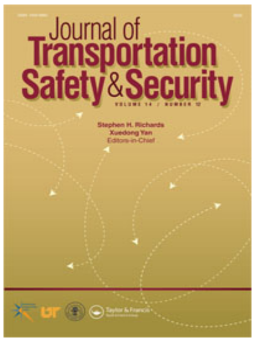

Journal of Transportation Safety & Security

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/utss20

Cluster analysis and multi-level modeling for evaluating the impact of rain on aggressive lane-changing characteristics utilizing naturalistic driving data

Anik Das, Md Nasim Khan, Mohamed M. Ahmed & Shaun S. Wulff

To cite this article: Anik Das, Md Nasim Khan, Mohamed M. Ahmed & Shaun S. Wulff (2022) Cluster analysis and multi-level modeling for evaluating the impact of rain on aggressive lane-changing characteristics utilizing naturalistic driving data, Journal of Transportation Safety & Security, 14:12, 2137-2165, DOI: 10.1080/19439962.2022.2069896

To link to this article: https://doi.org/10.1080/19439962.2022.2069896

Published online: 01 May 2022.

JOURNAL OF TRANSPORTATION SAFETY & SECURITY
2022, VOL. 14, NO. 12, 2137–2165
https://doi.org/10.1080/19439962.2022.2069896

Cluster analysis and multi-level modeling for evaluating the impact of rain on aggressive lane-changing characteristics utilizing naturalistic driving data

Anik Das[a], Md Nasim Khan[a], Mohamed M. Ahmed[b], and Shaun S. Wulff[c]

[a]Department of Civil and Architectural Engineering and Construction Management, University of Wyoming, Laramie, Wyoming, USA; [b]Department of Civil and Architectural Engineering and Construction Management, Williams and Person Professor, Wyoming Excellence Chair, University of Wyoming, Laramie, Wyoming, USA; [c]Department of Mathematics and Statistics, University of Wyoming, Laramie, Wyoming, USA

ABSTRACT

This study investigated lane-changing characteristics with regard to drivers' aggressiveness in rain and clear weather utilizing the SHRP2 Naturalistic Driving Study (NDS) dataset. An innovative methodology was developed to identify lane-changing events and extract corresponding parameters from the SHRP2 NDS database. Initially, K-means and K-medoids clustering methods were examined to classify drivers into non-aggressive and aggressive categories considering six features related to driving behavior, and K-means clustering was adopted based on the average silhouette width method (ASWM). Two-level mixed-effects linear regression models were calibrated to assess the contributing factors that affect lane-changing durations, which revealed that different vehicle kinematics, traffic, driver, and roadway characteristics, as well as weather conditions combined with other factors, were significant in the calibrated models for both driver types. The results revealed that the lane-changing duration associated with heavy rain decreased with a higher speed limit for aggressive drivers. Furthermore, the lane-changing duration associated with light/moderate rain decreased with the number of lanes for non-aggressive drivers. The study findings could be leveraged to incorporate drivers' aggressiveness into microsimulation lane-changing model calibration and validation as well as could have significant implications in improving safety in Connected and Autonomous Vehicles (CAV).

KEYWORDS

Lane-changing behavior; microsimulation modeling; SHRP2; naturalistic driving study; K-means clustering; K-medoids clustering; two-level mixed-effects linear regression models; lane-changing duration; Connected and Autonomous Vehicles

1. Introduction

Driving in inclement weather is critical and challenging from traffic safety perspective because of significant reduction of visibility, deteriorated roadway surroundings, as well as substantial impairment of driver behavior and

CONTACT Anik Das adas4@uwyo.edu Department of Civil and Architectural Engineering and Construction Management, University of Wyoming, 1000 E University Ave, Dept. 3295, Laramie, WY 82071, USA.

vehicle performance (Ariannezhad & Wu, 2019; S. Das, Dutta, & Sun, 2019). The statistics of the Federal Highway Administration (FHWA) demonstrate that about 21% of vehicle crashes occurred from 2007 to 2016 were adverse weather-related in which the majority of these crashes were due to wet pavement (70%) and presence of rainfall (46%) (FHWA, 2022). The intensity of rain, especially heavy rain, induces lane obstruction and limits driver behavior and performance through increasing visual obstruction and losing surface frictions (Jackson & Sharif, 2016). Lane-changing represents one of the most crucial and essential driver behavior affecting traffic safety (Shangguan, Wang, Fu, & Fang, 2021). Drivers' inaccurate and risky lane-changing maneuvers increase the probability of crash risk and might consequence in a lane-changing crash (Bakhit, Osman, & Ishak, 2017). Currently, a significant portion of motor vehicle crashes on highways are directly related to lane changing. In 2018, based on the National Highway Traffic Safety Administration (NHTSA), lane changing was accountable for nearly 543,000 motor vehicle crashes in the US (NHTSA, 2019). According to a report by Fitch et al., lane-changing events are responsible for a minimum of 60,000 people's injuries and a momentous quantity of property damages annually (Fitch et al., 2009). Outside of the US, additional studies revealed that lane-changing maneuvers cause roughly 12.6% of all crashes in the Netherlands and 9.8% of crash fatalities in Canada (You et al., 2015).

2. Related works

The related works section provides a brief outline of lane-changing behavior analysis based on various approaches and datasets, studies related to lane-changing durations, and studies focused on drivers' aggressiveness and its impact on lane-changing. Lane changing is one of the most common maneuvers on the roadways that occurs when a driver involves in the process of lateral movement of his/her vehicle from one lane to another and it depends on the decision made by the driver based on the neighboring environment (Zhang et al., 2022). Chovan et al. described lane-changing maneuvers as a cautious and considerable shift of a vehicle's lateral position (Chovan, Tijerina, Alexander, & Hendricks, 1994). Depending on drivers' motivation, the maneuvers are usually categorized as either mandatory or discretionary. Mandatory lane changing refers to the lane changing when the maneuver is required (e.g., to enter or exit from a highway, to avoid roadwork/obstacle, etc.). In contrast, discretionary lane changing represents the maneuver that is performed to achieve satisfactory driving conditions (e.g., to move from behind a slow leading vehicle).

Previous studies have analyzed lane-changing behavior by considering different approaches utilizing data from instrumented vehicles, connected vehicles, aerial surveys, simulations, trajectories, and naturalistic driving scenarios. Wang et al. utilized lane-changing frequency, turn signal, and rearview mirror to understand lane-changing characteristics using an experimental vehicle and concluded that longitudinal acceleration during the lane-changing period mainly ranged from $-1\,\mathrm{m/s^2}$ to $1\,\mathrm{m/s^2}$, and rear-view mirror usage is around 70% of the whole lane-changing times (J. Wang et al., 2013). Ali et al. investigated mandatory lane-changing maneuvers through a driving simulator experiment in a connected vehicle environment and concluded that drivers tend to increase initial speed, wait longer, and maintain a larger spacing to the lead vehicle during the process of lane-changing maneuvers (Ali, Zheng, & Haque, 2018). In another study, Qi et al. analyzed lane-changing behaviors on high occupancy vehicle (HOV) facilities with various access controls through aerial survey data and found that the spatial distributions of lane-changing intensity in the continuous access facility are much extended compared to the limited access HOV facility (Qi, Wu, Boriboonsomsin, & Barth, 2016). Zheng et al. examined the effects of lane-changing maneuvers in driver behavior using trajectory datasets from the Next Generation Simulation (NGSIM) and suggested that lane-changing decisions are related to driver characteristics and driver attitudes, including aggressiveness (Zheng, Ahn, Chen, & Laval, 2013). Chen et al. utilized vehicle kinematic data from a naturalistic dataset to predict lane-changing maneuvers via developing an automatic method. They summarized that the steering initiation for lane-changing events started within 5 s of the vehicle's lane line crossing (R. Chen, Gabler, & Sherony, 2017).

Lane-changing duration is one of the vital parameters of lane-changing maneuvers, although few studies utilized this parameter to investigate lane-changing behavior. A study observed 8,667 naturalistic lane-changing maneuvers from 16 participants and found a mean duration of 9.07 s (Lee, Olsen, & Wierwille, 2004). Hill et al. collected 726 freeway lane-changing events from 46 participants driving in an instrumented vehicle and found that lane-changing durations ranged from 2.30 s to 13.8 s with a mean of 5.48 s (Hill, Elefteriadou, & Kondyli, 2015). Using a driving simulator, Salvucci et al. identified an average lane-changing durations of 5.14 s from 11 participants (Salvucci & Liu, 2002). Toledo et al. used trajectory dataset and observed that the lane-changing durations ranged from 1 s to 13 s with a mean of 4.6 s from the 1,790 lane-changing events (Toledo & Zohar, 2007). Similarly, the study conducted by Moridpour et al. observed 42 lane-changing events of passenger cars using trajectory-level dataset and found the range of lane-changing durations from 2.1 s to 8.9 s, in addition to a mean of 4.8 s (Moridpour, Sarvi, & Rose, 2010). In another study, Wang et al. found an average lane-changing duration of car was 6.3 s using naturalistic lane change data (C. Wang, Li, Fu, Zhang, & Sun, 2019). Overall, the range of lane-changing durations varies from 1 s to 14 s, as suggested by these studies. One of the reasons for these variations of lane-changing durations could be the use of different datasets in different studies. As an example, driving simulator data are different from naturalistic driving study datasets since simulator studies are conducted in a controlled experimental environment. Thus, driver behavior and performance are highly likely to be biased and their behavior might not be the same as in real-life driving scenarios. In addition, studies have different definitions of lane-changing events and have different number and distribution of participants. Moreover, the variations could also be associated with driver heterogeneity (both inter as well as intra). These could be the probable reasons for substantial variations in lane-changing durations in several studies.

However, most of the previous studies considered driver as a homogeneous component. It is worth mentioning that drivers could exhibit heterogeneous driving behaviors based on their different levels of aggressiveness. Driver behavioral differences are extremely challenging to measure due to influence by various factors, such as vehicle kinematics, roadway geometry, and characteristics of surrounding drivers and traffic in addition to adverse weather conditions. All these factors contribute to significant negative impacts on lane-changing maneuvers. However, studies related to aggressive lane changing behavior suggested that aggressive drivers usually have a tendency of higher lane changing compared to less aggressive drivers (Golbabaei, Nejad, & Noory, 2014; Li, Jia, Gao, & Jiang, 2006). Despite recognizing the critical aspect of these behavioral situations, a few studies considered drivers' aggressiveness and their impact on lane-changing behavior. For instance, the study of Yang and Koutsopoulos adopted drivers' aggressiveness by incorporating an error term in the lane-changing logic at the MIcroscopic Traffic SIMulator (MITSIM) (Yang & Koutsopoulos, 1996). Sun and Elefteriadou conducted several studies of lane-changing behavior on urban streets based on focus groups to obtain driver characteristics (i.e., different aggressiveness levels) in an attempt to incorporate those into lane-changing models (Sun & Elefteriadou, 2010, 2011). Another study investigated relationships between various driver groups and their lane-changing characteristics considering variability on lane-changing decisions (Hill et al., 2015). Khattak et al. suggested that safety-critical events (SCE) were likely to be increased due to the increase in lane-changing and aggressive driving (Khattak, Fontaine, Li, Khattak, & Karnowski, 2021). The review of the previous studies undoubtedly suggests a need for a study to examine the role of drivers' aggressiveness in lane-changing behavior.

3. Research objective and contributions

As the review of the literature indicates, while various studies concentrated on lane-changing behavior from different viewpoints, there are a limited number of studies that analyzed drivers' aggressiveness during lane-changing maneuvers. More specifically, evaluating the impact of adverse weather on aggressive lane-changing characteristics is still lacking in majority of the current research. Hence, in an attempt to bridge the prevailing gaps in research and practice, the main goal of this paper was to evaluate lane-changing characteristics of various driver groups focusing on their aggressiveness in rain and clear weather situations. This study utilized microscopic trajectory data from SHRP2 Naturalistic Driving Study (NDS) to accomplish the research objective. Additionally, three key research tasks were conducted: (a) Developing a unique methodology to detect lane-changing events performed by the ego-vehicle, i.e., the NDS-vehicle, and extract corresponding parameters from the SHRP2 NDS database, (b) Classifying drivers into different groups based on their driving behavior via clustering approach, and (c) Developing appropriate statistical models for different driver types to identify contributing factors affecting lane-changing durations under different weather conditions.

This study offers three major contributions over the current practices, which are as follows: Firstly, given that SHRP2 NDS is a large-scale dataset, manual extraction of lane-changing events is time and labor intensive. Therefore, this study established an automated framework to extract lane-changing events from the SHRP2 database. Secondly, a methodology was proposed to classify driving behavior patterns based on aggressiveness during lane-changing maneuvers via clustering techniques. Finally, a multi-level modeling approach was presented to evaluate aggressive lane-changing maneuvers for various driver groups. To the best of the authors' knowledge, this study is one of the first attempts to leverage a multi-level modeling approach for assessing the factors contributing to lane-changing characteristics of classified driver groups considering a wide range of driver, kinematics, environmental, traffic, and roadway geometric features.

4. Data preparation and processing

The SHRP2 NDS collected a massive amount of real-time behavioral data from more than 3,400 participants between 2010 and 2013 in six US states (Wu, Dadashova, Geedipally, Pratt, & Shirazi, 2021). The participants' data were collected through an onboard Data Acquisition System (DAS), during these 3 years. DAS was developed by the Virginia Tech Transportation Institute (VTTI), which collected a wide variety of measurements, including speed, acceleration, yaw rate, lane position, as well as videos of the roadways from front-facing cameras (Ahmed et al., 2021; Campbell, 2012; Das, Ghasemzadeh, & Ahmed, 2019; Khan, Ghasemzadeh, & Ahmed, 2018; Khan, Das, & Ahmed, 2020). In addition, the Roadway Information Database (RID) was utilized, which includes roadway inventory data related to the six NDS states and supplementary data on weather information, crash histories, work zones, and active safety campaigns (Center for Transportation Research and Education (CTRE), 2015). Furthermore, driver demographic characteristics from the SHRP2 survey questionnaires were utilized along with the NDS and RID datasets.

In order to accurately classify weather-influenced SHRP2 NDS trips, various data acquisition methods were investigated initially. A windshield wiper could be used to classify weather events. However, because of the large proportion of older vehicles and errors in the DAS recording, wiper settings were not provided for most of the trips. The process of data acquisition was further complicated with the high inconsistency of various weather events under each NDS trip. In order to resolve these issues, two data acquisition methods were utilized for querying NDS trips occurring in rainy weather. The first one compiled weather data from the National Climatic Data Center (NCDC) database. An influence zone of five nautical miles around the airport weather stations was considered for identifying potential NDS trip locations occurring in rainy weather. The latter one utilized weather-influenced crash locations stated in the RID database to isolate rainy weather events. The detailed information regarding the two methods can be found in (Ahmed et al., 2017, 2018). For this study, 89 NDS trips in rainy weather and their matching 178 trips in clear weather were successfully acquired and processed via the two methods.

The primary stage of the data processing was to detect lane-changing events from the received NDS trips. Identifying and extracting lane-changing events from the NDS data was one of the challenging tasks in this study. An automated lane change extraction algorithm was developed in this regard. For this purpose, a lane position offset (LO) variable was taken as the indicator variable for identifying lane-changing events (InSight SHRP2 NDS, 2019). Several steps were considered to develop the algorithm. First, missing LO data in NDS trips were imputed with the "*fillmissing*" function of MATLAB. The method helped to impute missing data using linear interpolation of adjacent non-missing data points (MathWorks, 2021). Afterward, the peak (LO_P) of the LO was determined using the "*findpeaks*" function of MATLAB. For each possible lane-changing event, two peaks (i.e., positive and negative) were found. However, the algorithm was developed based on the positive peaks. In addition, the direction of lane-changing events (i.e., left to the right/right to the left) were identified using the positive peaks. After defining the peak, the change of LO (i.e.,

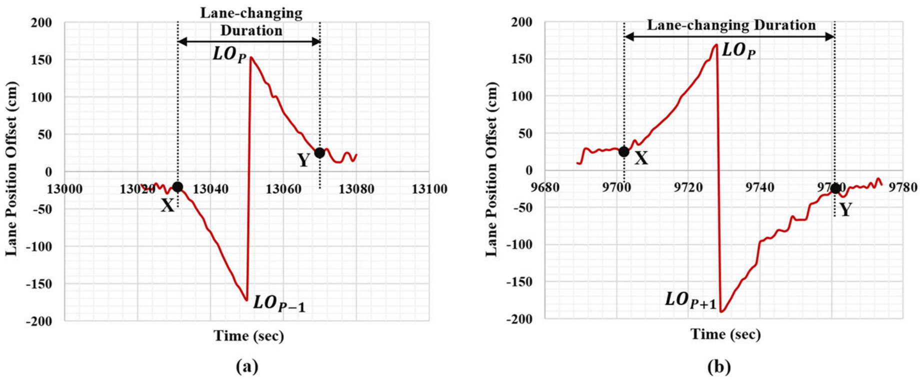

Figure 1. Automatic identification and extraction of lane-changing events from NDS database. (a) Lane-changing from right to the neighboring left and (b) lane-changing from left to the neighboring right.

absolute difference) between the positive peak (LO_P) and immediate before (LO_{P-1}) timestamp in the NDS time-series data was continuously monitored, as provided in Equation 1. Similarly, change of LO between the positive peak and after (LO_{P+1}) timestamp was monitored continuously from time-series data (Equation 2). The difference (m_1 or m_2) was checked with a threshold of 200 cm, as suggested by a previous study (Ghasemzadeh & Ahmed, 2018). Subsequently, three possible scenarios were investigated.

$$|LO_p - LO_{p-1}| = m_1 \tag{1}$$

$$|LO_p - LO_{p+1}| = m_2 \tag{2}$$

- Scenario #1: If both m_1 *and* m_2 were less than 200 *cm, no lane-changing event occurred*
- Scenario #2: If $m_1 > 200$ *cm, lane-changing occurred from right to left* (Figure 1a)
- Scenario #3: If $m_2 > 200$ *cm, lane-changing occurred from left to right* (Figure 1b)

Figure 1 shows two samples of a lane-changing maneuver extracted using the automated algorithm under heavy rain conditions. Once the lane-changing events were detected, the duration of the event was computed by observing the gradual decrease of LO from the two peaks. For instance, Figure 1a denotes a lane-changing event in which the driver changed his/her lane from right to the neighboring left lane. A gradual decrease from LO_P in the forward direction and gradual decrease from LO_{P-1} in the backward direction was perceived in the database, as shown in Figure 1a. The process of gradual decreases completed at X and Y, which was defined to

Figure 2. Light/moderate rain. Raindrops could be visible; wipers are active/low swipe rate; road surfaces are dry/wet; visibility is clear/moderate; road markings and information on road signs and vehicle(s) ahead could be recognized.

be the end of the lane-changing event where X and Y represent the start and end of the event, respectively. The duration between X and Y was measured as the lane-changing duration. On the contrary, Figure 1b illustrates a lane-changing event from left to the neighboring right lane. The gradual decrease from LO_P in the backward direction and gradual decrease from LO_{P+1} in the forward direction was monitored until the decrease ended at X and Y, as shown in Figure 1b. Subsequently, the lane-changing duration was calculated in the same manner. It is worth mentioning that the completion of a lane-changing event was defined when the gradual decrease was ended and the value of LO was stabilized within ±25 cm. This threshold was considered due to the fact that LO of the vehicles varied within ±25 cm while maintaining lanes from the NDS data.

The developed automatic identification algorithm was applied to the acquired NDS trips via MATLAB platform and corresponding features (i.e., mean, minimum, maximum, and standard deviation of speed, longitudinal acceleration, lateral acceleration, and yaw rate) during each lane changing event were obtained through the algorithm. Using the developed algorithm, in total 1,410 lane-changing events were effectively extracted. The subsequent step of data processing was to manually verify and confirm the identified lane-changing events. To do that, our research team developed a software named the Wyoming NDS Visualization and Visibility Identification Tool (Ahmed et al., 2015; Das, Ahmed, Ghasemzadeh, 2019; Das, Khan, & Ahmed, 2020; Ghasemzadeh, Hammit, Ahmed, & Eldeeb, 2019). During the manual verification and annotation, any events related to lane departures (e.g., vehicles swerving right and left of the lane) were identified via the NDS forward-facing videos and trajectory data using the developed visualization tool, and those events were discarded from the final dataset. A number of variables were verified and coded during the process, including different weather levels, visibility, surface, traffic states, and lane-changing event types associated with individual lane-changing events for every trip. Note that manual annotation was an expensive and time-consuming process. In order to achieve accurate annotation, several criteria were developed considering the qualitative based measures observed from the NDS front-facing video. For instance, weather conditions were classified

Figure 3. Heavy rain. Raindrops are clearly visible; wipers are active/high swipe rate; road surfaces are wet; visibility is affected; road markings and information on road signs and vehicle(s) ahead could not be recognized.

Figure 4. Clear weather. Visibility is high; road signs; markings; and surroundings are visible.

as light/moderate rain, heavy rain, and clear weather depending on roadway surroundings (i.e., guardrail, delineators, etc.), visibility of lane and shoulder markings, and identifiable road signs, as illustrated in Figures 2–4. Moreover, thorough training was provided to the research team, which included detailed written descriptions and numerous sample images to reduce any potential bias and to maintain consistency in the manual observation and annotation process. It is worth mentioning that windshield wipers' status and swipe rate were also considered to categorize the rainy weather conditions during the manual observation process. Weather conditions with active wipers and low swipe rate were considered as light/moderate rain. In contrast, heavy rain conditions were identified when the wipers were active at a high swipe rate. It is also worth pointing out that the validation of wipers' status (i.e., on/off) and presence of rain was conducted through leveraging our developed Wyoming NDS Visualization and Visibility Identification Tool. Utilizing the tool, the authors verified each of the trips in rain to ensure that the wipers are on. For more details about the data validation and reduction, please see (Ghasemzadeh et al., 2019).

Traffic conditions were also manually annotated from the selected NDS trips and categorized into three states including free-flow (Level of Service A and B), congested-flow (Level of Service C and D), and unstable-flow conditions (Level of Service E and F) considering the recommendations provided in Highway Capacity Manual (HCM). Once the manual data processing was completed, environmental and traffic conditions, as well as lane-changing event types were annotated for each verified lane-changing event segment to develop a comprehensive lane-changing events database. Afterward, RID and demographic characteristics of the SHRP2 participants

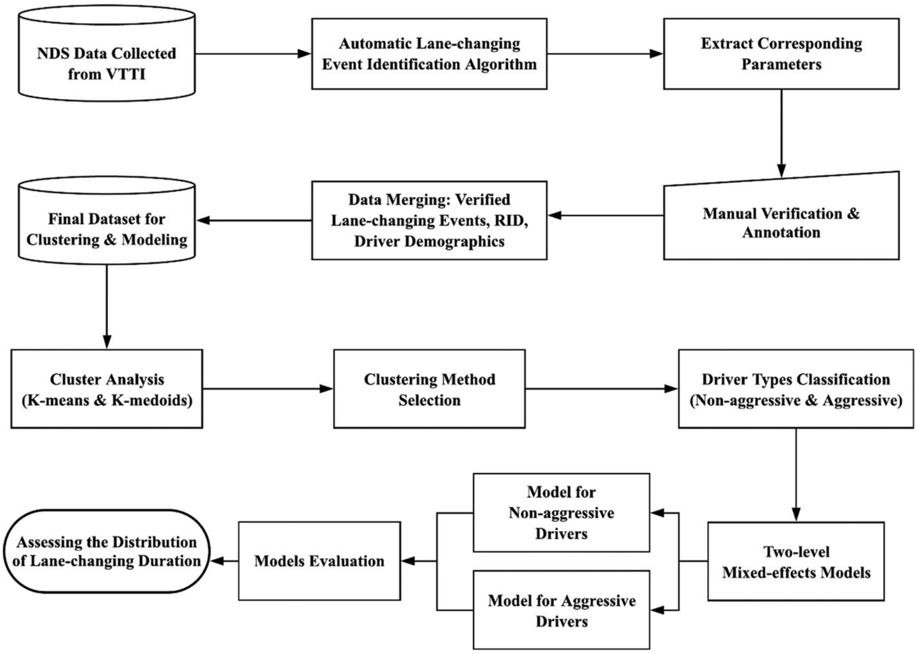

Figure 5. Flow chart summarizing the overall methodological approach from data preparation to analysis process.

were added to the database, which served as the basis for cluster analysis and statistical modeling. The overall methodological process from data preparation to analysis is presented in Figure 5.

5. Summary statistics

Considering the fact that unstable flow conditions (i.e., Level of Service E and F) might not represent driver discretionary lane-changing maneuvers, they were removed from the dataset as suggested by the authors' previous study (A. Das & Ahmed, 2021). As mentioned previously, a total of 1,410 lane-changing events were extracted using the automatic identification algorithm. All the lane-changing events were manually annotated as mandatory and discretionary considering various types of maneuvers identified from video observations including overtaking a slower lead vehicle, returning to desired driving lane, freeway entering and exiting, adding lane(s), and additional probable reasons, as stated in Lee et al. (2004). In total, 84 lane-changing events were characterized as mandatory lane-changing events. Since discretionary lane-changing behavior requires decision-making processes of different drivers interacting with each other that introduces heterogeneity, it is complicated with respect to safety, operations, and efficiency perspectives compared to mandatory behavior. Thus, mandatory maneuvers were discarded and all the following analyses were conducted based on

Table 1. Summary statistics of lane-changing maneuvers.

	Free-flow conditions				Congested-flow conditions			
	Light/ Moderate rain	Heavy rain	Clear weather	Total	Light/ Moderate rain	Heavy rain	Clear weather	Total
Num. of associated drivers	32	15	52	55	30	11	45	48
Total num. of lane-changing events segments	186	113	609	908	98	39	281	418
Total length of traversed routes (miles)	675.61	406.83	2114.64	3197.08	293.38	90.41	888.54	1272.33
Mean lane-changing events/mile	0.27	0.28	0.29	0.28	0.33	0.43	0.32	0.33
Total traveled duration (hr)	11.62	7.15	34.62	53.38	6.08	1.82	15.28	23.18

1,326 discretionary lane-changing events (284 in light/moderate rain, 152 in heavy rain, and 890 in clear weather) in free-flow and congested-flow conditions (Hill et al., 2015; Shao, Zhang, Feng, & Dong, 2020). Table 1 shows the breakdown of these 1,326 discretionary lane-changing events under various weather and traffic. After analyzing the segments of the lane-changing events in different traffic conditions, it was found that 908 lane-changing events (186 in light/moderate rain, 113 in heavy rain, and 609 in clear weather) occurred on free-flow and 418 lane-changing events (98 in light/moderate rain, 39 in heavy rain, and 281 in clear weather) occurred on congested-flow conditions. Table 1 also summarizes the summary statistics of lane-changing maneuvers with a total length of traversed routes and total travel time of the NDS trips considered for the study. In total, 908 lane-changing events corresponded to 3,197 miles on freeways and traveled over approximately 53 hr during free-flow conditions. In addition, 418 lane-changing events equivalent to 1,272 miles on freeways over nearly 23 hr of driving were found under congested-flow traffic conditions.

6. Methodology

6.1. Clustering approach

Cluster analysis was used to classify drivers into various groups. These groups were then characterized by their aggressive driving behavior when performing lane-changing maneuvers based upon the cluster means. This study utilized two well-known clustering techniques-K-means and K-medoids.

6.1.1. K-means clustering

The K-means is the widely used clustering technique that solves cluster problems utilizing unsupervised learning technique. The K-means clustering algorithm optimally clusters n observations into k groups. Subsequently, each cluster centroid and the number of observations within each cluster is determined through an iterative process (Cai, Chang, Gao, & Zhou, 2021). The iterative process is designed to minimize the within-cluster sum of squares (WSS) given by Equation 3.

$$\mathrm{WSS} = \sum_{l=1}^{q}\sum_{j=1}^{k}\sum\nolimits_{i\in C_j}(x_{il}-\overline{x}_l^{(j)})^2 \tag{3}$$

where x_{il} represents feature l for observation i, C_j denotes cluster j, and $\overline{x}_l^{(j)}$ is the mean of the observations in C_j (Everitt & Hothorn, 2011). The number of clusters k must be specified in order to implement K-means clustering.

6.1.2. K-medoids clustering

In contrast to the K-means clustering, the K-medoids method selects representative observations as centers, which are called medoids and observations cluster around those medoids. The K-medoids clustering can be used with random dissimilarity measures. The error function of this method is calculated by the summation of the dissimilarities between each observation and its respective medoid (Mohammadnazar, Arvin, & Khattak, 2021; Park & Jun, 2009), as shown in Equation 4.

$$E = \sum_{i=1}^{k}\sum_{x\in C_i}(x - M(C_i)) \tag{4}$$

where C_i denotes cluster k and $M(C_i)$ represents the medoid of the cluster C_i. Note that several algorithms were proposed for K-medoids clustering; however, this study considered one of the widely used K-medoids algorithms called Partitioning Around Medoids (PAM) (Kaufman & Rousseeuw, 2009).

6.1.3. Selection of clustering method and number of clusters

This study utilized the average silhouette width method (ASWM) to select the appropriate clustering for the analysis. ASWM defines how well each observation lies within its cluster ranging from -1 to $+1$. A higher value of ASWM represents a good clustering with regard to within-cluster homogeneity and between-cluster separation. If the dataset is clustered in k clusters, the silhouette coefficient of observation i can be calculated by Equation 5.

$$S_i = \frac{b_i-a_i}{\mathrm{Max}(a_i, b_i)} \tag{5}$$

where S_i is the silhouette width of the observation i, a_i indicates the average distance between i and all other observations in the same cluster, and b_i denotes the smallest average distance between observation i and all observations in any other clusters (Kaufman & Rousseeuw, 2009;

Mohammadnazar et al., 2021). Equations 6 and 7 compute the average silhouette width of each cluster (SWM) and average silhouette width of all clusters (i.e., ASWM), respectively.

$$SWM_j = \frac{1}{n}\sum_{i=1}^{n} S_i \tag{6}$$

$$ASWM = \frac{1}{k}\sum_{j=1}^{k} S_j \tag{7}$$

where n represents the number of observations in the same cluster and k indicates the number of all clusters (Kaufman & Rousseeuw, 2009; Mohammadnazar et al., 2021).

It is worth mentioning that determining the number of clusters is one of the critical steps of any clustering approach. In general, number of clusters are determined based on the measures of clustering quality or the subjective/engineering judgment of the researchers (Mohammadnazar et al., 2021). Following the latter approach, this study considered two clusters (i.e., non-aggressive and aggressive) corresponding to various groups of weather and traffic conditions, which is more consistent with previous research conducted on classifying driver types (Galovski & Blanchard, 2002; Gan, Weng, Li, & Han, 2020; Islam & Mannering, 2020; Liu & Lee, 2005; Zhou, Fu, & Wang, 2020). The details of these two categories, along with the feature characteristics, are discussed in the Results and Discussions section.

6.2. Multi-level mixed-effects linear regression model

A multi-level mixed-effects linear regression model was utilized to identify the contributing factors affecting driver lane-changing durations. This approach was necessary since a driver in this study could have multiple lane-changing events in different weather and traffic conditions. As a result, correlation would be expected on these repeated observations for the same driver. This study used a single random effect for driver, or random intercept, to account for this correlation (S. Chen, Chen, & Xing, 2022; Gan et al., 2020). Additional random effects involving the predictors, or random slopes, were also examined but were not found to improve the model according to the Akaike Information Criterion (AIC). A two-level random intercept model was thus developed consisting of the lane-changing event level and the driver level, which can be expressed as

Table 2. Comparisons of the performance of K-means and K-medoids clustering using ASWM.

		ASWM	
Traffic conditions	Weather conditions	K-means clustering	K-medoids clustering
Free-flow	Light/Moderate rain	0.475	0.174
	Heavy rain	0.490	0.205
	Clear weather	0.393	0.218
Congested-flow	Light/Moderate rain	0.529	0.193
	Heavy rain	0.466	0.368
	CLEAR weather	0.471	0.155

Level 1: Lane-changing event

$$Y_{ij} = \beta_{0j} + \sum_{l=1}^{q} \beta_{jl} x_{ijl} + \varepsilon_{ij} \tag{8}$$

Level 2: Driver level

$$\beta_{0j} = \gamma_{00} + u_{0j}, \beta_{jl} = \gamma_{0l} \tag{9}$$

Equations (8) and (9) can be combined to obtain

$$Y_{ij} = \gamma_{00} + \sum_{l=1}^{q} \gamma_{0l} x_{ijl} + u_{0j} + \varepsilon_{ij} \tag{10}$$

where Y_{ij} is the lane-changing duration for event i and driver j, γ_{00} is the combined model intercept, γ_{0l} is the fixed slope coefficient for explanatory variable x_l, u_{0j} is the driver level random intercept, ε_{ij} is the lane-changing event level error term (Blissett, 2017; Goldstein, 2011). Equation 10 represents the random intercept model with fixed slope coefficients. It is assumed that the random terms u_{0j} and ε_{ij} are statistically independent. An intraclass correlation coefficient (ICC) can be used to measure the degree of correlation among observations within a specific group (Deligianni, Quddus, Morris, Anvuur, & Reed, 2017). For a two-level random intercept model, ICC can be calculated using Equation 11.

$$ICC = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_{\varepsilon}^2} \tag{11}$$

where $\sigma_{u_0}^2$ is the intercept variance of the random intercept model at level 2 (i.e., driver level) and σ_{ε}^2 is the variance of the level 1 residuals (i.e., lane-changing event).

7. Results and discussions

7.1. Cluster analysis

K-means and K-medoids clustering were performed utilizing "*cluster*" and "*factoextra*" packages in R® statistical software. It is worth mentioning that each observation indicated a lane-changing event performed by the driver that was categorized based on their aggressiveness. Table 2 illustrates the ASWM values for two cluster classifications (i.e., $k = 2$) in clear and rainy weather under various traffic conditions using K-means and K-medoids clustering methods. As shown in Table 2, higher ASWM values were obtained in K-means clustering representing that this clustering algorithm produced relatively better classification compared to K-medoids. Hence, drivers were classified into non-aggressive (NA) and aggressive (A) categories based upon their driving behaviors considering K-means clustering instead of K-medoids clustering. It is worth pointing out that identifying which cluster represented which driver type was determined by the researcher's judgment and the majority of the voting in the classification features. Utilizing these approaches, two clusters were allocated as aggressive and non-aggressive driver types.

Selecting the necessary features in determining aggressive driving behavior was a crucial step. The lane-changing event database involved a wide range of variables that could be considered to categorize drivers. Additionally, various features can be found in the literature to identify driver types. Therefore, variable importance from unsupervised random forests was applied to select the variables to be used in the cluster analysis. Utilizing the unsupervised random forests and following the findings from previous studies, six different features were chosen for driver classification through clustering, including mean speed, maximum longitudinal acceleration, maximum lateral acceleration, and mean yaw rate along with the mean number of lane-changing events and mean speed differences (i.e., the average speed of the driver minus the posted speed limit) (Higgs & Abbas, 2013; Hill et al., 2015). Note that drivers could also be grouped based on their demographics (Hossain et al., 2022). However, only variables directly associated with driving behavior were utilized since the purpose underlying the cluster analysis was to obtain groups of drivers who were non-aggressive or aggressive in their lane-changing maneuvers under different weather and traffic states. Table 3 provides the cluster means of non-aggressive and aggressive drivers with respect to the six features. It can be observed from Table 3 that comparatively higher feature values were obtained for aggressive drivers in majority of the cases compared to non-aggressive drivers. In this regard, Multivariate Analysis of Variance (MANOVA) was used to test the equality of the population means among non-aggressive and aggressive drivers using the six features. According to the MANOVA test, it was found that the population means of the six features were significantly different between the two driver groups in various weather and traffic states. Therefore, the identified driver types were determined to be different on average with respect to these features according to the statistical test and could be

Table 3. Cluster means of non-aggressive and aggressive driver.

		Mean of the features							
Traffic conditions	Weather conditions	Mean speed (mph)	Maximum longitudinal acceleration (g)	Maximum lateral acceleration (g)	Mean yaw rate (deg/s)	Mean num. of lane-changing per mile	Mean speed differences (mph)	MANOVA-Pillai test (Pr > F)	Driver category
Free-flow	Light/Moderate rain	56.173	0.091	0.1025	−0.496	0.432	−2.504	<0.0001	NA
		68.828	0.081	0.131	−0.291	0.288	3.206		A
	Heavy rain	53.985	0.081	0.094	−1.672	0.392	−2.746	0.0160	NA
		65.712	0.255	0.148	0.081	0.466	2.157		A
	Clear weather	59.769	0.162	0.142	−0.155	0.294	1.034	<0.0001	NA
		69.517	0.085	0.138	−0.307	0.409	3.809		A
Congested-flow	Light/Moderate rain	30.938	0.118	0.094	−0.287	0.411	−18.976	<0.0001	NA
		57.693	0.071	0.117	−0.052	0.556	−6.280		A
	Heavy rain	63.521	0.081	0.090	−0.195	0.569	−4.205	0.0766*	NA
		46.534	0.129	0.144	0.324	0.634	−7.746		A
	Clear weather	49.233	0.138	0.148	−0.388	0.295	−6.548	<0.0001	NA
		64.677	0.092	0.137	−0.127	0.479	−0.498		A

NA = Non-aggressive, A = Aggressive.
*Significant at 90% confidence level.

effectively considered for further analysis. Table 3 also shows the MANOVA results for the two driver types under various weather and traffic conditions.

7.2. Modeling lane-changing durations

Once the drivers were classified by the K-means cluster analysis, the associated driver type was added to the lane-changing event database containing the weather and traffic conditions. A two-level mixed-effects linear regression model was then used to develop the lane-changing duration model for each driver type. The study hypothesis was that the contributing factors to lane-changing durations would be different for each driver type. Thus, separate models were used to more easily illustrate these differences and to simplify the interpretations. The explanatory variables are the possible causal factors that could impact the lane-changing durations. These include vehicle kinematics, environmental and traffic characteristics, driver demographics, and roadway characteristics, as shown in Table 4. Note that variable importance obtained from supervised random forests was used to select the predictors from among the numerous kinematic variables where the response variable was lane-changing duration. Table 4 exhibits the selected kinematic variables along with additional variables for modeling lane-changing duration corresponding to non-aggressive and aggressive drivers.

The Variance Inflation Factor (VIF) was checked to assess multicollinearity among the predictors. The VIF of all the predictors used in the two models fell below 3, which provided evidence that multicollinearity was not an issue for these models (Kutner, Nachtsheim, & Neter, 2004).

7.2.1. Lane-changing duration model for non-aggressive drivers

The lane-changing durations are frequently assumed to follow a lognormal distribution (Toledo & Zohar, 2007) This distributional assumption was found to also be reasonable in this study based upon residual diagnostic checks. Therefore, the response variable (i.e., lane-changing duration in seconds) was log-transformed before developing the models for both driver types. For convenience, lane-changing duration in this study refers to lane-changing duration in log seconds. Several random effect terms were evaluated. However, a single random effect with driver only was used based on AIC. This mixed-effect model for the lane-changing durations was developed using the "*lme4*" package in the R® statistical software. The importance of the fixed-effect terms was evaluated using the type III Analysis of Variance (ANOVA) with Satterthwaite degrees of freedom (Kuznetsova, Brockhoff, & Christensen, 2019). Only statistically significant fixed-effect terms were retained in the final model. The heterogeneity associated with

Table 4. Variables descriptions for lane-changing duration models.

Variable	Description	Type	Levels (*Reference)
Response variable			
Lane-changing Duration (s)	The time requires to perform a lane-changing maneuver	Continuous	–
Explanatory variables			
Vehicle kinematics factors			
Std. speed (mph)	Standard deviation of speed during lane-changing event	Continuous	–
Min. longitudinal acceleration (g)	Minimum acceleration/deceleration in the longitudinal direction versus time during lane-changing event	Continuous	–
Std. lateral acceleration (g)	Standard deviation of acceleration/deceleration in the lateral direction versus time during lane-changing event	Continuous	–
Std. yaw rate (deg/s)	Standard deviation of vehicle angular velocity around the vertical axis during lane-changing event	Continuous	–
Environmental & traffic characteristics			
Weather condition	Weather condition during lane-changing event	Categorical	Clear* Light/Moderate rain Heavy rain
Traffic conditions	Traffic condition during lane-changing event	Categorical	Free-flow (LOS A&B)* Congested-flow (LOS C&D)
Driver demographics			
Gender	The participant's gender	Categorical	Male* Female
Driving experience	The participant's number of driving years	Categorical	$\leq$10 years* $>$10 years
Driver mileage last details	The approximate number of miles the participant drove last year	Categorical	$<$10,000 miles* 10,000–20,000 miles $>$20,000 miles
Roadway characteristics			
Presence of curve	Whether the participants drove curve or tangent during lane-changing event	Categorical	Tangent* Curve
Speed limit (mph)	Speed limit during lane-changing event	Categorical	$\leq$60 mph* $>$60 mph
Number of freeway lanes	Number of lanes during lane-changing event	Continuous	–

Table 5. Two-level mixed-effects linear regression model of lane-changing duration for non-aggressive drivers.

Fixed-effect parameter	Description	Estimate	Std. error	t value	Pr(>|t|)	Confidence interval 2.5%	Confidence interval 97.5%
Intercept	–	1.5179	0.1283	11.833	0.0000	1.2665	1.7693
Std. speed	–	0.0544	0.012	4.5205	0.0000	0.0308	0.078
Min. longitudinal acc.	–	−1.2322	0.2551	−4.8306	0.0000	−1.7322	−0.7323
Std. lateral acc.	–	−2.4312	1.1879	−2.0466	0.0412	−4.7594	−0.1029
Std. yaw rate	–	0.0879	0.0494	1.7767	0.0762*	−0.0091	0.1848
Weather cond.	Light/Moderate rain	0.3921	0.1164	3.3702	0.0008	0.1641	0.6202
Traffic cond.	Congested-flow	−0.1787	0.0628	−2.8449	0.0048	−0.3019	−0.0556
Gender	Female	−0.2433	0.1075	−2.2623	0.0303	−0.4541	−0.0325
Driver mileage last details	>20,000 miles	−0.4787	0.1622	−2.9517	0.0047	−0.7965	−0.1608
Driving experience	>10 years	0.2022	0.0684	2.9578	0.0074	0.0682	0.3362
Presence of curve	Tangent	−0.1467	0.0721	−2.0336	0.0425	−0.2881	−0.0053
Number of freeway lanes	–	0.0528	0.0292	1.8056	0.0717*	−0.0045	0.1101
Weather cond. × Number of freeway lanes	Light/Moderate rain, Number of freeway lanes	−0.1359	0.0443	−3.0698	0.0023	−0.2226	−0.0491
Gender × Driver mileage last details	Female, >20,000 miles	0.4582	0.1831	2.502	0.0168	0.0993	0.8172
Gender × presence of curve	Female, Curve	0.1505	0.0808	1.8639	0.0629*	−0.0078	0.3088

*Significant at 90% confidence level.

the drivers was incorporated through increased standard errors associated with the effects of the contributory factors, which impact the t-values, p-values, and confidence intervals. Table 5 shows the results of the two-level mixed-effect model for non-aggressive drivers.

The ICC value was found to be 8%, indicating that an estimated 8% of the variation in the lane-changing duration is explained by incorporating driver heterogeneity in the model for non-aggressive drivers. In other words, 8% of unobserved heterogeneity in lane-changing durations resulted from between driver variance. As can be seen in Table 5, several fixed-effect terms had significant effects on the lane-changing duration for non-aggressive drivers. For instance, an increase in standard deviation of speed increases the lane-changing duration of non-aggressive drivers. This increase might be because drivers adjusted their speeds from several seconds before beginning to the end of lane-changing maneuver according to surrounding traffic conditions, which resulted in a higher variation of speed. The standard deviation of lateral acceleration had a significant impact on lane-changing duration of non-aggressive drivers. The negative coefficient of the standard deviation of lateral acceleration indicated that higher variation of lateral acceleration decreased the lane-changing duration of non-aggressive drivers. Traffic conditions were found to have a significant negative effect on the lane-changing durations. During congested traffic conditions, drivers had to change lanes with smaller gaps. Therefore, they were forced to change lanes quickly resulting in lower lane-changing durations (X. Wang, Yang, & Hurwitz, 2019). Interestingly, driving experience had a positive effect on the lane-changing duration. The results revealed that experienced non-aggressive drivers took more time to change lanes compared to less experienced drivers indicating the cautious behavior of the experienced drivers resulted in longer time to complete the maneuver.

Interactions terms were also examined in the model to understand the combined effect of those variables in the lane-changing durations. The interaction between weather conditions and number of freeway lanes, gender and driver mileage last year, and gender and presence of curve were found to affect the lane-changing duration for non-aggressive drivers. For example, the interaction between weather conditions and number of freeway lanes indicated that the effect of weather conditions on the lane-changing duration changes with the number of freeway lanes. The result is not surprising as by increasing the number of lanes, drivers are not restricted and have more opportunity and flexibility for changing lanes rapidly in adverse weather conditions. In addition, the interaction between gender and driver mileage revealed that there is an increase in lane-changing duration for female drivers who drove more than 20,000 miles last year compared to male drivers who drove more than 20,000 miles last year. Finally, a positive estimate of the interaction between gender and presence of curve indicated that the lane change durations on curve were higher for female drivers compared to male drivers on straight segments. The findings of the higher lane-changing durations of female drivers could represent their better risk perception than male drivers as observed in the naturalistic driving environment (Lee et al., 2004).

7.2.2. Lane-changing duration model for aggressive drivers

A two-level mixed-effect model of the lane-changing duration for aggressive drivers was also developed. Similar to the non-aggressive drivers, type III ANOVA with Satterthwaite degrees of freedom method was used to check the importance of the fixed-effect terms. The developed model is presented in Table 6. Only statistically significant fixed-effect terms were retained in the final model.

The ICC value was found to be 11%, indicating that an estimated 11% of the variation in lane-changing duration is explained by incorporating driver heterogeneity in the model for aggressive drivers. The results provided in Table 6 revealed that several fixed-effect terms, including main-effect and interaction terms, had effects on the lane-changing duration for aggressive drivers. Similar to the non-aggressive drivers, the standard deviation of speed, minimum longitudinal acceleration, traffic conditions, and driving experience were found to affect the lane-changing durations for these types of drivers. For instance, a positive estimate of the standard deviation of speed implied that lane change durations of aggressive drivers increased

Table 6. Two-level mixed-effects linear regression model of lane-changing duration for aggressive drivers.

Fixed-effect parameter	Description	Estimate	Std. error	t value	Pr(>|t|)	Confidence interval 2.5%	Confidence interval 97.5%
Intercept	–	1.6616	0.089	18.6687	0.0000	1.4872	1.8361
Std. speed	–	0.1049	0.0118	8.8526	0.0000	0.0816	0.1281
Min. longitudinal acc.	–	−1.4997	0.2685	−5.5855	0.0000	−2.026	−0.9735
Std. lateral acc.	–	−10.5877	1.9458	−5.4413	0.0000	−14.4013	−6.774
Weather cond.	Light/Moderate rain	−0.1765	0.1018	−1.735	0.0831*	−0.376	0.0229
Traffic cond.	Congested-flow	−0.1327	0.0369	−3.5912	0.0004	−0.2051	−0.0603
Driving experience	>10 years	0.1638	0.0557	2.9402	0.0060	0.0546	0.273
Speed limit	>60 mph	0.2723	0.1039	2.6206	0.0090	0.0687	0.476
Weather cond. × Std. lateral acc.	Light/Moderate rain, Std. lateral acc.	14.1593	3.5214	4.0209	0.0001	7.2574	21.0611
Weather cond. × Std. yaw rate	Light/Moderate rain, Std. yaw rate	−0.4995	0.155	−3.2219	0.0013	−0.8034	−0.1956
Weather cond. × Speed limit	Heavy rain, >60 mph	−0.4035	0.1483	−2.7212	0.0067	−0.6941	−0.1129
Speed limit × Number of freeway lanes	>60 mph, Number of freeway lanes	−0.0822	0.035	−2.3474	0.0192	−0.1508	−0.0136

*Significant at 90% confidence level.

due to higher variation of speed. Note that this increase in lane changing duration is higher compared to non-aggressive drivers. It was also observed that lane-changing durations of aggressive drivers tended to decrease during congested-flow conditions. In addition, driving experience had a positive effect on the lane-changing duration of aggressive drivers indicating that experienced aggressive drivers were more cautious in changing lanes similar to non-aggressive experienced drivers.

It is worth noting that several interaction terms including weather conditions with standard deviation of lateral acceleration, standard deviation of yaw rate, and speed limit, as well as speed limit and number of freeway lanes were also found to be significant. As can be seen in Table 6, the effect of weather conditions on the lane-changing durations for aggressive drivers changes with the standard deviation of lateral acceleration. For example, the positive estimate of the interaction between light/moderate rain and standard deviation of lateral acceleration indicated that lane-changing durations during light/moderate rain increased with the increase of standard deviation of lateral acceleration compared to clear weather. This could be explained by the fact that drivers had poor steering control in adverse weather conditions. Therefore, the variation of lateral acceleration was higher, which resulted in a higher lane-changing duration. This finding could be supported by the result of a recent study that described the influence of adverse weather on drivers' perceived risk and revealed that drivers had difficulties in-vehicle operation under poor visibility in adverse weather conditions (C. Chen, Zhao, Liu, Ren, & Liu, 2019). Additionally, the interaction between weather conditions and speed limit had a negative estimate, which indicated that the lane-changing durations decreased on freeways with a higher speed limit under heavy rain conditions compared to clear weather. This finding indicated aggressive driving behavior specifically since drivers took less time to change lanes even in heavy rain conditions on roadways with higher speed limits. A recent study concentrated on assessing the impacts of driving environment on driving behavior suggested that even though drivers tended to drive less aggressively for higher rain intensities, the impact of rain intensity on driving behavior depends on roadways with a variety of speed limits (Faria, Baptista, Farias, & Pereira, 2020). This could justify the lower lane-changing durations of aggressive drivers considering their behavioral characteristics on freeways with a higher speed limit under heavy rain. Considering the interaction between speed limit and number of freeway lanes, it was found that drivers had lower lane-changing durations with the increased number of lanes in higher speed limit.

7.2.3. Assessing the distribution of lane-changing duration

The distribution of the lane-changing durations is of particular interest for the calibration and validation of microsimulation models. As mentioned previously, studies have suggested that lane-changing duration has a log-normal distribution. Since this study used lane-changing duration in log seconds, then it is anticipated that these lane-changing durations should be normally distributed. To correctly assess this distributional assumption, it is important to first remove the effects due to driver, vehicle kinematics, environmental and traffic characteristics, driver demographics, and roadway characteristics so that the observations are close to independent and identically distributed. Thus, standardized residuals were used to assess the assumption of normality based upon the multi-level model fits for the non-

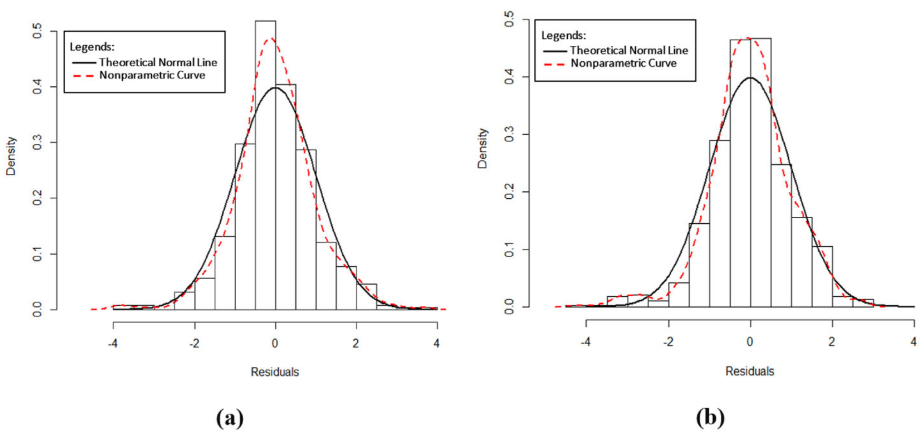

Figure 6. Standardized residuals from the model fits for (a) non-aggressive drivers and (b) aggressive drivers.

aggressive and aggressive drivers. Figure 6 presents the distributions of model fit for both non-aggressive and aggressive drivers based upon these standardized residuals. The black solid line indicates the theoretical standard normal line and the red dashed line represents no assumption on any distribution (i.e., a nonparametric curve). As can be seen in Figure 6, the model of lane-changing durations for aggressive drivers has a marginal deviation at the left tail. In general, Figure 6 provides a good evidence that the lane-changing durations are normally distributed as the theoretical standard normal lines approximate the histogram of the standardized residuals fairly well for the non-aggressive and aggressive drivers.

8. Conclusions

The study explored lane-changing characteristics of different driver types during rain and clear weather conditions by leveraging naturalistic driving data. The study is one of the first to comprehensively investigate drivers' aggressiveness in lane-changing behavior and to identify factors affecting lane-changing characteristics for different driver types. An innovative methodology was developed in this study to automatically extract lane-changing events from the SHRP2 NDS database. Utilizing the developed methodology, 1,326 discretionary lane-changing events, which included 284 in light/moderate rain, 152 in heavy rain, and 890 in clear weather, were effectively identified from 89 trips in rain and additional 178 trips in clear weather conditions.

The K-means and K-medoids clustering methods were performed and K-means clustering was adopted based on the average silhouette width method to classify drivers into non-aggressive and aggressive considering six different features, including mean speed, maximum longitudinal acceleration, maximum lateral acceleration, and mean yaw rate along with the mean number of lane-changing events and mean speed differences from the speed limit in different weather and traffic conditions. The findings from cluster analysis results revealed that the majority of the features produced higher values for aggressive drivers in contrast with non-aggressive drivers. Using the MANOVA test, the clustering outcomes were verified, which confirmed that the population means associated with the two clusters were significantly different for the various weather and traffic conditions.

The two-level mixed-effects linear regression models were calibrated for both driver types to assess the contributing factors that affect lane-changing durations of each driver type under rain and clear weather conditions. The findings suggested that various vehicle kinematics, traffic, driver, and roadway features, as well as weather conditions combined with other factors, were significant in the calibrated models for both driver groups. For example, the results indicated that lane-changing durations were expected to decrease with congested traffic conditions, which indicated that drivers were forced to change lanes due to smaller gaps. In addition, it was found that the lane-changing duration decreased with a higher speed limit under heavy rain compared to clear weather conditions for aggressive drivers. Moreover, lane-changing durations decreased in light/moderate rain conditions with the increase of the number of lanes for non-aggressive drivers. The distribution of the lane-changing durations was evaluated based upon standardized residuals from these model fits. These evaluations demonstrated that lane-changing durations could be modeled with the normal distribution on the log second scale or the lognormal distribution on the second scale.

The clustering analysis coupled with multi-level modeling conducted in this study unlocks a new perspective to understand drivers' aggressiveness during performing lane-changing maneuvers and the findings have several practical implications, especially in the area of microsimulation modeling. Based on the analysis of standardized residuals from the multi-level model fits associated with non-aggressive and aggressive drivers, the study found that normal distribution (in log seconds) should be used in modeling lane-changing durations. This finding could be effectively leveraged to calibrate and validate lane-changing microsimulation models. More specifically, the microsimulation model could be developed from the estimates obtained in this study according to Equation (10), i.e., the lane-changing event-level error term and driver level random intercept could be simulated from normal distributions. These terms could be added with the estimated values for the relevant covariates (e.g., different weather and traffic conditions) to obtain a simulated lane-changing duration (in log seconds) for non-aggressive and aggressive drivers. Recently, Connected and Autonomous Vehicles (CAV) have emerged as one of the potential sources of trajectory-level data like naturalistic driving studies. The clustering results could be applied to classify driver types (i.e., non-aggressive and aggressive) in CAV in order to enhance the safety of the vehicles considering various conditions. In this regard, car manufacturers, as well as various public and private agencies working on the development of CAV technology, could leverage the information to improve the design of Advanced Driving Assistance Systems (ADAS) in monitoring driving behavior and providing appropriate feedback regarding their characteristics during lane-changing maneuvers. Additionally, with the clustering results, Traffic Management Centers (TMCs)/DOT personnel could flag locations with aggressive driving through ranking drivers' behavioral performance in a CAV environment and drivers could be provided with proper warnings to drive more defensive manners if they are driving aggressively. Lastly, based on the findings, policymakers could develop sets of guidelines for Cooperative Automated Transportation (CAT) deployment in mixed-traffic environments (i.e., human-driven and autonomous vehicles sharing the same roads) to prevent aggressive driving and promote safety.

However, future studies could focus on more advanced clustering methods to classify driver types. In addition, future studies are essential to develop both parametric and nonparametric modeling approaches to investigate the impact of additional inclement weather events, including fog and snow on lane-changing characteristics of diverse driver groups. Finally, the methodology and findings presented in this study could be generalized to other roadway segments in conducting future research based on safety and operational perspectives.

Funding

This work was sponsored by the Wyoming Department of Transportation (WYDOT). Grant Number: RS07221.

ORCID

Anik Das http://orcid.org/0000-0003-4674-5334
Md Nasim Khan http://orcid.org/0000-0001-5996-091X
Mohamed M. Ahmed http://orcid.org/0000-0002-1921-0724
Shaun S. Wulff http://orcid.org/0000-0002-5695-4925

References

Ahmed, M. M., Das, A., Khan, M. N., Hammit, B., & Ali, E. (2021). Driver Performance and Behavior in Adverse Weather Conditions: Microsimulation and Variable Speed Limit Implementation of the SHRP2 Naturalistic Driving Study Results. Final Report WY-2105F, Wyoming Department of Transportation.

Ahmed, M. M., Ghasemzadeh, A., Eldeeb, H., Gaweesh, S., Clapp, J., Ksaibati, K., & Young, R. (2015). *Driver performance and behavior in adverse weather conditions: An investigation using the SHRP2 naturalistic driving study data—phase 1*. Publication FHWA-WY-16/08F. Wyoming Department of Transportation, vol. 307, 1–73.

Ahmed, M., Young, R., Ghasemzadeh, A., Hammit, B., Ali, E., Khan, N., Das, A., & Eldeeb, H. (2017). *Implementation of SHRP2 results within the Wyoming connected vehicle variable speed limit system: Phase 2 early findings report and phase 3 proposal*. U.S. Department of Transportation.

Ahmed, M. M., Ghasemzadeh, A., Hammit, B., Khan, N., Das, A., Ali, E., Young, R. K., & Eldeeb, H. (2018). *Driver performance and behavior in adverse weather conditions: An investigation using the SHRP2 naturalistic driving study data - phase 2*. Final Report WY-18/05F, U.S. Department of Transportation.

Ali, Y., Zheng, Z., & Haque, M. M. (2018). Connectivity's impact on mandatory lane-changing behaviour: Evidences from a driving simulator study. *Transportation Research Part C: Emerging Technologies*, *93*(June), 292–309. doi:10.1016/j.trc.2018.06.008

Ariannezhad, A., & Wu, Y.-J. (2019). Effects of heavy rainfall in different light conditions on crash severity during Arizona's monsoon season. *Journal of Transportation Safety & Security*, *11*(6), 579–594. doi:10.1080/19439962.2018.1454561

Bakhit, P. R., Osman, O. A., & Ishak, S. (2017). Detecting imminent lane change maneuvers in connected vehicle environments. *Transportation Research Record: Journal of the Transportation Research Board*, *2645*(1), 168–175. doi:10.3141/2645-18

Blissett, R. (2017). *Multilevel modeling in R*. Retrieved from https://rpubs.com/rslbliss/r_mlm_ws

Cai, X., Chang, W., Gao, L., & Zhou, C. (2021). Design and application of real-time monitoring system for service status of continuously welded turnout on the high-speed railway bridge. *Journal of Transportation Safety & Security*, *13*(7), 735–758. doi:10.1080/19439962.2019.1673858

Campbell, K. L. (2012). The SHRP 2 naturalistic driving study: Addressing driver performance and behavior in traffic safety. *Transportation Research News*, *282*, 30–35.

Center for Transportation Research and Education (CTRE). (2015). *SHRP2-roadway information database - background*. Retrieved from http://www.ctre.iastate.edu/shrp2-rid/

Chen, C., Zhao, X., Liu, H., Ren, G., & Liu, X. (2019). Influence of adverse weather on drivers' perceived risk during car following based on driving simulations. *Journal of Modern Transportation*, *27*(4), 282–292. doi:10.1007/s40534-019-00197-4

Chen, R., Gabler, H. C., & Sherony, R. (2017). *Predicting driver lane change maneuvers using vehicle kinematic data*. 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV), 1–9.

Chen, S., Chen, Y., & Xing, Y. (2022). Comparison and analysis of crash frequency and rate in cross-river tunnels using random-parameter models. *Journal of Transportation Safety & Security*, *14*(2), 280–304. doi:10.1080/19439962.2020.1779420

Chovan, J. D., Tijerina, L., Alexander, G., & Hendricks, D. L. (1994). *Examination of lane change crashes and potential IVHS countermeasures*. Final Report, National Highway Traffic Safety Administration, U.S. Department of Transportation (Issue March).

Das, A., & Ahmed, M. M. (2021). Exploring the effect of fog on lane-changing characteristics utilizing the SHRP2 naturalistic driving study data. *Journal of Transportation Safety & Security*, *13*(5), 477–502. doi:10.1080/19439962.2019.1645777

Das, A., Ahmed, M. M., & Ghasemzadeh, A. (2019). Using trajectory-level SHRP2 naturalistic driving data for investigating driver lane-keeping ability in fog: An association rules mining approach. *Accident Analysis and Prevention, 129*, 250–262.

Das, A., Ghasemzadeh, A., & Ahmed, M. M. (2019). Analyzing the effect of fog weather conditions on driver lane-keeping performance using the SHRP2 naturalistic driving study data. *Journal of Safety Research, 68*(February), 71–80.

Das, A., Khan, M. N., & Ahmed, M. M. (2020). Nonparametric Multivariate Adaptive Regression Splines Models for Investigating Lane-Changing Gap Acceptance Behavior Utilizing Strategic Highway Research Program 2 Naturalistic Driving Data. *Transportation Research Record: Journal of the Transportation Research Board, 2674*(5), 223–238.

Das, S., Dutta, A., & Sun, X. (2019). Patterns of rainy weather crashes: Applying rules mining. *Journal of Transportation Safety and Security*, 12(9), 1083–1105. doi:10.1080/19439962.2019.1572681

Das, A., Ghasemzadeh, A., & Ahmed, M. M. (2019). Analyzing the effect of fog weather conditions on driver lane-keeping performance using the SHRP2 naturalistic driving study data. *Journal of Safety Research, 68*(February), 71–80.

Deligianni, S. P., Quddus, M., Morris, A., Anvuur, A., & Reed, S. (2017). Analyzing and modeling drivers' deceleration behavior from normal driving. *Transportation Research Record: Journal of the Transportation Research Board*, *2663*(1), 134–141. doi:10.3141/2663-17

Everitt, B., & Hothorn, T. (2011). *An introduction to applied multivariate analysis with R*. New York, USA: Springer Science & Business Media.

Faria, M. V., Baptista, P. C., Farias, T. L., & Pereira, J. M. S. (2020). Assessing the impacts of driving environment on driving behavior patterns. *Transportation*, *47*(3), 1311–1337. doi:10.1007/s11116-018-9965-5

FHWA. (2022). *How do weather events impact roads?* Retrieved from https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm

Fitch, G. M., Lee, S. E., Klauer, S., Hankey, J., Sudweeks, J., & Dingus, T. (2009). *Analysis of lane-change crashes and near-crashes*. Final Report DOT HS 811 147, US Department of Transportation, National Highway Traffic Safety Administration (Issue June).

Galovski, T., & Blanchard, E. B. (2002). Psychological characteristics of aggressive drivers with and without intermittent explosive disorder. *Behaviour Research and Therapy*, *40*(10), 1157–1168. doi:10.1016/S0005-7967(01)00083-3

Gan, X., Weng, J., Li, W., & Han, M. (2020). Spatial-temporal varying coefficient model for lane-changing behavior in work zone merging areas. *Journal of Transportation Safety & Security*, 1–24. doi:10.1080/19439962.2020.1864075

Ghasemzadeh, A., & Ahmed, M. M. (2018). Utilizing naturalistic driving data for in-depth analysis of driver lane-keeping behavior in rain: Non-parametric MARS and parametric logistic regression modeling approaches. *Transportation Research Part C: Emerging Technologies*, *90*(March), 379–392. doi:10.1016/j.trc.2018.03.018

Ghasemzadeh, A., Hammit, B. E., Ahmed, M. M., & Eldeeb, H. (2019). Complementary methodologies to identify weather conditions in naturalistic driving study trips: Lessons learned from the SHRP2 naturalistic driving study & roadway information database. *Safety Science*, *119*, 21–28. doi:10.1016/j.ssci.2019.01.006

Golbabaei, F., Nejad, F. M., & Noory, A. R. (2014). A microscopic analysis of speed deviation impacts on lane-changing behavior. *Transportation Planning and Technology*, *37*(4), 391–407. doi:10.1080/03081060.2014.897128

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). UK: Wiley.

Higgs, B., & Abbas, M. (2013 *A two-step segmentation algorithm for behavioral clustering of naturalistic driving styles*. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, Itsc, 857–862. doi:10.1109/ITSC.2013.6728339

Hill, C., Elefteriadou, L., & Kondyli, A. (2015). Exploratory analysis of lane changing on freeways based on driver behavior. *Journal of Transportation Engineering*, *141*(4), 04014090. doi:10.1061/(ASCE)TE.1943-5436.0000758

Hossain, M. J., Ivan, J. N., Zhao, S., Wang, K., Sharmin, S., Ravishanker, N., & Jackson, E. (2022). Considering demographics of other involved drivers in predicting the highest driver injury severity in multi-vehicle crashes on rural two-lane roads in California. *Journal of Transportation Safety & Security*, 1–16. doi:10.1080/19439962.2022.2033899

InSight SHRP2 NDS. (2019). *SHRP2 NDS data access*. Retrieved from https://insight.shrp2nds.us/

Islam, M., & Mannering, F. (2020). A temporal analysis of driver-injury severities in crashes involving aggressive and non-aggressive driving. *Analytic Methods in Accident Research*, *27*, 100128. doi:10.1016/j.amar.2020.100128

Jackson, T. L., & Sharif, H. O. (2016). Rainfall impacts on traffic safety: Rain-related fatal crashes in Texas. *Geomatics, Natural Hazards and Risk*, *7*(2), 843–860. doi:10.1080/19475705.2014.984246

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis* (Vol. 344). Hoboken, New Jersey, USA: John Wiley & Sons.

Khan, M. N., Ghasemzadeh, A., & Ahmed, M. M. (2018). Investigating the impact of fog on freeway speed selection using the shrp2 naturalistic driving study data. *Transportation Research Record: Journal of the Transportation Research Board, 2672*(16), 93–104.

Khan, M. N., Das, A., & Ahmed, M. M. (2020). Non-parametric association rules mining and parametric ordinal logistic regression for an in-depth investigation of driver speed selection behavior in adverse weather using SHRP2 naturalistic driving study data. *Transportation Research Record: Journal of the Transportation Research Board, 2674*(11), 101–119.

Khattak, Z. H., Fontaine, M. D., Li, W., Khattak, A. J., & Karnowski, T. (2021). Investigating the relation between instantaneous driving decisions and safety critical events in naturalistic driving environment. *Accident; Analysis and Prevention*, *156*(September 2020), 106086. doi:10.1016/j.aap.2021.106086

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied linear regression models* (4th ed.). McGraw-Hill, Chicago, IL, USA.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2019). *Package 'lmerTest'*.

Lee, S. E., Olsen, E. C. B., & Wierwille, W. W. (2004). *A comprehensive examination of naturalistic lane-changes*. Final Report DOT Hs, 809(702), March, 1–236.

Li, X.-G., Jia, B., Gao, Z.-Y., & Jiang, R. (2006). A realistic two-lane cellular automata traffic model considering aggressive lane-changing behavior of fast vehicle. *Physica A: Statistical Mechanics and Its Applications*, *367*, 479–486. doi:10.1016/j.physa.2005.11.016

Liu, B. S., & Lee, Y. H. (2005). Effects of car-phone use and aggressive disposition during critical driving maneuvers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(4–5), 369–382. doi:10.1016/j.trf.2005.04.019

MathWorks. (2021). *Fillmissing*. Retrieved from https://www.mathworks.com/help/matlab/ref/fillmissing.html

Mohammadnazar, A., Arvin, R., & Khattak, A. J. (2021). Classifying travelers' driving style using basic safety messages generated by connected vehicles: Application of unsupervised machine learning. *Transportation Research Part C: Emerging Technologies*, *122*(June 2020), 102917. doi:10.1016/j.trc.2020.102917

Moridpour, S., Sarvi, M., & Rose, G. (2010). Modeling the lane-changing execution of multiclass vehicles under heavy traffic conditions. *Transportation Research Record: Journal of the Transportation Research Board*, *2161*(1), 11–19. doi:10.3141/2161-02

NHTSA. (2019). *Traffic safety facts: A compilation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system*. U.S. Department of Transportation.

Park, H. S., & Jun, C. H. (2009). A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications*, *36*(2), 3336–3341. doi:10.1016/j.eswa.2008.01.039

Qi, X., Wu, G., Boriboonsomsin, K., & Barth, M. J. (2016). Empirical study of lane-changing characteristics on high-occupancy-vehicle facilities with different types of access control based on aerial survey data. *Journal of Transportation Engineering*, *142*(1), 04015034. doi:10.1061/(ASCE)TE.1943-5436.0000803

Salvucci, D. D., & Liu, A. (2002). The time course of a lane change: Driver control and eye-movement behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*(2), 123–132. doi:10.1016/S1369-8478(02)00011-6

Shangguan, Q., Wang, J., Fu, T., & Fang, S. (2021). Quantification of cut-in risk and analysis of its influencing factors: A study using random parameters ordered probit model. *Journal of Transportation Safety & Security*, 1–26. doi:10.1080/19439962.2021.1994683

Shao, H., Zhang, M., Feng, T., & Dong, Y. (2020). A discretionary lane-changing decision-making mechanism incorporating drivers' heterogeneity: A signalling game-based approach. *Journal of Advanced Transportation*, *2020*, 1–16. doi:10.1155/2020/8892693

Sun, D., & Elefteriadou, L. (2010). Research and implementation of lane-changing model based on driver behavior. *Transportation Research Record: Journal of the Transportation Research Board*, *2161*(1), 1–10. doi:10.3141/2161-01

Sun, D., & Elefteriadou, L. (2011). Lane-changing behavior on urban streets: A focus group-based study. *Applied Ergonomics*, *42*(5), 682–691. doi:10.1016/j.apergo.2010.11.001

Toledo, T., & Zohar, D. (2007). Modeling duration of lane changes. *Transportation Research Record: Journal of the Transportation Research Board*, *1999*(1), 71–78. doi:10.3141/1999-08

Wang, C., Li, Z., Fu, R., Zhang, M., & Sun, Q. (2019). Lane change safety assessment of coaches in naturalistic driving state. *Safety Science*, *119*(December 2017), 126–132. doi:10.1016/j.ssci.2018.09.009

Wang, J., Dang, R., Zhang, F., Wang, J., Yi, S., & Li, K. (2013). *Analysis of Chinese driver's lane change characteristic based on real vehicle tests in highway*. Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), 100084(October), 1917–1922. doi:10.1109/ITSC.2013.6728509

Wang, X., Yang, M., & Hurwitz, D. (2019). Analysis of cut-in behavior based on naturalistic driving data. *Accident; Analysis and Prevention*, *124*(March), 127–137. doi:10.1016/j.aap.2019.01.006

Wu, L., Dadashova, B., Geedipally, S., Pratt, M. P., & Shirazi, M. (2021). Using naturalistic driving study data to explore the association between horizontal curve safety and operation on rural two-lane highways. *Journal of Transportation Safety & Security*, *13*(8), 896–913. doi:10.1080/19439962.2019.1697776

Yang, Q., & Koutsopoulos, H. N. (1996). A microscopic traffic simulator for evaluation of dynamic traffic management systems. *Transportation Research Part C: Emerging Technologies*, *4*(3), 113–129. doi:10.1016/S0968-090X(96)00006-X

You, F., Zhang, R., Lie, G., Wang, H., Wen, H., & Xu, J. (2015). Trajectory planning and tracking control for autonomous lane change maneuver based on the cooperative vehicle infrastructure system. *Expert Systems with Applications*, *42*(14), 5932–5946. doi:10.1016/j.eswa.2015.03.022

Zhang, W., Wang, C., Chen, Q., Liu, J., Feng, Z., Wang, K., & Shen, Y. (2022). Lane-change behavior in low illumination: Research based on a questionnaire investigation. *Journal of Transportation Safety & Security*, *14*(1), 130–151. doi:10.1080/19439962.2020.1744050

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: Emerging Technologies*, *26*, 367–379. doi:10.1016/j.trc.2012.10.007

Zhou, Y., Fu, R., & Wang, C. (2020). Learning the car-following behavior of drivers using maximum entropy deep inverse reinforcement learning. *Journal of Advanced Transportation*, *2020*, 1–13. doi:10.1155/2020/4752651