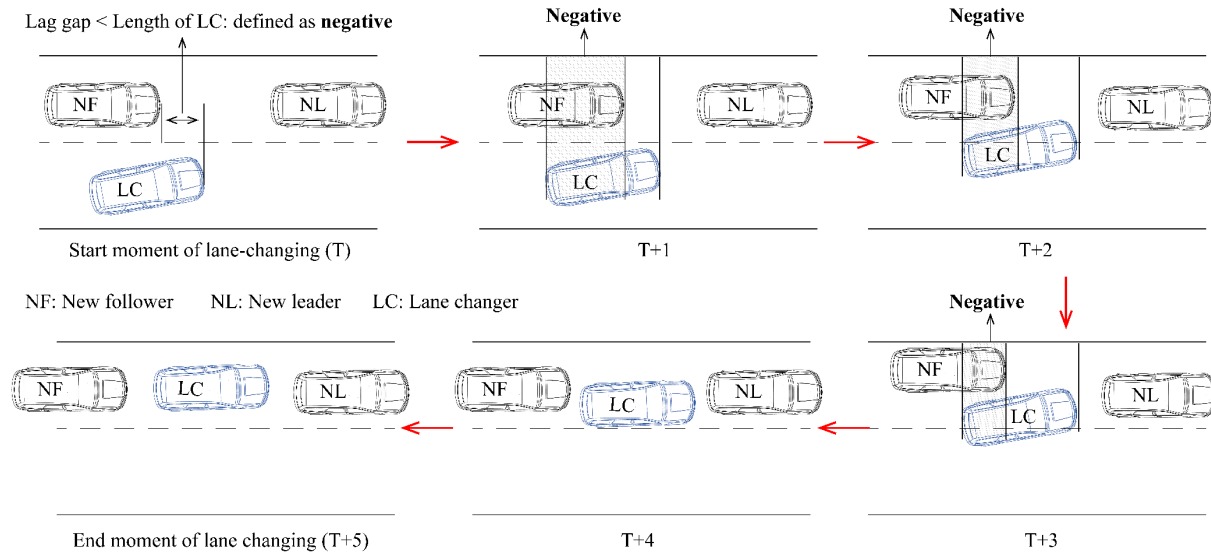


Figure 1 Illustration of forced lane-changing with a negative gap

Although NGFLC forms the basis of sideswipe crashes, the literature has not fully investigated its identification procedure and influence on crash risk. Given that both the new follower and the lane changer undergo an aggressive LC maneuver, it is unclear how driver

behavior changes and affects the traffic flow efficiency. Naturally, if NGFLC negatively impacts traffic flow compared to normal LCs, understanding the triggers leading to NGFLC becomes imperative.

Motivated by these research gaps, the primary objective of this study is to comprehensively examine NGFLC from its identification, consequences, and formulation. More specifically, we utilize three trajectory datasets from different locations and times to investigate the frequency of NGFLC. Five widely used SSMs are extended to two-dimensional to examine the impact of NGFLC on crash risk. Then, a method to quantify how the NGFLC affects the traffic flow efficiency is proposed. We also compare these findings with those derived from normal LC samples to gain a deeper understanding of the outcomes of NGFLC. Finally, Binary logit models are developed to investigate which factors and how they affect an LC decision to become an NGFLC. The research results are the first attempt to illuminate the impact of NGFLC on microscopic-level traffic flow, which has been previously overlooked. It can aid in driving assessment systems to better understand, predict, and avoid NGFLC.

2. Literature review

This section presents a detailed review of previous studies from two perspectives. In the first subsection, we review studies concerning forced LC, given its close relation to NGFLC. In the second subsection, we review methods employed to measure the influence of an LC maneuver on traffic flow.

2.1 Modeling forced Lane-changing maneuver

Hidas (2002) coined the term “forced LC (FLC)” to describe an LC maneuver based on the principle of driver courtesy. In this context, the lane changer sends a forced request to the new follower, who subsequently reduces speed and increases the initial gap to accommodate the request. His seminal work involved developing and implementing an FLC algorithm within the Simulation of Intelligent TRAnsport System (SITRAS). Simulation results from SITRAS suggested that the integration of FLC significantly enhances the realism of the flow-speed relationship. A subsequent study by Sun and Kondyli (2010) corroborated the effectiveness of FLC by developing a framework for an urban LC system that incorporated FLC, demonstrating its high accuracy in predicting travel time under congested flow conditions.

Considering the significance of FLC and its prevalence in congested flow, several rule-based models have been developed to classify and predict FLC. For instance, Sun and Elefteriadou (2014) proposed an FLC model using a pre-defined rule that guides the new follower to either reject or accept the lane changer’s request when the gap in the target lane is less than the critical gap. Qu and Wang (2015) developed an FLC model in which the new follower slows down and increases the gap based on the location of the lane changer. Yang et al. (2016) categorized FLC and other LC maneuvers according to the time gap in the target lane. Specifically, an LC maneuver is deemed as FLC if the time gaps in the target lane remain constant or diminish during the LC maneuver. Similarly, Chauhan et al. (2022) suggested that the LC maneuver can be classified as FLC if the spacing headway between the new follower and the new leader expands during the LC maneuver.

These studies above have made great contributions to modeling FLC, especially when the initial gap is larger than the critical gap. However, in real-world traffic situations, the critical gap may fluctuate according to different traffic conditions (Yang et al., 2019). Research is lacking on such FLC situations where the initial gap or the gap during the LC maneuver is continuously less than the critical gap (see Figure 1). To address this gap, Zhao et al. (2013) introduced the

concept of transline ride (TLR), in which the lane changer may encounter multiple new followers until the LC maneuver is completed. However, they did not propose a method for identifying NGFLC, and instead, the NGFLC samples in this study were manually extracted by watching video footage. Hence, developing a procedure for automatically classifying the NGFLC from the entire traffic flow trajectory is crucial, which will benefit future analysis.

2.2 Lane-changing impact

The impact of LC maneuvers on crash risk has always been an area of interest in the field of transportation. Numerous models have been proposed to assess the crash risk associated with LC maneuvers. For example, [Yang et al. \(2019\)](#) investigated the crash risk between the new follower and the lane changer based on TTC and found that the crash risk is higher before the lane changer crosses the lane line. [Park et al. \(2018\)](#) proposed a lane change risk index using surrogate safety measures to evaluate the crash risk of LC maneuvers. Comparison results revealed that the lane changer tends to be involved in a rear-end collision with the original leader in a work zone section. [Xing et al. \(2019\)](#) analyzed the LC crash risk on toll roads based on a two-dimensional TTC and concluded that manual toll collection vehicles notably heighten the crash risk. Considering various LC types, [Chen et al. \(2021\)](#) divided the LC process into sixteen patterns based on the vehicle type, discovering a robust correlation between crash risk and factors such as gap distance, longitudinal speed, and acceleration. Discretionary and mandatory LC have been investigated separately for their crash risk ([Ali et al., 2020](#), [Ali et al., 2019](#)). In addition to examining the impact of LC on crash risk, various models have been proposed to identify the factors contributing to risky LC behavior, including statistic models ([Adanu et al., 2021](#), [Ali et al., 2022](#), [Yang et al., 2019](#), [Ouyang et al., 2023](#)), and data-driven models ([Chen et al., 2019](#), [Shi et al., 2019](#)).

On the other hand, LC maneuver in congested flow can disrupt traffic flow capacity due to slow insertion and abundant acceleration ([Duret et al., 2011](#); [Marczak et al., 2015](#), [Leclercq et al., 2016](#), [Chen et al., 2023](#)). Empirical studies have indicated that LC maneuvers might induce speed reductions of the new follower, leading to oscillation. 12 out of 18 oscillations in the NGSIM dataset are caused by LC maneuvers ([Zheng et al. \(2011b\)](#)). [Chen et al. \(2022\)](#) observed that some new followers might reduce their speed when they notice a lane changer who starts to shift its lateral position in the adjacent lane.

The abovementioned studies could serve as valuable references for evaluating the impact of LC maneuvers. [Lee et al. \(2006\)](#) suggested that lateral overlap between the lane changer and the new follower occurs due to inappropriate decision-making by the lane changer. For simplicity, they treated this as the difference in flow ratio between the target and original lanes. They incorporated it into the Average Flow Ratio index proposed by [Chang and Kao \(1991\)](#) to estimate the crash risk. Simulation results reported by [Ben-Akiva et al. \(2006\)](#) implied that FLC significantly influences the macroscopic outcomes of traffic flow. [Zhao et al. \(2013b\)](#) found that the basic LC characteristics, such as LC time, LC velocity, and the number of affected vehicles, significantly differ between FLC and other LC maneuvers. However, a detailed understanding of the microscopic consequences and causes of NGFLC remains unclear.

3. Data and negative gap forced lane-changing detection

3.1 Data collection

The research team conducted drone video recordings at three merging segments in China. The first site, Yintian Road in Nanjing, was recorded over a week from December 12 to December 19, 2023. The recording period for each day was chosen during the morning peak

hours from 7:30 am to 9:30 am to capture a higher frequency of LC maneuvers. The data from this site serves as the modeling dataset, namely dataset 1. The other two sites aim to assess whether the presence, impact, and occurrence mechanism of NGFLC, as obtained in the modeling dataset, can be observed at different locations and cities. Specifically, the second site was also located in Nanjing but on a different expressway known as Xianlin Road. For this site, the drone recordings were conducted on July 28, 2021, from 7:30 am to 9:00 am. The third merging segment was located in another city, Chongqing. The data collection for this site was conducted during peak hours on December 15, 2023, from 7:30 am to 10:30 am. All recordings were made under sunny weather conditions. Drones were flown at an altitude of 300m, capturing 400m road segments. **Figure 2** illustrates the locations of these three study sites. The following paragraphs provide a brief overview of the trajectory data extraction procedure. For detailed methodology, readers are referred to our previous studies ([Chen et al., 2021](#), [Wan et al., 2020](#)). The trajectory extraction method includes five steps: video stabilization, vehicle detecting, vehicle tracking, lane detecting, and raw data smoothing.

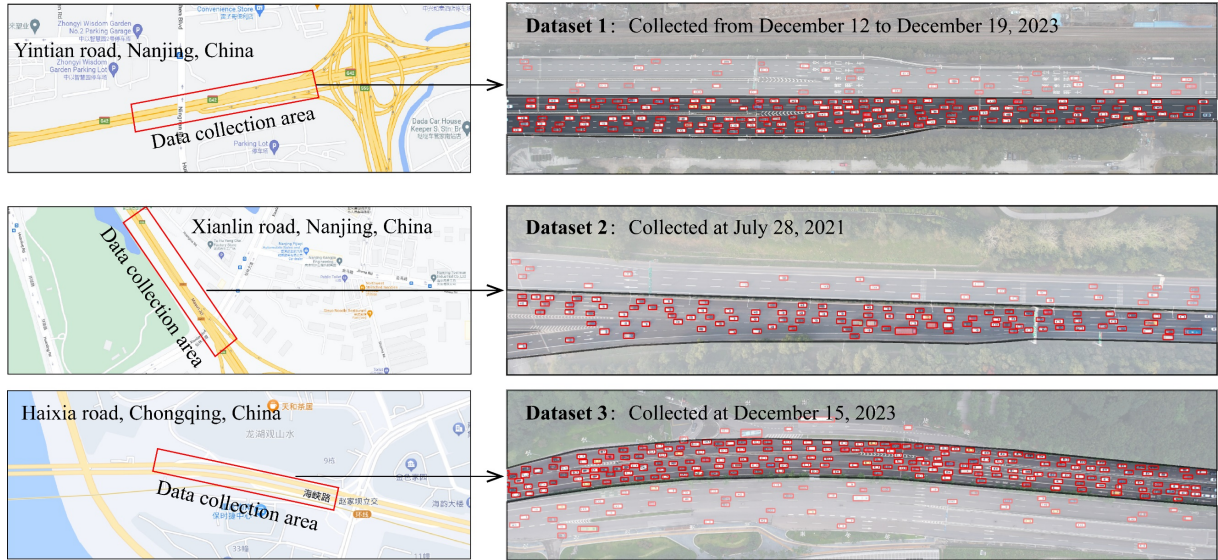


Figure 2 Three study sites

Firstly, the stabilization of raw drone video is required due to potential drone camera shaking caused by control maneuvers or wind during high-altitude recording. Each video's initial frame is used as a baseline frame. Then, the Temporally Robust Global Motion Compensation model is adopted ([Li et al. 2019](#)) to calculate transformation matrixes, aligning subsequent frames to the baseline frame. This process yields adjusted frames, ensuring vehicle movements in the video remain unaffected by the instability of the camera. **Figure 3** illustrates an example of this process, where the left image is the baseline frame, the center shows a mid-recording frame, and the right image is the adjusted frame. In **Figure 3(b)**, a noticeable rotation of the road compared to the baseline frame is observed. The stabilization process aligns the road position in the adjusted frame with the baseline frame, as shown in **Figure 3(c)**.

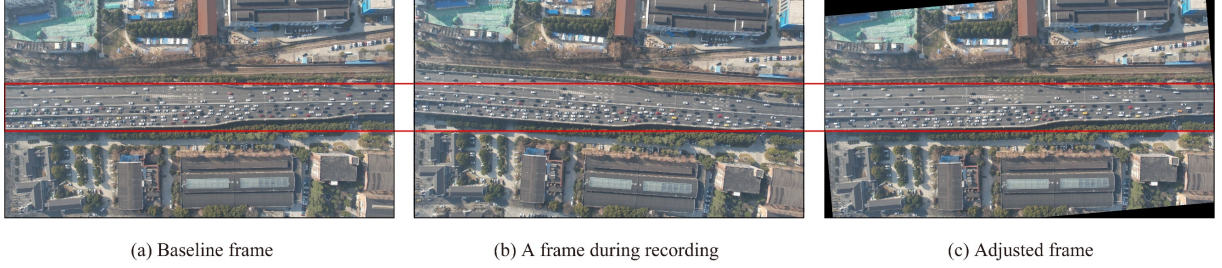


Figure 3 An example of the stabilization process

Next, the You Only Look Once v4 (YOLOv4) algorithm is employed for vehicle detection in each frame. The YOLOv4 model is trained on 500 drone pictures from the first recording site, with vehicles manually marked within rectangular boundaries. Using the well-trained model, the four corner points of the rectangle bounding box for each vehicle can be obtained. To quantify its detection accuracy, one vehicle will be marked with two bounding boxes, one for manually marking and one for automatically detecting. The detection error of the trained YOLOv4 model is measured by the Mean Squared Error (MSE) of the differences in the center points between detected and ground truth bounding boxes. For the first recording site, 20 drone images are randomly selected from drone videos. In these images, a total of 1,059 vehicles are manually marked and automatically detected by YOLOv4 with bounding boxes, respectively. Of these, the MSE for the trained YOLOv4 is 0.04m. For the other two sites, the MSE values were 0.06m and 0.09m across 924 and 1,142 vehicles in randomly selected images, respectively. These test results demonstrate that our detecting model is good enough to detect vehicle positions accurately.

Subsequently, the movement of each vehicle is tracked frame by frame. Considering the drone's camera records at 30 fps, the time step for tracking the vehicle trajectory is set at 1/30 second. Traffic lanes are identified using the method proposed by [Liu et al. \(2020\)](#). Thus, each vehicle can be assigned to its corresponding lane, which facilitates the identification of the LC event.

The raw trajectories derived from the previous steps contain considerable white noise. Minor positional deviations over short time intervals can cause fluctuations in speed and acceleration profiles. To eliminate noise in these raw trajectories, we employ the Wavelet Transform method recommended by [Chen et al. \(2021\)](#). This approach ensures that the location, speed, and acceleration of each vehicle at every time interval remain within realistic limits. To check the accuracy of the smoothed trajectories, we randomly selected two hours of trajectory data, and compared the range of acceleration before and after smoothing with the recommended range (-6m/s^2 to 5m/s^2) ([Montanino and Punzo, 2013](#), [Thiemann et al., 2008](#)). It is found that 98.25% of time intervals of acceleration are out of the recommended range before smooth. After smoothing, only 0.067% of time intervals are outside the recommended range, indicating that noises in the raw data have been well addressed, resulting in a smoothed trajectory conducive to high-precision analysis. **Figure 4** shows an example of obtained vehicle trajectories.

In total, we obtain 67,519 vehicles' trajectories from 14 hours of recordings at study site 1. Additionally, 7,071 vehicles' trajectories from 1.5 hours of recordings at study site 2 and 13,154 vehicles' trajectories from 3 hours of recordings at study site 3 are obtained. Each vehicle trajectory data includes the lateral location (m), longitudinal location (m), speed (m/s), acceleration (m/s^2), lane number, space headway (m), and time (s). The speed and acceleration for each vehicle are calculated based on the location information at each time interval.

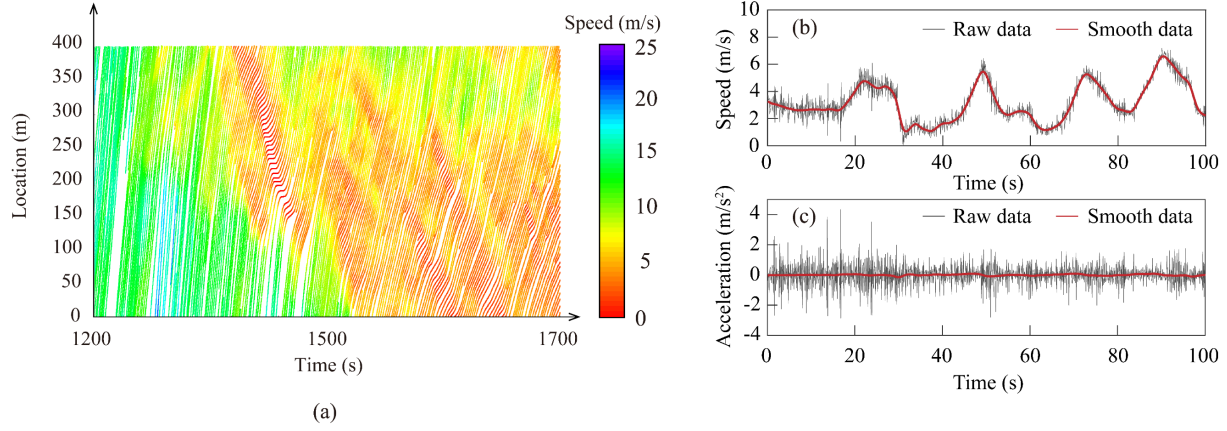


Figure 4 (a) An example of vehicles' trajectories in a single lane; (b) speed and (c) acceleration of a vehicle before and after smoothing

3.2 Identification of negative gap forced lane-changing

This section presents a procedure for identifying NGFLG. The identification process consists of two steps: (1) extracting LC samples from the trajectory dataset and (2) establishing a criterion to determine the occurrence of negative gaps in the extracted LC samples.

3.2.1 Lane-changing identification

This subsection details the extraction procedure for LC samples. Given that an LC process is a time series trajectory, we set three critical time points to locate it, which are the start, insertion, and end moments. These time points are relevant and will be defined below. **Figure 5 (a)** provides an example of these points. After describing these points, we will show how to extract them numerically.

- The start moment of the LC maneuver is marked by "A" in **Figure 5 (a)**. Consistent with previous studies (Ali et al., 2018; Yang et al., 2019), we characterized the LC start moment as the instance when the lane changer suddenly changes its lateral movement in the original lane.
- The insertion of the target lane is marked by "B" in **Figure 5 (a)**. This is defined as the instance when the center of the LC vehicle reaches the edge of the target lane.
- The end moment of the LC maneuver is marked by "C" in **Figure 5 (a)**. After the lane changer is successfully inserted into the target lane, it will adjust the speed and lateral position to the center of the target lane. Finally, the LC maneuver ends when the lane changer's movement attains stability in the target lane. Therefore, we determine this stabilization point as the end moment of the LC maneuver.

The Wavelet Transform with a Mexican hat function is employed to identify the start and end moments of the LC maneuver automatically (time points A and C). This method has demonstrated excellent efficiency in capturing spatial information in time series data (Zheng et al., 2011; Chen et al., 2014, Ali et al., 2020). Specifically, the lateral moment of the lane changer is transformed into a wavelet-based curve, as shown in **Figure 5 (b)**. The curve's maximum and minimum points correspond to the first and second sudden shifts in the lateral moment, representing the start and end moments of the LC maneuver, respectively. Since our trajectory extraction provides the location of each lane edge, the time point B for each LC sample can be easily obtained by calculating the lane changer's distance to the lane edge equation. Once the

distance reaches zero, time point B is identified. Following this procedure, we obtained 20,748 LC events in dataset 1, 2,232 LC events in dataset 2, and 1,967 LC events in dataset 3.

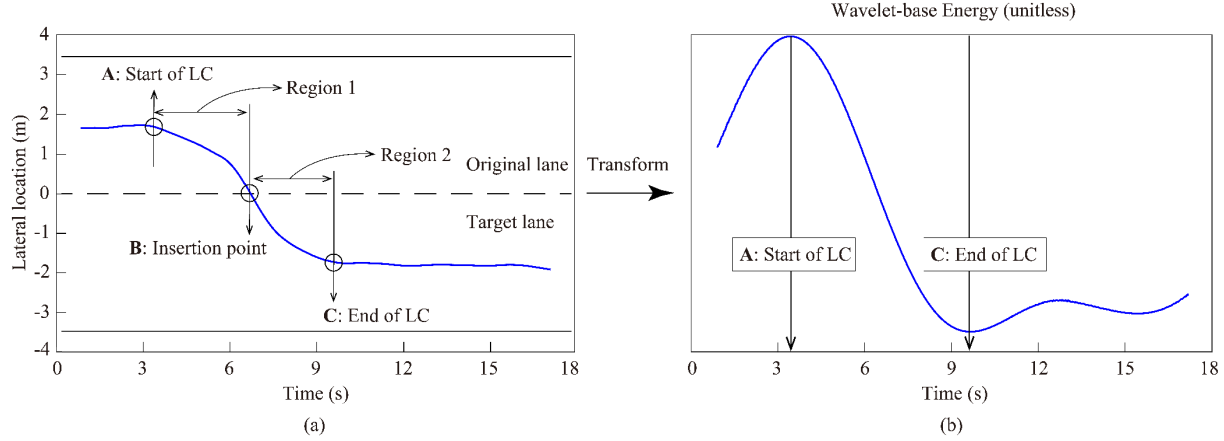


Figure 5 Determination of three critical time points during the LC maneuver: (a) lane changer's lateral trajectory; (b) transferred curve based on Wavelet-Transform

3.2.2 Identification of negative gap forced lane-changing

With a focus on the LC's impact, a complete LC sample should include the lane changer and the surrounding vehicles. Therefore, for each LC event extracted previously, we correspondingly retrieve vehicles' trajectories in the target and original lanes. Hence, each complete LC sample incorporated four vehicles: the lane changer, the new follower, the new leader, and the original leader. For all the candidate LC samples, we implement the following two rules to filter out unsuitable samples:

- Rule 1: The new follower and the new leader must be the same vehicles from time points A to C.
- Rule 2: The original leader must be in the same vehicle from time points A to B.

Regarding an LC sample, rule 1 ensures that interactions between the lane changer and the vehicles in the target lane are not affected by other LC maneuvers. As a result, 3,841 LC samples from dataset 1, 567 LC samples from dataset 2, and 349 LC samples from dataset 3 are removed after applying this rule.

Rule 2 is a relaxed version of Rule 1 that allows us to gain insights into how the original leader affects the lane changer's behavior. By applying these rules, we end up with 12,815 LC samples in dataset 1, 1,617 LC samples in dataset 2, and 1,378 LC samples in dataset 3.

Next, we develop an indicator called ΔLon to identify whether the LC vehicle has a negative gap with the new follower:

$$\Delta Lon(t) = X_{LC}(t) - X_{NF}(t) \quad (1)$$

where $X_{LC}(t)$ is the rear point of the LC vehicle at time t , and $X_{NF}(t)$ is the front point of the new follower at time t .

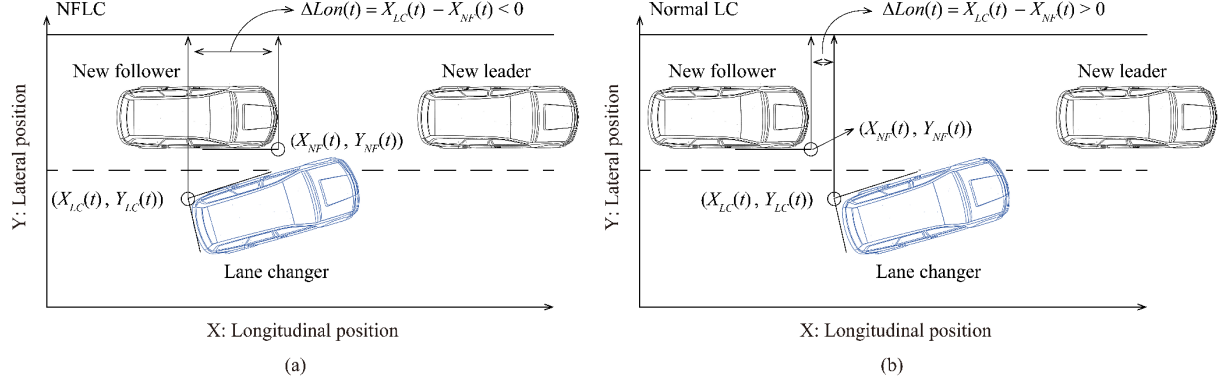


Figure 6 Two examples: (a) an example of NGFLC; and (b) an example of normal LC

Figure 6 (a) and (b) illustrate two examples of $\Delta Lon(t)$ with negative and positive values, respectively. A negative value of $\Delta Lon(t)$ indicates that the lane changer's rear bumper is behind the new follower's front bumper at time t , implying that the two vehicles are laterally overlapped. If any negative value of ΔLon exists throughout the entire LC maneuver, we classify this LC maneuver as NGFLC. On the contrary, if ΔLon remains positive throughout the entire LC maneuver, it is classified as a normal LC.

Table 1 presents the classification results of each trajectory dataset. We find that 33%, 36%, and 28% of LC samples are identified as NGFLC in dataset 1, dataset 2, and dataset 3, respectively. These similar and considerable proportions of NGFLCs highlight that NGFLCs are prevalent in the traffic flow. In the following section, we will delve into the consequences of NGFLC and explore how it differs from normal LC.

Table 1 Identification results of NGFLC at three datasets

	Dataset 1	Dataset 2	Dataset 3
Total LC samples	12815	1617	1378
Normal LC samples	8586	1026	992
NGFLC samples	4229	591	386
Percentage of NGFLC	33%	36%	28%

4. Methodologies

To achieve a comprehensive understanding of the NGFLC behavior and its differences from normal LC behavior, safety and efficiency metrics are initially introduced in this section. Specifically, five surrogate safety measures (SSMs) are introduced in Section 4.1, and a method to quantify the impact of LC on traffic flow efficiency is proposed in Section 4.2. To unveil which factors and how they characterize an LC decision as an NGFLC maneuver, binary logit models are established in Section 4.3.

4.1 Safety measurements

4.1.1 Time-based SSMs

In previous studies, the conventional TTC has been widely utilized to assess the potential crash risk between two consecutive vehicles (Biswas et al., 2021; Ding et al., 2019; Gu et al., 2019). However, this approach assumes that the two vehicles are moving in the same lane, which may not always apply in some scenarios where collisions occur at different angles. To address

this limitation, researchers have extended safety measures into two dimensions, yielding satisfactory results (Venthuruthiyil and Chunchu, 2022; Ward et al., 2015; Xing et al., 2019). Given that our study focuses on the LC scenario, where conflict with angles could arise, we extend the conventional TTC into two-dimensional to evaluate crash risk during the LC maneuver.

In the two-dimensional TTC, two successive vehicles are represented as rectangular bounding boxes. Similar to the conventional TTC, these bounding boxes move in accordance with their current speed and direction. The two-dimensional TTC measures the time it would take until the first intersection moment of these two bounding boxes, quantifying the current risk level. Clearly, a smaller TTC value suggests a more dangerous situation. On the other hand, if two bounding boxes do not intersect, the TTC value is considered to be infinite. A threshold of TTC, abbreviated as TTC^* , is used to differentiate between safe and conflict situations. Herein, we use four TTC^* values ranging from 1s to 4s with 1s intervals. An LC sample is categorized as a conflict if the TTC value is less than the TTC^* at any moment during the LC maneuver.

Considering that LC maneuvers typically last from 3s to 15s (Moridpour et al., 2010b), we adopt two variants of two-dimensional TTC called Time Exposed TTC (TET) and Time Integrated TTC (TIT) to evaluate the crash risk for a complete LC sample. Specifically, the TET value is determined by the total dangerous duration wherein the TTC value falls below the TTC^* during the LC maneuver (Minderhoud and Bovy, 2001), which can be calculated as follows:

$$TET_i = \sum_{t=0}^N \delta_i(t) \cdot \tau \quad (2)$$

where N is the number of time intervals, each time interval is τ (0.03s in our study), during the i -th LC maneuver. $\delta_i(t)$ is a switching variable between 0 and 1. Specifically, if $TTC_i(t)$ less than TTC^* , $\delta_i(t)$ equals 1, otherwise, $\delta_i(t)$ equals 0.

TIT measures the integral of the TTC profile below the threshold during the LC maneuver, allowing to express the crash risk severity, which can be calculated as follows:

$$TIT_i = \sum_{t=0}^N [TTC^* - TTC_i(t)] \cdot \tau, \quad \forall 0 \leq TTC_i(t) \leq TTC^* \quad (3)$$

4.1.2 Deceleration-based SSMs

This study uses two deceleration-based SSMs to evaluate the kinematic characteristics for crash risk avoidance during the LC maneuver (Cunto and Saccomanno, 2008). The first conflict indicator, Deceleration Rate to Avoid a Crash (DRAC), measures the required deceleration rate of the following vehicle to match the speed of the leading vehicle to avoid a crash. Traditionally, this indicator quantifies the crash risk between two vehicles moving in the same direction, which can be calculated as follows:

$$DRAC = \begin{cases} \frac{(V_f - V_l)^2}{D}, & \text{if } V_f > V_l \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where V_f and V_l are the speed of the following and leading vehicles, respectively, D is the distance between the front bumper of the following vehicle and the rear bumper of the leading vehicle.

In the LC scenario, the formula of DRAC needs to be redefined. Specifically, when the rectangular bounding boxes of two vehicles do not intersect along their current direction and speed, meaning that their two-dimensional TTC value is infinite, the DRAC equals zero. On the other hand, if there is a potential intersection position, the distance from the following bounding box to that intersection point will be considered as the remaining distance, denoted as D_{2d} . Under this context, the DRAC is defined as the required deceleration rate for the following vehicle to completely stop before reaching the potential intersection position, which can be rewritten as follows:

$$DRAC_i(t) = \begin{cases} \frac{[V_f(t)]^2}{D_{2d}(t)}, & \text{if } TTC_i(t) \in \mathbb{R} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

$$D_{2d}(t) = TTC_i(t) \cdot V_f(t) \quad (6)$$

When the value of DRAC exceeds the maximum available deceleration rate (MADR), the following vehicle is considered to be in a traffic conflict. Similar to previous studies, four MADR values ranging from 2m/s^2 to 5m/s^2 with 1m/s^2 interval are used in this study. An LC sample is categorized as a conflict LC sample if the DRAC value is larger than the MADR value at any moment during the LC maneuver.

An improved indicator is adopted to evaluate the probability of conflict moments identified by DRAC during the LC maneuver, namely the Crash Potential Index (CPI). The formula of CPI is given as follows:

$$CPI_i = \frac{\sum_{t=0}^N \theta_i(t) \cdot \tau}{N\tau} \quad (7)$$

where $\theta_i(t)$ is a switching variable between 0 and 1, specifically, if $DRAC_i(t)$ is large than MADR, $\theta_i(t)$ equals 1, otherwise, $\theta_i(t)$ equals 0.

4.2 Efficiency measurements

In previous studies, the term “void” refers to the extra time gap that arises during the LC maneuver. It has been used to assess the impact of LC on traffic flow capacity (Chen and Ahn, 2018; Laval and Daganzo, 2006; Leclercq et al., 2016). In practice, the void may be closed or even reduced for several reasons, such as aggressive behavior or congested flow (Chen et al., 2020). Building on the definition of the void, we measure the change in time gap during the LC maneuver to capture the impact of LC on traffic flow efficiency for both the lane changer and the new follower.

We define the void created by the lane changer as ΔO_{LC} , which can be calculated as below:

$$\Delta O_{LC} = O_{NL-LC} - I_{NL-LC} \quad (8)$$

where O_{NL-LC} is the final time gap between the new leader and the lane changer at the end moment of LC, and I_{NL-LC} is the initial time gap between the new leader and the lane changer at the insertion moment.

Figure 7 provides two examples for Equation (8). **Figure 7 (a)** shows a scenario where the lane changer increases the time gap after the LC maneuver, resulting in a positive value of ΔO_{LC} . This positive value reduces the traffic flow capacity. **Figure 7 (b)** shows an opposite situation where the lane changer decreases the time gap, resulting in a negative value of ΔO_{LC} . This negative value enhances the traffic flow capacity.

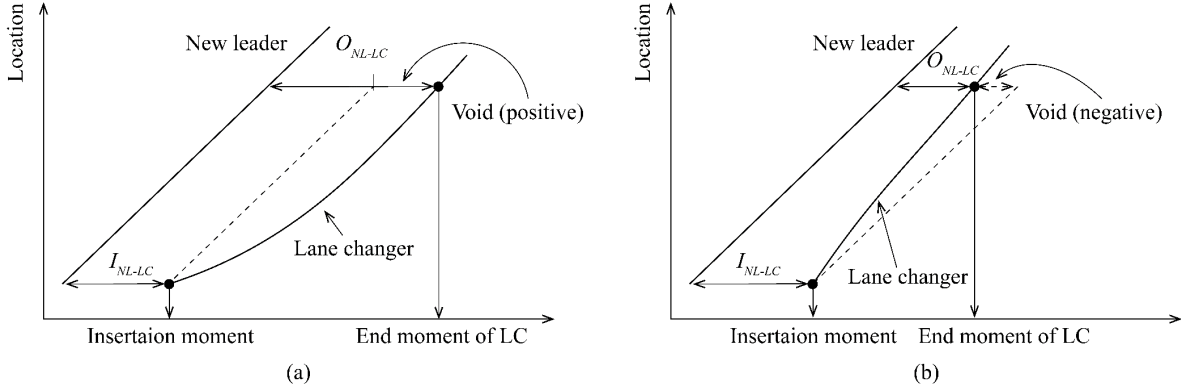


Figure 7 Two examples for the quantification of the changing process of the void created by the lane changer (a) example 1; (b) example 2

The second metric, denoted as ΔO_{NF} , quantifies the void caused by the new follower, which can be calculated as below:

$$\Delta O_{NF} = O_{LC-NF} - I_{NL-NF} \quad (9)$$

where O_{LC-NF} is the final time gap between the lane changer and the new follower at the end moment of LC, and I_{NL-NF} is the initial time gap between the new leader and the new follower at the start moment of LC.

Figure 8 provides two examples for Equation (9). It follows that a larger value of ΔO_{NF} indicates that the new follower is more likely to create an extra time gap compared to its initial driving behavior, which can negatively affect the traffic flow efficiency.

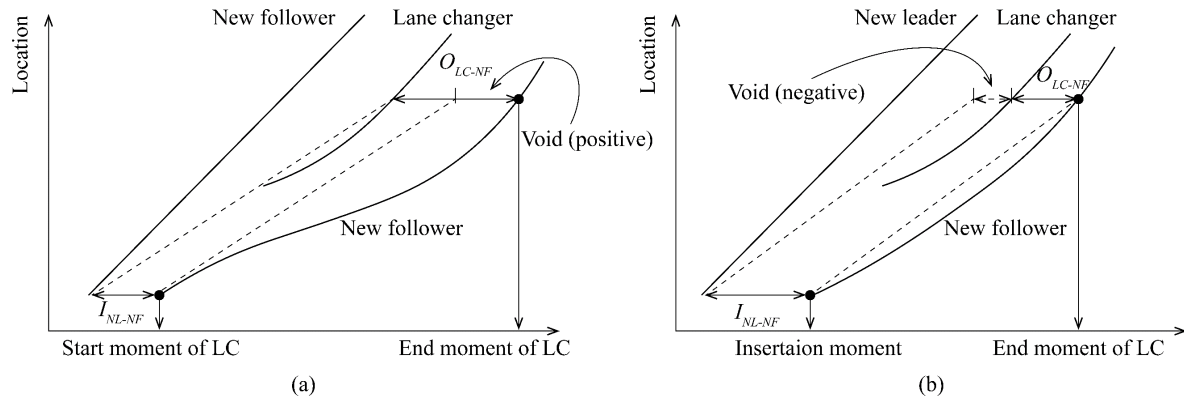


Figure 8 Two examples for the quantification of the changing process of the void created by the new follower (a) example 1; (b) example 2

4.3 Binary logit model

Considering the fact that the outcome of an LC decision in our case is either an NGFLC or a normal LC, we use a binary variable Y to capture this outcome. Y equals 1 if the LC decision leads to NGFLC and 0 otherwise. Let X represent the vector of independent variables, which corresponds to the current traffic conditions when the lane changer makes its LC decision. A logit function is used to generate a predicted value of Y under the impact of X . Thus, we have the following expressions for the binary logit model.

$$\ln \frac{p(1|X)}{1-p(1|X)} = \alpha + \beta X + \varepsilon \quad (10)$$

$$p(1|X) = \frac{\exp(\alpha + \beta X + \varepsilon)}{1 + \exp(\alpha + \beta X + \varepsilon)} \quad (11)$$

where α is the intercept, X is the vector of independent variables consisting of x_1, x_2, \dots, x_n (e.g., relative speed, time headway, etc.), β is the vector of estimated coefficients, ε is the error term, and $p(1|X)$ is the predicted probability that the LC decision under a given X will lead to NGFLC.

We evaluate the proposed models from their fitting performance and prediction performance. Specifically, the Akaike information criterion (AIC), as suggested by [Araike \(1998\)](#), is used to compare the fitting performance of different models, which can be computed as follows:

$$AIC = 2n - 2L(\theta) \quad (12)$$

where n is the number of independent variables, $L(\theta)$ is the log-likelihood at convergence. The model with a lower AIC is considered the better model.

To compare the predicted values of the models with their ground truth, we use a confusion matrix, which includes four basic elements: true positive (TP), false positive (FP), false negative (FN), and true negative (TN). Based on these four elements, three indicators, *Recall*, *False Alarm Rate (FAR)*, and *F-score*, are calculated to describe and compare the models' prediction performance.

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

$$FAR = \frac{TN}{FP + TN} \quad (14)$$

$$F-score = \frac{2TP}{2TP + FP + FN} \quad (15)$$

5. Results

In this section, we first present the consequences of NGFLC on both crash risk and traffic flow efficiency and also compare these outcomes with those of normal LC, which could facilitate an in-depth analysis of NGFLC (Section 5.1). Then, the contributing factors and how they characterize an LC decision as an NGFLC maneuver are discussed in Section 5.2. The analysis in this section is only based on dataset 1.

5.1 Crash risk analysis

This study categorizes the potential crash risks present in both original and target lanes during the LC maneuver into four distinct groups. These groups are:

- Group 1: the new follower and the lane changer;
- Group 2: the new follower and the new leader;
- Group 3: the lane changer and the new leader;
- Group 4: the lane changer and the original leader.

For each Group, TTC and DRAC are first applied to all LC samples to identify conflict LC samples, respectively. The results are given in **Table 2**. From these conflict LC samples identified through TTC and DRAC, the percentages of NGFLC and normal LC samples are shown in **Figure 9** and **Figure 10**, respectively. Detailed discussions of these findings are provided below.

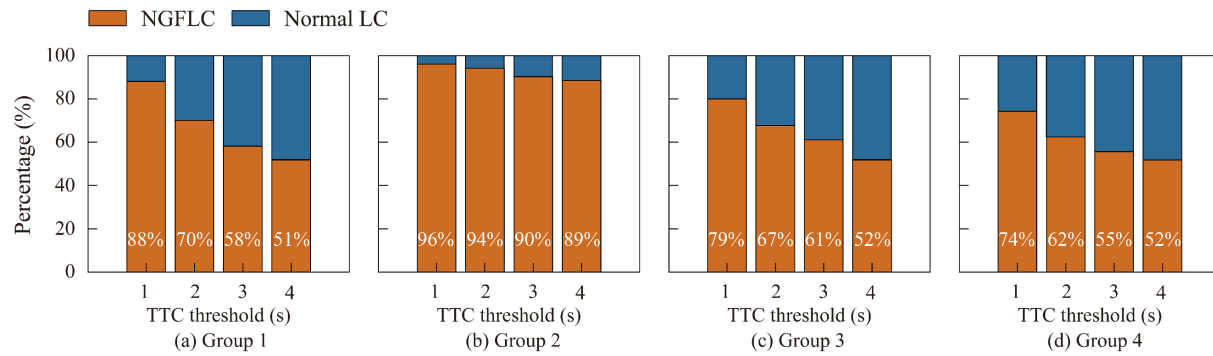


Figure 9 Percentages of NGFLC and normal LC in conflict LC samples identified by TTC: (a) Group 1; (b) Group 2; (c) Group 3; (d) Group 4

For Group 1, **Table 2** suggests that 10.53% of LC samples result in conflict situations when the TTC^* is set at 1s. Among these conflict LC samples, 88% are identified as NGFLCs, as shown in **Figure 9(a)**. Given that a TTC value below 1s is indeed dangerous since it may not provide enough time to avoid a crash, this result highlights that NGFLC is responsible for a predominant part of the highly risky Group 1 during LC maneuvers.

Table 2 Percentage of conflict LC samples identified by TTC and DRAC

SSMs	Threshold	Percentage of conflict LC samples to total LC samples (%)			
		Group 1	Group 2	Group 3	Group 4
TTC	1s	10.53	0.41	6.01	0.82
	2s	16.69	2.57	10.61	4.87
	3s	26.14	5.85	17.01	8.91
	4s	35.11	10.69	23.09	15.12
DRAC	$2m/s^2$	19.24	6.07	12.8	9.20
	$3m/s^2$	11.35	3.46	8.29	5.06
	$4m/s^2$	6.21	1.37	4.19	2.73
	$5m/s^2$	2.98	0.02	1.92	0.88

Regarding Group 2, **Table 2** reveals that only 10.69% of LC samples are dangerous when we set the TTC^* at 4s. This proportion further reduces to 0.41% when TTC^* is reduced to 1s. **Figure 9(b)** suggests that the proportion of normal LC samples to these conflict LC samples is

relatively small, ranging from 3.84% to 11.38%. This finding is reasonable because the new follower must maintain a large gap from its leader to accommodate the lane changer during normal LC maneuvers. However, when the lane changer conducts an NGFLC maneuver, the new follower suffers a higher crash risk with its leader, which can be inferred by the significant difference between the orange and blue bars in **Figure 9(b)**. This finding provides crucial microscopic insight into a novel rear-end crash formation mechanism, diverging from previous studies that primarily concentrated on two crash targets in the target lane: (1) potential crash risk between the new follower and the lane changer; (2) potential crash risk between the lane changer and the new leader (Gu et al., 2019; Yang and Ozbay, 2011). Our study indicates that a lane changer in the adjacent lane could trigger a rear-end collision in the target lane. A possible explanation for this result is that the new follower pays much attention to the lane changer since they are laterally adjacent. Consequently, the movement of the new leader has been largely ignored by the new follower, leading to an increase in the crash risk.

Figure 9(c) presents the crash risk of Group 3. As we can see from this figure, the orange bar is consistently higher than the blue bar for all TTC* values. This pattern indicates that NGFLCs are more likely to create risky situations between the lane changer and the new leader. Similarly, **Figure 9 (d)** shows comparable trends in Group 4, suggesting that during NGFLCs, the interactions between the LC and the original leader are more dangerous before insertion than in normal LCs.

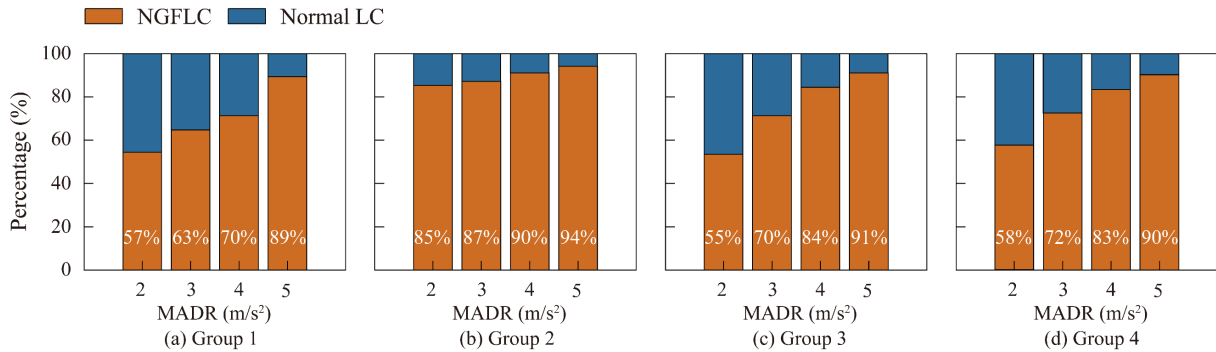


Figure 10 Proportion of NGFLC and normal LC in conflict LC samples identified by DRAC: (a) Group 1; (b) Group 2; (c) Group 3; (d) Group 4

Figure 10 reveals that the percentage of NGFLC samples in each Group notably exceeds that of normal LC samples. This observation is consistent with the LC crash risk assessment using TTC. Moreover, it implies that vehicles involved in NGFLC require a more significant deceleration effort to avoid crashes effectively.

To evaluate the severity and frequency of crash risk during a complete LC maneuver, we calculate the TIT, TET, and CPI for all conflict LC samples. The results are given in **Figure 11**. Generally, the average TET values for NGFLC under various thresholds are notably higher than those for normal LC across all Groups. This result indicates that NGFLCs tend to induce longer hazardous situations for surrounding vehicles compared to normal LCs. Similarly, in each Group, the average TIT and CPI values for NGFLC are both larger than those for normal LC, indicating a higher potential for crash severity during NGFLC maneuver.

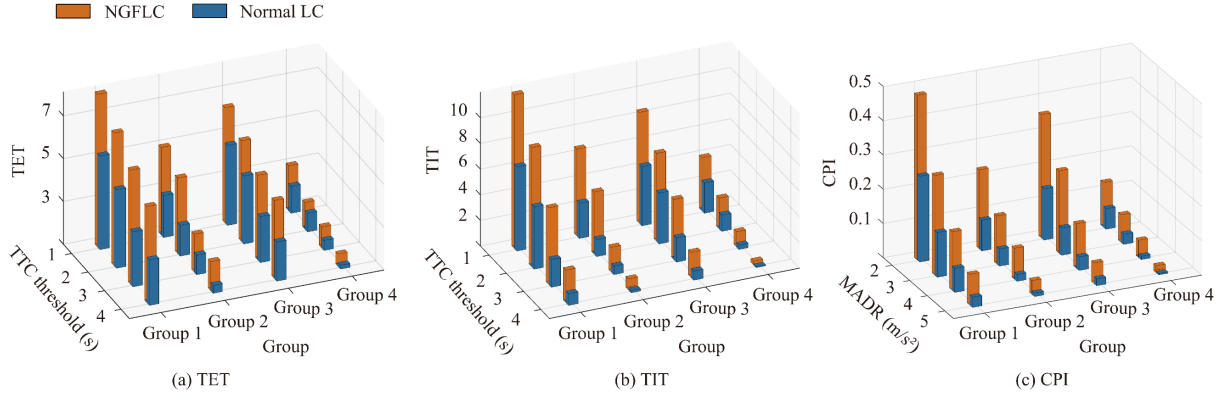


Figure 11 TET, TIT, and CPI for different Groups

5.2 Traffic flow efficiency analysis

This section evaluates the impact of NGFLC and normal LC on traffic flow efficiency using the measures outlined in Section 4.2. **Figure 12 (a)** shows the void created by the lane changer upon entering the target lane. The red and green boxplots in this figure represent NGFLC and normal LC, respectively. For NGFLCs, the average time gap for the lane changer is 1.4s at the insertion moment. Then, it rises to 2.6s at the end moment. In contrast, normal LCs significantly increase the average time gap from 1.4s to 1.8s. A paired t-test is applied to check whether the time gap increases significantly between NGFLC and normal LC samples. The test result confirms that the lane changer is significantly more likely to create voids after NGFLC (Average $\Delta O_{LC} = 1.2s$) than after normal LC (Average $\Delta O_{LC} = 0.2s$), thus exerting a more substantial effect on traffic flow efficiency.

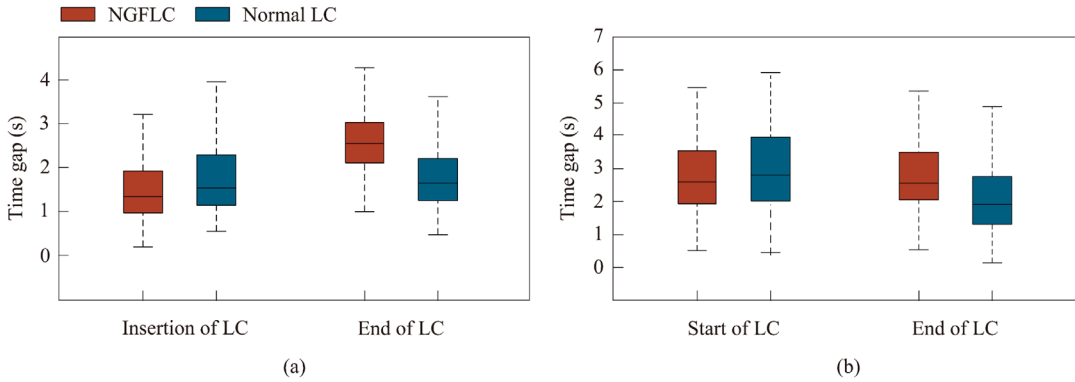


Figure 12 Changing pattern of time gap under NGFLC and normal LC on dataset 1 for (a) the lane changer and (b) the new follower

Figure 12 (b) illustrates changes in the new follower's time gap during LC maneuvers. It can be observed that the average time gap of the new follower at the start moment of NGFLCs (2.4s) is smaller than that of normal LCs (2.9s). Furthermore, the new follower tends to maintain a larger time gap after NGFLC (2.3s) compared to that after normal LC (1.9s). A paired t-test reveals that the average change value in the new follower's time gap (ΔO_{NF}) after NGFLC (-0.1s) is significantly lower than after normal LC (-1.0s). This finding suggests that the willingness of the new follower to catch up with the lane changer is more likely to be restricted after experiencing NGFLC.

Indeed, several previous studies have also stated that the lane changer, as well as its new follower, both have a larger probability of creating voids in the target lane during the LC maneuver, resulting in the reduction of traffic flow capacity (Chen and Ahn, 2018; Marczak and Buisson, 2014; Oh and Yeo, 2015). Our above findings are similar to these empirical findings. The findings suggest that even when the assumption of the bounded acceleration and the free speed do not meet in the real traffic flow, the voids can still be induced during the LC maneuver. Most importantly, the current results here provide new insights into the literature by demonstrating that NGFLC, compared to normal LC, maybe the primary reason for the void creation. It is recommended that future efforts should be devoted to investigating its relation to some critical traffic phenomena, such as oscillation and capacity drop.

5.3 Performance of binary logit models

In the literature, the focus for analyzing LC maneuvers has typically been on the vehicles' information in the target lane, while information pertaining to the original lane is often overlooked. For the purpose of comparison, we develop two binary logit models. **Table 3** presents the definitions and description statistics of the candidate independent variables used in these two models. The rejected gap in this table is calculated by counting the number of gaps presented to the lane changer before it initiates the LC maneuver. Specifically, the first model considers only information from the target lane (model 1), while the second enhanced model includes the leader's information in the original lane (model 2). The Pearson correlation test is adopted for each pair of candidate variables to avoid multicollinearity. The results show that all correlation values are less than 0.4, indicating non-significant collinearity is found among the candidate variables.

Table 3 Definition and description of selected independent variables

Variables	Description	Mean	S.D.	Min.	Max.
Variables for the target lane					
T_{NL-NF}	Time headway between the new leader and new follower on the target lane (s)	2.1	2.9	0.11	10.2
R_g	The number of rejected gaps	3.7	2.8	0	11
M_{gaps}	The average value of rejected gaps (m)	7.2	5.7	2.8	22.3
V_{NL-NF}	The speed difference between the new leader and the new follower on the target lane (m/s)	0.6	2.9	-11.3	11.6
V_{LC-NF}	The speed difference between the lane changer and the new follower on the target lane (m/s)	0.3	4.1	-9.3	10.4
Variables for the original lane					
T_{LO-LC}	The time headway between the leader in the original lane and the lane changer (s)	1.3	1.5	0.1	8.5
V_{LO-NL}	The speed difference between the leader in the original lane and the leader in the target lane (m/s)	-0.3	2.5	-8.9	11.3
A_{LO-LC}	The acceleration difference between the leader in the original and the lane changer (m/s^2)	-0.4	1.2	-4.3	2.1

Table 4 presents the results related to the developed binary logit models. The parameter time headway in the target lane (T_{NL-NF}) is negatively associated with the probability of NGFLC. It indicates that NGFLC is more likely to occur when the lane changer decides to change lanes

with a small time headway in the target lane. This result is consistent with the earlier findings in **Figure 10 (b)**, which demonstrate that the average T_{NL-NF} for NGFLCs is smaller than that for normal LCs. In most of the previous studies, a small gap in the target lane leads to two consequences: either the lane changer rejects this gap (Marczak et al., 2013), or the lane changer accepts it but aborts the LC maneuver when it deems the conditions in the target lane are unsuitable to complete the following LC maneuver (Ali et al., 2020). Our finding offers new insight into the outcome of accepting a small gap. Specifically, the lane changer will accept it and attempt to yield the new follower to a slow speed to provide a larger gap by laterally overlapping with the new follower.

Table 4 Estimation results mode 1 and model 2

Variables	Model 1			Model 2		
	Mean	SE.	P-value	Mean	SE.	P-value
Constant	3.4	0.2	<0.001	3.8	0.2	<0.001
T_{NL-NF}	-0.8	0.1	<0.001	-1.5	0.1	<0.001
Rg	0.6	0.3	<0.001	0.9	0.2	<0.001
M_{gaps}	-0.3	0.02	<0.001	-0.5	0.04	0.002
V_{NL-NF}	1.7	0.4	<0.001	2.1	0.3	<0.001
V_{LC-NF}	0.9	0.1	0.002	0.6	0.1	<0.001
T_{LO-LC}	-	-	-	-0.7	0.1	<0.001
V_{LC-LO}	-	-	-	1.6	0.2	<0.001
A_{LC-LO}	-	-	-	3.2	0.5	<0.001
V_{NL-LC}	-	-	-	-1.5	0.1	<0.001
Model fit statistics						
Number of observations	12815			12815		
AIC	4517			3142		

The positive coefficient of the number of rejected gaps (Rg) indicates that the probability of NGFLC increases with an increase in Rg . Moreover, the negative coefficient of the average value of rejected gaps (M_{gaps}) suggests that the lower M_{gaps} lead to a higher probability of NGFLC. These findings imply that when a lane changer, after repeatedly rejecting small gaps in the target lane, tends to become more impatient and is more likely to conduct an NGFLC. Previously, some studies have focused on why lane changers reject or accept a presented gap in the target lane. The results of this study emphasize that the lane changer's rejection behavior may affect its following LC maneuver. Hence, the gap acceptance theory should further incorporate the lane changer's historical rejection behavior to improve the accuracy of predictions.

Now, we examine the relative speed of the new follower to the new leader (V_{NL-NF}) and to the lane changer (V_{LC-NF}). Large values of these variables indicate that the new follower is quickly approaching both the new leader and the lane changer. In this context, Yang et al. (2019) found that a large gap in the target lane is required for the lane changer to decide on an LC decision. In our study, we are interested in how the two variables affect the LC maneuver if the lane changer insists on changing lanes. A positive relationship between the variables of V_{FV-LV} and V_{FV-LC} and the probability of NGFLC is observed. It indicates that an LC decision is more likely to result in an NGFLC when the values of V_{FV-LV} or V_{FV-LC} are higher. This finding is

consistent with our earlier results in Section 4.2.1, which indicated that NGFLC is more hazardous than normal LC since a larger value of speed difference poses a significant crash risk.

Regarding the relation between the lane changer and the original leader, we find that the coefficient of their time headway (T_{LO-LC}) is negative. This implies that a smaller time headway to the original leader increases the likelihood of NGFLC. Furthermore, the positive coefficients of their relative speed (V_{LC-LO}) and relative acceleration (A_{LC-LO}) indicate that the faster the lane changer, the larger probability the LC decision will result in NGFLC. It is reasonable that when the lane changer is quickly approaching the original leader, it is more likely to execute an urgent LC maneuver to avoid collision with the original leader. This LC intention increases the frequency of NGFLC.

The relative speed between the new leader and the original leader (V_{NL-LO}) is positively associated with the probability of NGFLC. It indicates that if the new leader is traveling at a higher speed than the original leader, the probability of NGFLC increases. This phenomenon occurs because the slower original leader restricts the lane changer's longitudinal movement. As a result, the lane changer can only conduct a sharp lateral movement and create an overlap with the new follower. Similarly, the original leader's movement has been found to be related to the risk level of the LC maneuver (Chen et al., 2022). Our results here reveal that the original leader's behavior also significantly affects the lane changer's LC behavior and should be considered in the future LC analysis.

Table 4 presents the fitting results of the proposed models. By comparing the AIC values, we find that the model with considering the original leader has the lower AIC value and is more suitable to analyze the NGFLC maneuver. **Table 5** and **Table 6** show the prediction performance, revealing significant improvements in the model that include the original leader's information. Specifically, this enhanced model achieves a 14% higher accuracy and an 8% reduction in false detections compared to the model without considering the original leader. These comparison results highlight the importance of the original leader's role in analyzing the LC maneuvers.

Table 5 Confusion matrix of two developed models

Confusion matrix	Model 1		Model 2	
	Actual NGFLC	Actual normal LC	Actual NGFLC	Actual normal LC
Predicted NGFLC	TP = 3418	FP = 989	TP = 3851	FP = 781
Predicted normal LC	FN = 811	TN = 7617	FN = 378	TN = 7805

Table 6 Prediction quality of two developed models

Measure	Derivation	Model 1	Model 2
Recall	TP/(TP+FN)	81%	91%
False alarm rate	FP/(FP+TN)	11%	9%
F1 score	2TP/(2TP+FP+FN)	79%	87%

6. Transferability test

In **Table 1**, we observe a similar percentage of NGFLC samples in both dataset 2 and dataset 3. This similarity suggests that NGFLCs are a common occurrence across various locations and cities. Consequently, this section is dedicated to examining the extent to which the impacts of NGFLCs on traffic flow can be found in these two datasets. Also, we aim to assess the transferability of the prediction model across different datasets.

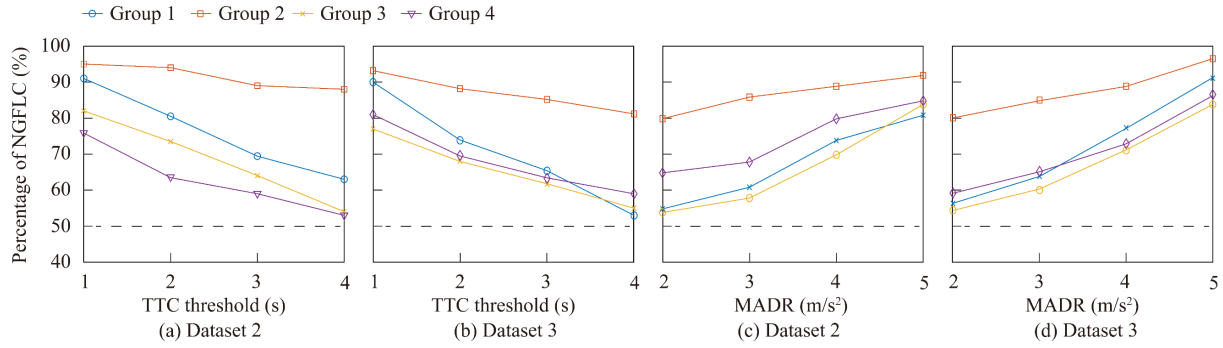


Figure 13 Percentage of NGFLC in conflict LC samples at dataset 2 and dataset 3

The analytical methods applied in Section 5.1 and Section 5.2 are replicated for dataset 2 and dataset 3. **Figure 13** presents the percentage of NGFLC in conflict LC samples for these two datasets. It is evident that NGFLC samples constitute over 50% of the total conflict LC samples. This finding is not affected by different datasets, Groups, and SSMs (TTC or DRAC). **Table 7** presents the results of TET, TIT, and CPI for these conflict LC samples. It demonstrated that the average duration and severity of crash risk during NGFLCs are both higher compared to normal LCs in dataset 2 and dataset 3.

Table 7 Results of TET, TIT, and CPI for NGFLCs and normal LCs at dataset 2 and dataset 3

SSM	Threshold	Dataset 2				Dataset 3			
		NGFLC (normal LC)				NGFLC (normal LC)			
		Group1	Group2	Group3	Group4	Group1	Group2	Group3	Group1
TET (s)	1s	4.1(1.7)	1.2(0)	2.8(0)	0.8(0)	3.6(1.2)	1.4(0)	3.2(0.5)	0.5(0)
	2s	5.3(3.6)	2.4(0)	3.5(1.2)	1.2(0)	4.1(2.3)	2.6(0.5)	3.7(1.3)	0.9(0)
	3s	6.1(3.8)	3.3(0.9)	4.7(2.5)	1.7(0.5)	5.7(3.1)	3.2(1.2)	4.2(1.9)	1.1(0.2)
	4s	6.9(4.2)	4.2(1.7)	5.6(3.4)	2.3(1.3)	6.3(3.9)	4.1(1.9)	5.4(2.6)	1.9(0.6)
TIT (s)	1s	1.9(0.6)	0.7(0)	1.4(0)	0.6(0)	1.3(0.4)	0.5(0)	2.1(0.2)	0.2(0)
	2s	3.6(1.8)	1.6(0)	2.6(0.9)	1.3(0)	3.3(1.9)	1.1(0.9)	3.9(1.1)	0.6(0)
	3s	6.2(3.4)	3.8(1.2)	4.1(2.7)	2.2(1.6)	6.5(3.6)	2.9(1.8)	5.3(2.6)	1.5(0.8)
	4	9.5(6.1)	6.4(2.8)	7.7(4.8)	3.8(2.5)	8.5(5.7)	5.6(3.4)	8.1(4.8)	2.8(1.7)
CPI (%)	2m/s2	37(26)	19(8)	31(24)	12(6)	34(22)	15(6)	27(18)	10(6)
	3m/s2	21(18)	12(6)	23(15)	6(4)	19(14)	11(3)	21(11)	8(2)
	4m/s2	15(9)	5(2)	14(6)	4(1)	11(5)	8(1)	15(7)	5(0.4)
	5m/s2	7(4)	2(1)	8(2)	1(0.5)	5(2)	4(0.4)	4(2)	2(0.2)

As expected, **Table 8** suggests that the adjustments made to the time headways by the new follower and the lane changer during NGFLC in dataset 2 and dataset 3 are similar to those discussed in Section 5.2. These consistent observations across all three trajectory datasets emphasize the significant impact of NGFLCs on the safety of surrounding vehicles and traffic flow efficiency, highlighting their importance and the need for careful consideration in future research.

Table 8 Impact of NGFLCs and normal LCs on traffic flow efficiency at dataset 2 and dataset 3

Efficiency measurement	Vehicles	Moments	Dataset 2	Dataset 3
			NGFLC (normal LC)	NGFLC (normal LC)
Time gap (s)	New	Start moment of LC	1.3(1.5)	1.8(1.7)
	follower	End moment of LC	2.2(1.9)	3.1(1.9)
	Lane	Insertion moment of LC	2.2(3.4)	3.6(3.9)
	changer	End moment of LC	2.0(2.1)	3.2(2.6)

To assess the transferability of the binary logit model developed with considering the original leader (model 2), we directly applied this model to the LC samples in dataset 2 and dataset 3. The prediction performance of model 2 is detailed in **Table 9**. It can be seen from this table that the recall, false alarm rate, and F1 score of model 2 in dataset 2 and dataset 3 are comparable to those observed in dataset 1. This similarity results suggest that model 2 is effectively transferable across different locations.

Table 9 Transferability results of model 2 at dataset 2 and dataset 3

Measure	Model with original lane information	
	Dataset 2	Dataset 3
Recall	87%	89%
False alarm rate	10%	11%
F1 score	84%	86%

7. Conclusion

This study investigates a critical LC maneuver known as negative forced lane-changing (NGFLC). Three vehicle trajectory datasets from different locations and cities are used, namely datasets 1 to 3. Specifically, dataset 1 is used for modeling, and datasets 2 and 3 are used for transferability testing. A classification procedure is proposed to distinguish between NGFLC and normal LC samples. The results show that NGFLC samples constitute a significant percentage of the LC samples, accounting for 33%, 36%, and 28% in each dataset. Based on these data, the empirical analysis and predictive models offer four major contributions to the understanding of NGFLC:

(a) Crash risk analysis: Five widely used SSMs are extended to a two-dimensional scope for NGFLC and normal LC samples. The comparison results yield three key insights. First, NGFLCs account for approximately 90% of extremely hazardous LC maneuvers, characterized by a TTC value less than 1s or a DRAC value larger than 5m/s^2 . These are identified as the primary contributors to LC crashes. Second, a notable potential crash risk is observed between the new follower and the new leader during NGFLCs, while this risk is comparatively lower during normal LCs. Lastly, NGFLCs tend to prolong the duration of danger and increase the severity of crash risk during LC maneuvers.

(b) Traffic flow efficiency: This study quantifies the impact of LC maneuvers on traffic flow efficiency by examining the changes in time gaps. Findings highlight that the lane changers are more likely to increase the initial time gap after performing NGFLC maneuvers. Although we observe that the new follower attempts to reduce the time gap, which holds for NGFLC and normal LC, the extent of reduction in normal LCs is significantly greater than in NGFLCs. This

implies that both the lane changer and the new follower involved in NGFLCs are more likely to disrupt traffic flow.

(c) Contribution factors: With an awareness of the profound impact of NGFLC on traffic flow, two binary logit models are developed to identify the critical factors affecting the probability of an LC decision resulting in an NGFLC. These factors, such as time headway, the number of rejected gaps, the average value of rejected gaps, the relative speed of the new follower to the lane changer, and the relative speed of the new follower to the new leader significantly impact the occurrence of NGFLCs. Moreover, incorporating the information about the original leader improves the prediction accuracy. Notably, this information has been largely ignored in previous studies.

(d) Transferability: Our insights and prediction models developed for NGFLC using dataset 1 are validated with dataset 2 and dataset 3. The results confirm that the negative outcomes of NGFLC can be observed across different locations and cities. Also, the prediction model achieves satisfactory accuracy when directly applied to these two datasets.

Despite this study providing a comprehensive view of the consequences and causes of NGFLC, some limitations could merit further research. For example, our analysis focused on only one type of NGFLC. Indeed, a variety of NGFLC could be observed in the traffic flow, such as the lane changer overlapping with the new follower while the new follower successfully closed the gap. In such a situation, the lane changer has to wait for the next gap. Thus, this NGFLC's impact may differ from the FLC we introduced in this study. Further efforts are needed to examine different types of FLC to present a full understanding of the overlapping behavior during the LC maneuver. Also, we only focused on the safety and efficiency issues on the target lane. A more comprehensive evaluation could be conducted to explain the consequences of NGFLC, such as the cooperation LC induced by NGFLC and its impact on the original lane. The lateral overlap between two vehicles seems closely linked to sideswipe crashes. This study only identified the NGFLC out of the normal LC. Further research is needed to investigate which NGFLC is more likely to lead to a sideswipe crash.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (51925801, 52232012, 52272331) and the Fundamental Research Funds for the Central Universities (2242019R40060, 2242020K40063). The authors thank the anonymous reviewers for their time to review our article and their constructive comments.

REFERENCES

- Adanu, E.K., Lidbe, A., Tedla, E., Jones, S., 2021. Factors associated with driver injury severity of lane changing crashes involving younger and older drivers. *Accident Analysis & Prevention* 149, 105867. <https://doi.org/10.1016/j.aap.2020.105867>
- Ali, Y., Bliemer, M.C., Haque, M.M., Zheng, Z., 2022. Examining braking behaviour during failed lane-changing attempts in a simulated connected environment with driving aids. *Transportation Research Part C: Emerging Technologies* 136, 103531.
- Ali, Y., Sharma, A., Haque, Md.M., Zheng, Z., Saifuzzaman, M., 2020b. The impact of the connected environment on driving behavior and safety: A driving simulator study. *Accident Analysis & Prevention* 144, 105643. <https://doi.org/10.1016/j.aap.2020.105643>

- 1 Ali, Y., Zheng, Z., Haque, Md.M., 2018. Connectivity's impact on mandatory lane-changing
2 behaviour: Evidences from a driving simulator study. *Transportation Research Part C:*
3 *Emerging Technologies* 93, 292–309. <https://doi.org/10.1016/j.trc.2018.06.008>
- 4 Ali, Y., Zheng, Z., Haque, Md.M., Wang, M., 2019. A game theory-based approach for
5 modelling mandatory lane-changing behaviour in a connected environment.
6 *Transportation Research Part C: Emerging Technologies* 106, 220–242.
7 <https://doi.org/10.1016/j.trc.2019.07.011>
- 8 Ali, Y., Zheng, Z., Mazharul Haque, Md., Yildirimoglu, M., Washington, S., 2020a. Detecting,
9 analysing, and modelling failed lane-changing attempts in traditional and connected
10 environments. *Analytic Methods in Accident Research* 28, 100138.
11 <https://doi.org/10.1016/j.amar.2020.100138>
- 12 Ben-Akiva, M.E., Choudhury, C.F., Lee, G., Rao, A., Toledo, T., 2006. Verification and
13 validation plan: forced lane change and cooperative merging model. NGSIM Group
14 Report, MIT 1–5.
- 15 Biswas, R.K., Friswell, R., Olivier, J., Williamson, A., Senserrick, T., 2021. A systematic review
16 of definitions of motor vehicle headways in driver behaviour and performance studies.
17 *Transportation Research Part F: Traffic Psychology and Behaviour* 77, 38–54.
18 <https://doi.org/10.1016/j.trf.2020.12.011>
- 19 Chang, G.-L., Kao, Y.-M., 1991. An empirical investigation of macroscopic lane-changing
20 characteristics on uncongested multilane freeways. *Transportation Research Part A:*
21 *General* 25, 375–389. [https://doi.org/10.1016/0191-2607\(91\)90015-I](https://doi.org/10.1016/0191-2607(91)90015-I)
- 22 Chauhan, P., Kanagaraj, V., Asaithambi, G., 2022. Understanding the mechanism of lane
23 changing process and dynamics using microscopic traffic data. *Physica A: Statistical*
24 *Mechanics and its Applications* 593, 126981. <https://doi.org/10.1016/j.physa.2022.126981>
- 25 Chen, D., Ahn, S., 2018. Capacity-drop at extended bottlenecks: Merge, diverge, and weave.
26 *Transportation Research Part B: Methodological* 108, 1–20.
27 <https://doi.org/10.1016/j.trb.2017.12.006>
- 28 Chen, D., Ahn, S., Laval, J., Zheng, Z., 2014. On the periodicity of traffic oscillations and
29 capacity drop: the role of driver characteristics. *Transportation research part B:*
30 *methodological* 59, 117–136.
- 31 Chen, D., Srivastava, A., Ahn, S., Li, T., 2020. Traffic dynamics under speed disturbance in
32 mixed traffic with automated and non-automated vehicles. *Transportation Research Part*
33 *C: Emerging Technologies* 113, 293–313. <https://doi.org/10.1016/j.trc.2019.03.017>
- 34 Chen, K., Knoop, V.L., Liu, P., Li, Z., Wang, Y., 2022. How gaps are created during anticipation
35 of lane changes. *Transportmetrica B: Transport Dynamics* 1–21.
36 <https://doi.org/10.1080/21680566.2022.2152129>
- 37 Chen, K., Liu, P., Li, Z., Wang, Y., Lu, Y., 2021. Modeling Anticipation and Relaxation of Lane
38 Changing Behavior Using Deep Learning. *Transportation Research Record* 2675, 186–
39 200. <https://doi.org/10.1177/03611981211028624>
- 40 Chen, K., Knoop, V.L., Liu, P., Li, Z., Wang, Y., 2023. Modeling the impact of lane-changing's
41 anticipation on car-following behavior. *Transportation Research Part C: Emerging*
42 *Technologies* 150, 104110. <https://doi.org/10.1016/j.trc.2023.104110>
- 43 Chen, K., Li, Z., Yang, Y., Qi, L., Ke, K. High-Resolution Vehicle Trajectory Extraction and
44 Denoising From Aerial Videos. *IEEE Transactions on Intelligent Transportation Systems*,
45 vol. 22, no. 5, pp. 3190-3202, May 2021, doi: 10.1109/TITS.2020.3003782.

- Chen, Q., Huang, H., Li, Y., Lee, J., Long, K., Gu, R., Zhai, X., 2021. Modeling accident risks in different lane-changing behavioral patterns. *Analytic Methods in Accident Research* 30, 100159. <https://doi.org/10.1016/j.amar.2021.100159>
- Chen, X., Li, Z., Yang, Y., Qi, L., Ke, R., 2021. High-Resolution Vehicle Trajectory Extraction and Denoising From Aerial Videos. *IEEE Trans. Intell. Transport. Syst.* 22, 3190–3202. <https://doi.org/10.1109/TITS.2020.3003782>
- Chen, T., Shi, X., Wong, Y.D., 2019. Key feature selection and risk prediction for lane-changing behaviors based on vehicles' trajectory data. *Accident Analysis & Prevention* 129, 156–169. <https://doi.org/10.1016/j.aap.2019.05.017>
- Cunto, F., Saccomanno, F.F., 2008. Calibration and validation of simulated vehicle safety performance at signalized intersections. *Accident Analysis & Prevention* 40, 1171–1179. <https://doi.org/10.1016/j.aap.2008.01.003>
- Ding, N., Jiao, N., Zhu, S., Liu, B., 2019. Structural equations modeling of real-time crash risk variation in car-following incorporating visual perceptual, vehicular, and roadway factors. *Accident Analysis & Prevention* 133, 105298. <https://doi.org/10.1016/j.aap.2019.105298>
- Duret, A., Ahn, S., Buisson, C., 2011. Passing rates to measure relaxation and impact of lane-changing in congestion. *Computer-Aided Civil and Infrastructure Engineering* 26, 285–297.
- Gu, X., Abdel-Aty, M., Xiang, Q., Cai, Q., Yuan, J., 2019. Utilizing UAV video data for in-depth analysis of drivers' crash risk at interchange merging areas. *Accident Analysis & Prevention* 123, 159–169. <https://doi.org/10.1016/j.aap.2018.11.010>
- Hess, S., Choudhury, C.F., Bliemer, M.C.J., Hibberd, D., 2020. Modelling lane changing behaviour in approaches to roadworks: Contrasting and combining driving simulator data with stated choice data. *Transportation Research Part C: Emerging Technologies* 112, 282–294. <https://doi.org/10.1016/j.trc.2019.12.003>
- Hidas, P., 2002. Modelling lane changing and merging in microscopic traffic simulation. *Transportation Research Part C: Emerging Technologies* 10, 351–371.
- Hou, Y., Edara, P., Sun, C., 2015. Situation assessment and decision making for lane change assistance using ensemble learning methods. *Expert Systems with Applications* 42, 3875–3882. <https://doi.org/10.1016/j.eswa.2015.01.029>
- Julian, E., Damerow, F., 2015. Complex Lane Change Behavior in the Foresighted Driver Model, in: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. Presented at the 2015 IEEE 18th International Conference on Intelligent Transportation Systems - (ITSC 2015), IEEE, Gran Canaria, Spain, pp. 1747–1754. <https://doi.org/10.1109/ITSC.2015.284>
- Laval, J.A., Daganzo, C.F., 2006. Lane-changing in traffic streams. *Transportation Research Part B: Methodological* 40, 251–264.
- Leclercq, L., Knoop, V.L., Marczyk, F., Hoogendoorn, S.P., 2016. Capacity drops at merges: New analytical investigations. *Transportation Research Part C: Emerging Technologies* 62, 171–181.
- Lee, C., Abdel-Aty, M., Hsia, L., 2006. Potential Real-Time Indicators of Sideswipe Crashes on Freeways. *Transportation Research Record* 1953, 41–49. <https://doi.org/10.1177/0361198106195300105>
- Li, Y., Gu, R., Lee, J., Yang, M., Chen, Q., Zhang, Y., 2021. The dynamic tradeoff between safety and efficiency in discretionary lane-changing behavior: A random parameters logit

- approach with heterogeneity in means and variances. *Accident Analysis & Prevention* 153, 106036. <https://doi.org/10.1016/j.aap.2021.106036>
- Li, Z., Chen, X., Ling, L., Wu, H., Zhou, W., Qi, C., 2019. Accurate Traffic Parameter Extraction from Aerial Videos with Multi-Dimensional Camera Movements. In: *The Transportation Research Board (TRB) 98th Annual Meeting*, Washington, D.C., No. 19-02817.
- Liu, Y.-B., Zeng, M., Meng, Q.-H., 2020. Heatmap-based Vanishing Point boosts Lane Detection. <https://doi.org/10.48550/ARXIV.2007.15602>
- Marczak, F., Buisson, C., 2014. Analytical Derivation of Capacity at Diverging Junctions. *Transportation Research Record* 2422, 88–95. <https://doi.org/10.3141/2422-10>
- Marczak, F., Daamen, W., Buisson, C., 2013. Merging behaviour: Empirical comparison between two sites and new theory development. *Transportation Research Part C: Emerging Technologies* 36, 530–546. <https://doi.org/10.1016/j.trc.2013.07.007>
- Marczak, F., Leclercq, L., Buisson, C., 2015. A Macroscopic Model for Freeway Weaving Sections: A macroscopic model for freeway weaving sections. *Computer-Aided Civil and Infrastructure Engineering* 30, 464–477. <https://doi.org/10.1111/mice.12119>
- Minderhoud, M.M., Bovy, P.H.L., 2001. Extended time-to-collision measures for road traffic safety assessment. *Accident Analysis & Prevention* 33, 89–97. [https://doi.org/10.1016/S0001-4575\(00\)00019-1](https://doi.org/10.1016/S0001-4575(00)00019-1)
- Moridpour, S., Sarvi, M., Rose, G., 2010a. Lane changing models: a critical review. *Transportation Letters* 2, 157–173. <https://doi.org/10.3328/TL.2010.02.03.157-173>
- Moridpour, S., Sarvi, M., Rose, G., 2010b. Modeling the Lane-Changing Execution of Multiclass Vehicles under Heavy Traffic Conditions. *Transportation Research Record* 2161, 11–19. <https://doi.org/10.3141/2161-02>
- Montanino, M., Punzo, V., 2013. Making NGSIM Data Usable for Studies on Traffic Flow Theory: Multistep Method for Vehicle Trajectory Reconstruction. *Transportation Research Record* 2390, 99–111. <https://doi.org/10.3141/2390-11>
- Oh, S., Yeo, H., 2015. Impact of stop-and-go waves and lane changes on discharge rate in recovery flow. *Transportation Research Part B: Methodological* 77, 88–102. <https://doi.org/10.1016/j.trb.2015.03.017>
- Qu, X., Wang, S., 2015. Long-Distance-Commuter (LDC) Lane: A New Concept for Freeway Traffic Management: LDC lane. *Computer-Aided Civil and Infrastructure Engineering* 30, 815–823. <https://doi.org/10.1111/mice.12102>
- Ouyang, P., Liu, P., Guo, Y., Chen, K., 2023. Effects of configuration elements and traffic flow conditions on Lane-Changing rates at the weaving segments. *Transportation Research Part A: Policy and Practice* 171, 103652. <https://doi.org/10.1016/j.tra.2023.103652>
- Park, H., Oh, C., Moon, J., Kim, S., 2018. Development of a lane change risk index using vehicle trajectory data. *Accident Analysis & Prevention* 110, 1–8. <https://doi.org/10.1016/j.aap.2017.10.015>
- Pei, Y., Wang, Y., Zhang, Y., 2010. Microscopic Model of Automobile Lane-changing Virtual Desire Trajectory by Spline Curves. *PROMET* 22, 203–208. <https://doi.org/10.7307/ptt.v22i3.276>
- Rafati Fard, M., Shariat Mohaymany, A., Shahri, M., 2017. A new methodology for vehicle trajectory reconstruction based on wavelet analysis. *Transportation Research Part C: Emerging Technologies* 74, 150–167. <https://doi.org/10.1016/j.trc.2016.11.010>

- 1 Reinolsmann, N., Alhajyaseen, W., Brijs, T., Pirdavani, A., Hussain, Q., Brijs, K., 2021.
2 Investigating the impact of a novel active gap metering signalization strategy on driver
3 behavior at highway merging sections. *Transportation Research Part F: Traffic*
4 *Psychology and Behaviour* 78, 42–57. <https://doi.org/10.1016/j.trf.2021.01.017>
- 5 Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. ORB: An efficient alternative to SIFT or
6 SURF, in: 2011 International Conference on Computer Vision. Presented at the 2011 IEEE
7 International Conference on Computer Vision (ICCV), IEEE, Barcelona, Spain, pp. 2564–
8 2571. <https://doi.org/10.1109/ICCV.2011.6126544>
- 9 Sen, B., Smith, J.D., Najm, W.G., 2003. Analysis of Lane Change Crashes, Performed by John A.
10 Volpe National Transportation System Center, Cambridge, MA. Sponsored by National
11 Highway Traffic Safety Administration, Washington DC.
- 12 Shi, X., Wong, Y.D., Li, M.Z.-F., Palanisamy, C., Chai, C., 2019. A feature learning approach
13 based on XGBoost for driving assessment and risk prediction. *Accident Analysis &*
14 *Prevention* 129, 170–179. <https://doi.org/10.1016/j.aap.2019.05.005>
- 15 Sun, D. (Jian), Elefteriadou, L., 2014. A Driver Behavior-Based Lane-Changing Model for Urban
16 Arterial Streets. *Transportation Science* 48, 184–205.
17 <https://doi.org/10.1287/trsc.1120.0435>
- 18 Sun, D. (Jian), Elefteriadou, L., 2010. Research and Implementation of Lane-Changing Model
19 Based on Driver Behavior. *Transportation Research Record* 2161, 1–10.
20 <https://doi.org/10.3141/2161-01>
- 21 Sun, D.J., Kondyli, A., 2010. Modeling Vehicle Interactions during Lane-Changing Behavior on
22 Arterial Streets: Modeling vehicle interactions during lane-changing behavior on arterial
23 streets. *Computer-Aided Civil and Infrastructure Engineering* 25, 557–571.
24 <https://doi.org/10.1111/j.1467-8667.2010.00679.x>
- 25 Thiemann, C., Treiber, M., Kesting, A., 2008. Estimating Acceleration and Lane-Changing
26 Dynamics from Next Generation Simulation Trajectory Data. *Transportation Research*
27 *Record* 2088, 90–101. <https://doi.org/10.3141/2088-10>
- 28 Venthuruthiyil, S.P., Chunchu, M., 2022. Anticipated Collision Time (ACT): A two-dimensional
29 surrogate safety indicator for trajectory-based proactive safety assessment. *Transportation*
30 *Research Part C: Emerging Technologies* 139, 103655.
31 <https://doi.org/10.1016/j.trc.2022.103655>
- 32 Wan, Q., Peng, G., Li, Z., Inomata, F.H.T., 2020. Spatiotemporal trajectory characteristic
33 analysis for traffic state transition prediction near expressway merge bottleneck.
34 *Transportation Research Part C: Emerging Technologies* 117, 102682.
- 35 Wang, X., Yang, M., Hurwitz, D., 2019. Analysis of cut-in behavior based on naturalistic driving
36 data. *Accident Analysis & Prevention* 124, 127–137.
37 <https://doi.org/10.1016/j.aap.2019.01.006>
- 38 Ward, J.R., Agamennoni, G., Worrall, S., Bender, A., Nebot, E., 2015. Extending Time to
39 Collision for probabilistic reasoning in general traffic scenarios. *Transportation Research*
40 *Part C: Emerging Technologies* 51, 66–82. <https://doi.org/10.1016/j.trc.2014.11.002>
- 41 Xing, L., He, J., Abdel-Aty, M., Cai, Q., Li, Y., Zheng, O., 2019. Examining traffic conflicts of
42 up stream toll plaza area using vehicles' trajectory data. *Accident Analysis & Prevention*
43 125, 174–187. <https://doi.org/10.1016/j.aap.2019.01.034>
- 44 Yang, D., Zhu, L., Ran, B., Pu, Y., Hui, P., 2016. Modeling and Analysis of the Lane-Changing
45 Execution in Longitudinal Direction. *IEEE Trans. Intell. Transport. Syst.* 17, 2984–2992.
46 <https://doi.org/10.1109/TITS.2016.2542109>

- 1 Yang, H., Ozbay, K., 2011. Estimation of Traffic Conflict Risk for Merging Vehicles on
2 Highway Merge Section. *Transportation Research Record* 2236, 58–65.
3 <https://doi.org/10.3141/2236-07>
- 4 Yang, M., Wang, X., Quddus, M., 2019. Examining lane change gap acceptance, duration and
5 impact using naturalistic driving data. *Transportation Research Part C: Emerging*
6 *Technologies* 104, 317–331. <https://doi.org/10.1016/j.trc.2019.05.024>
- 7 Yuan, J., Abdel-Aty, M., Cai, Q., Lee, J., 2019. Investigating drivers' mandatory lane change
8 behavior on the weaving section of freeway with managed lanes: A driving simulator
9 study. *Transportation Research Part F: Traffic Psychology and Behaviour* 62, 11–32.
10 <https://doi.org/10.1016/j.trf.2018.12.007>
- 11 Yun, M., Zhao, Jing, Zhao, Jianzhen, Weng, X., Yang, X., 2017. Impact of in-vehicle navigation
12 information on lane-change behavior in urban expressway diverge segments. *Accident*
13 *Analysis & Prevention* 106, 53–66. <https://doi.org/10.1016/j.aap.2017.05.025>
- 14 Zhang, Y., Ioannou, P.A., 2017. Coordinated Variable Speed Limit, Ramp Metering and Lane
15 Change Control of Highway Traffic. *IFAC-PapersOnLine* 50, 5307–5312.
16 <https://doi.org/10.1016/j.ifacol.2017.08.982>
- 17 Zhao, L., Sun, J., Zhang, H.M., 2013. Observations and Analysis of Multistep-Approach Lane
18 Changes at Expressway Merge Bottlenecks in Shanghai, China. *Transportation Research*
19 *Record* 2395, 73–82. <https://doi.org/10.3141/2395-09>
- 20 Zheng, L., Sayed, T., 2020. A bivariate Bayesian hierarchical extreme value model for traffic
21 conflict-based crash estimation. *Analytic Methods in Accident Research* 25, 100111.
22 <https://doi.org/10.1016/j.amar.2020.100111>
- 23 Zheng, Z., Ahn, S., Chen, D., Laval, J., 2011a. Freeway Traffic Oscillations: Microscopic
24 Analysis of Formations and Propagations using Wavelet Transform. *Procedia - Social and*
25 *Behavioral Sciences* 17, 702–716. <https://doi.org/10.1016/j.sbspro.2011.04.540>
- 26 Zheng, Z., Ahn, S., Chen, D., Laval, J., 2011b. Applications of wavelet transform for analysis of
27 freeway traffic: Bottlenecks, transient traffic, and traffic oscillations. *Transportation*
28 *Research Part B: Methodological* 45, 372–384. <https://doi.org/10.1016/j.trb.2010.08.002>
- 29