TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
1–16

DOI: 10.1177/0361198120914293
journals.sagepub.com/home/trr
SAGE

Nonparametric Multivariate Adaptive Regression Splines Models for Investigating Lane-Changing Gap Acceptance Behavior Utilizing Strategic Highway Research Program 2 Naturalistic Driving Data

Anik Das[1], Md Nasim Khan[1], and Mohamed M. Ahmed[1,2]

Abstract
Gap acceptance is one of the crucial components of lane-changing analysis and an important parameter in microsimulation modeling. Drivers' poor gap judgment, and failure to accept a necessary safety gap, make it one of the major causes of lane-changing crashes on roadways. Several studies have been conducted to investigate lane-changing gap acceptance behavior; however, very few studies examined the behavior in complex real-world situations, such as in naturalistic settings. This study examined lane-changing gap acceptance behavior from the big Strategic Highway Research Program 2 (SHRP2) Naturalistic Driving Study (NDS) datasets using a nonparametric multivariate adaptive regression splines (MARS) approach to better understand the complex effects of different factors in gap acceptance behavior. The study developed a unique methodology to identify lane-changing events of the non-NDS-vehicles using the front-mounted radar data from NDS vehicles and extract necessary parameters for analyzing gap acceptance behavior. In addition, surrogate measures of safety, that is, time-to-collision (TTC), was utilized to understand the impact of lane-changing on the NDS following vehicle safety. Moreover, different distributions of gap acceptance were fitted to identify the trend of gap acceptance behavior. The results from the MARS model revealed that different factors including relative speed between lane-changing vehicle (LCV) and lead vehicle (LV)/following vehicle (FV), traffic conditions, acceleration of LCV and FV, and roadway geometric characteristics have significant effects on gap acceptance behavior. The results of this study have significant implications, which could be used in microsimulation model calibration and safety improvements in connected and autonomous vehicles (CAV).

Driver lane-changing maneuvers are one of the most essential maneuvers on the roadways and have a substantial impact on roadway safety. The data of the National Highway Traffic Safety Administration (NHTSA) estimates that lane-changing maneuver was responsible for about 451,000 motor vehicle crashes in the U.S.A. in 2015 (*1*). According to You et al., lane-changing maneuvers accounted for roughly 13,939 traffic accidents and the death of 24,565 people in the U.S.A. from 1994 to 2005 (*2*). In another study, Hou et al. implied that improper lane-changing and merge maneuvers are responsible for approximately 5% of all crashes and 7% of all crash fatalities in the U.S.A. (*3*).

A driver has to consider several factors while changing a lane including speed, position of the vehicle, and vehicle(s) in the target lane, as well as different vehicular characteristics, geographical characteristics of the roadway, and other factors including weather and traffic characteristics (*3*). To execute a lane-changing maneuver, drivers assess the adjacent gap in the target lane, that is, they evaluate the lead and lag gaps. The lane-changing decision is determined based on the availability of a safe gap in the target lane. A driver's inaccurate gap judgment and failure to accept a necessary safety gap after

[1]Department of Civil & Architectural Engineering, University of Wyoming, Laramie, WY
[2]Federal Highway Administration, U.S. Department of Transportation, Turner-Fairbank Highway Research Center, McLean, VA

Corresponding Author:
Mohamed M. Ahmed, mahmed@uwyo.edu

initiating a lane-changing maneuver might introduce high-risk driving maneuvers and could eventually result in a lane-changing crash (*3*). Therefore, gap acceptance is one of the critical elements of lane-changing analysis.

Different approaches for analyzing lane-changing gap acceptance can be found in the literature. Ahmed proposed a systematic approach for modeling lane-changing behavior using a discrete choice framework, where a gap acceptance model was utilized to characterize the execution of lane-changing maneuvers. Ahmed concluded that several important factors including gap length, relative speed, distance remaining to the point at which lane change must be completed, and so forth, affect drivers' gap acceptance behavior (*4*). Toledo et al. modeled lane-changing gap acceptance as a binary choice problem by comparing the available space gaps with the critical gaps (minimum acceptable gap). They assumed that an available gap (i.e., lead and lag gap) was acceptable if it was greater than the critical gap (*5*). Hill and Elefteriadou analyzed the lane-changing behavior on freeways by collecting data on drivers' desire speed, lane-changing duration, and gap acceptance, and found that drivers were more likely to accept smaller lag gaps during congested traffic conditions (*6*). Lee et al. developed a probability model for discretionary lane-changing maneuver (i.e., a maneuver that is intended to improve the perceived driving condition) in highways and found that both relative velocity and relative lead gap are the main criteria for discretionary lane-changing maneuvers and have similar positive influences on the choice probability model (*7*). In another study, Wang et al. developed multilevel mixed-effects linear models to examine the influencing factors of lane-changing gap acceptance and found that acceptance of lead and lag gaps were significantly affected by several factors including environmental variables, vehicle type, and kinematic parameters (*8*).

In recent years, the use of numerous nonparametric techniques has increased because of their advantages over the traditional parametric techniques in investigating driver behavior (*9–11*). Multivariate adaptive regression splines (MARS) is becoming one of the most popular nonparametric approaches in transportation fields. However, researchers mostly used this technique to develop crash prediction models and to identify crash-contributing factors. Haleem et al. utilized MARS model to develop crash modification factors for urban freeway interchange influence areas (*12*). Park and Abdel-Aty assessed the safety effects of multiple roadside treatments by estimating crash modification factors (CMF) using MARS model (*13*). In another study, MARS model was utilized to predict rear-end crashes at unsignalized intersections (*14*). A study focused on the analysis of freeway accidents applied MARS model to explore the effects of non-behavioral factors including roadway geometric characteristics, traffic factors, and environmental conditions on the frequency of freeway accidents (*15*). Another study conducted by Gaweesh et al. applied MARS model to develop crash prediction models for a case study of Wyoming Interstate 80 (*16*). Apart from the crash analysis perspective, MARS model has been recently utilized in other transportation fields (e.g., traffic flow prediction, vehicular emission prediction, fuel consumption estimation, etc.) to model complex relationships and interactions (*17–19*).

Even though many studies utilized MARS model for safety analysis and other fields, there is a lack of studies that utilized MARS model to analyze complex driver behavioral performance, especially lane-changing gap acceptance behavior. In addition, very few studies examined the behavior comprehensively using naturalistic data. MARS has the ability to model complex relationships and interactions. At the same time, it can provide interpretable models (*16*). Additionally, MARS can describe and evaluate causal links between different factors (*20*). Therefore, providing a better explanatory predictive model fit for the naturalistic data using MARS would help understanding the different contributing factors affecting lane-changing gap acceptance behavior. The main objective of this study is to investigate lane-changing gap acceptance behavior from the big Strategic Highway Research Program 2 (SHRP2) Naturalistic Driving Study (NDS) and Roadway Information Database (RID) datasets using MARS approach. This will be accomplished by conducting the following key research tasks:

- Developing an effective and comprehensive methodology using the NDS front-mounted radar data to identify lane-changing events of the vehicles adjacent to the NDS vehicle's lane (i.e., the subject/ego vehicle is not the NDS vehicle in this study) as well as extracting necessary parameters for analyzing gap acceptance behavior.
- Identifying the key contributing factors that affect lane-changing gap acceptance behavior using the nonparametric MARS approach.

Data Acquisition and Preparation

The study utilized a subset of the SHRP2 NDS data that were acquired from the Virginia Tech Transportation Institute (VTTI). The SHRP2 collected the first large-scale and most comprehensive NDS data, where more than 3,400 drivers participated from six U.S. states between 2010 and 2013 (*21*). A total of 2 petabytes of NDS data were collected from in-vehicle data acquisition systems (DAS). The DAS were capable of collecting various vehicle kinematics from CAN-bus and additional

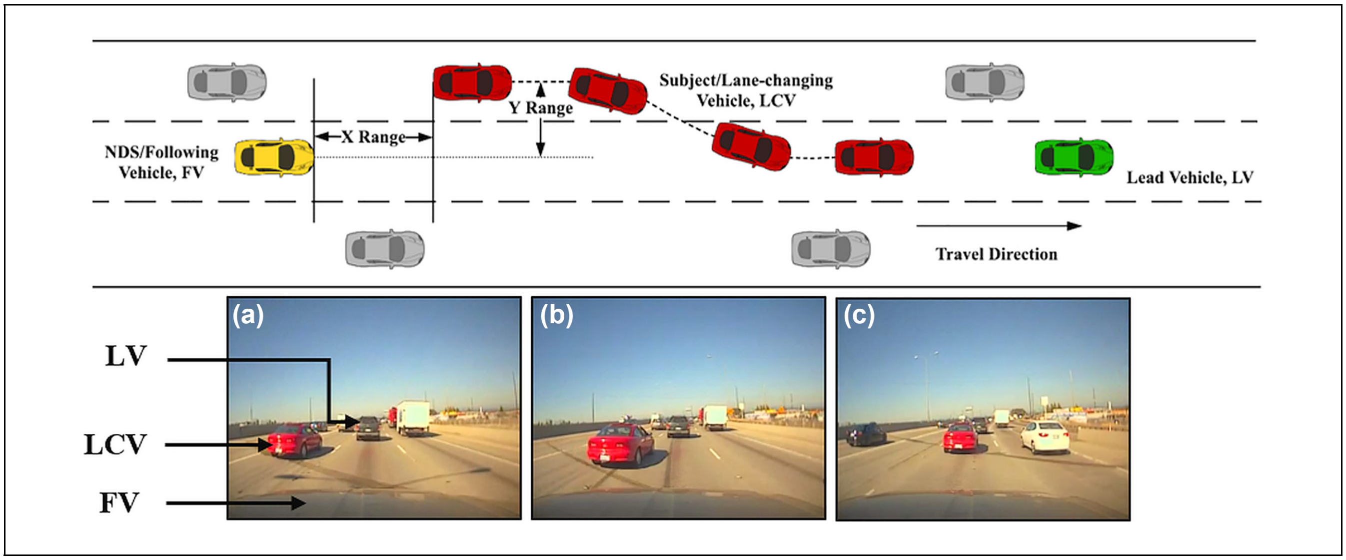

Figure 1. Typical lane-changing event considered in this study: (*a*) Naturalistic Driving Study radar detected a lane-changing vehicle coming from the left of the Naturalisic Driving Study vehicle's lane, (*b*) lane-changing vehicle started changing lane, and (*c*) lane-changing vehicle completed the lane-changing maneuver and occupied the Naturalistic Driving Study vehicle's lane.

data using its forward radar, video views of the front and rear roadways, accelerometers, GPS, and onboard computer vision systems (*22*). Additionally, another complementary Roadway Information Database (RID) was utilized that contains dataset from the SHRP2 mobile data collection project and other existing dataset from public and private sources (*23*). Note that these unprecedented datasets have been used in many studies to investigate driver behavior in addition to traffic safety and operations (*24–30*).

To effectively extract NDS trips that occurred in different weather conditions, two innovative methodologies were developed. In the first method, weather data from the National Climatic Data Center (NCDC) database were utilized to identify different weather events. In the second method, weather-related crash locations were considered to detect weather conditions. More information regarding the data acquisition procedure can be found in (*31*, *32*). Using the data acquisition methodologies, many NDS trips in different weather conditions were effectively and systematically identified and acquired for the first time.

The front-mounted radar data were processed by VTTI to improve the usability of the collected data (*33*). DAS-equipped NDS vehicles had only radar data from a single front-mounted radar unit, which can track up to eight vehicles in front of the NDS ego-vehicle and can provide information regarding distance to lead vehicle(s) (LVs) and relative speeds between the NDS ego-vehicle and targets (i.e., other LVs). More details regarding SHRP2 NDS radar data can be found in (*33*). Considering that front-mounted radar cannot detect the presence of a lag vehicle (i.e., the vehicle behind the NDS vehicle's lane), it is not possible to consider the NDS vehicle as a subject vehicle for gap acceptance analysis. Therefore, to investigate the lane-changing gap acceptance behavior, the vehicles adjacent to the NDS vehicle's lane were considered as the lane-changing vehicle (LCV) in this study, where the NDS vehicle served as a following vehicle (FV) and provided the opportunity to analyze gap acceptance behavior of the vehicle (i.e., LCV) in front of it, as shown in Figure 1. Figure 1 demonstrates a lane-changing event in a trip performed by an LCV in clear weather conditions.

It is worth mentioning that the authors had only access to data of LCV recorded by the NDS vehicle's front radar. In fact, gap acceptance analysis requires a lot of data processing for the front-mounted NDS radar data since these data are not readily available in the SHRP2 NDS data. Additionally, working with radar data is not straightforward because of its noisy nature and the limited distinction between detected object categories. Moreover, several NDS trips might exist with missing radar data or other erroneous values. Therefore, an effective algorithm was needed to be developed for processing radar data efficiently. This study developed an automatic algorithm to identify the lane-changing event and corresponding parameters for analyzing gap acceptance behavior. To develop the algorithm, the following steps were considered:

1. Some received NDS trips had missing radar data. NDS trips with missing radar data were imputed using *fillmissing* function of MATLAB, where missing data were imputed through linear

interpolation of non-missing neighboring values (*34*). It is worth mentioning that no imputations were made if more than 10% of data were missing, and those data were discarded from the algorithm (*35*, *36*).

2. Afterward, radar data were smoothed with *smoothdata* function of MATLAB using "movemedian" method. The method helped to reduce the periodic trends in the data because of some outliers (*37*).
3. The presence of potential LCV on the nearest lane was determined considering the following criteria. The algorithm proceeded further only if the first criterion was met.
   - If $|Y_{lcv}|>|Y_l|$, potential LCV is on the nearest lane with respect to the NDS vehicle
   - If $|Y_{lcv}|<|Y_l|$, potential LCV is on the NDS vehicle's lane

   where, Y_{lcv} = left/right lateral distance between potential LCV and FV
   Y_l = Distance from vehicle centerline to inside of left/right side lane marker based on a vehicle-based machine vision technique (*38*)

4. To ensure FV (i.e., NDS vehicle) did not move in a lateral direction, that is, change a lane, corresponding events were eliminated. The lane-changing maneuver of FV was identified using lane position offset variable from the NDS data. A threshold of ±100 cm lateral shift (i.e., left and right) in the position of an FV was considered as a lane-changing maneuver (*10*). Figure 2 shows a sample of lane-changing maneuver (left to right) with lane position offset value above 100 cm, and the peaks in the figure represent the occurrence of lane-changing maneuver of the FV. As can be seen in Figure 2, the value of lane position offset started to increase, indicating that FV started to move laterally from left to right of the lane center until the value reached a maximum point. A jump was then occurred representing the FV reaching the far right of the driver's adjacent lane.
5. The presence of LV on the NDS vehicle's lane was ensured considering the following criteria.
   - If $|Y_{lv}|<|Y_l|$, LV is on the NDS vehicle's lane
   - If $|Y_{lv}|>|Y_l|$, LV is not on the NDS vehicle's lane

   where, Y_{lv} = left/right lateral distance between LV and FV
   These criteria were checked for all possible Y ranges. The velocity of LCV and FV (more than 1 m/s) were checked in this step to ensure that the two vehicles are in motion (*8*). However, FVs were included in all lane changes. It is worth noting that to investigate gap acceptance behavior using the SHRP2 NDS data, the FV has been always an NDS vehicle in the target lane.

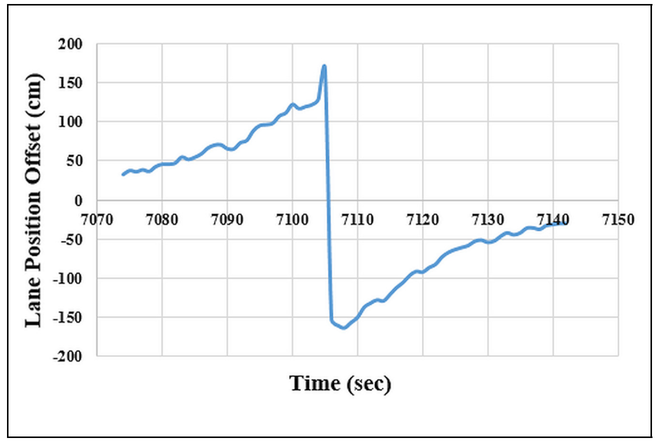

Figure 2. Demonstration of automated identification process of following vehicle's lane-changing maneuver using lane position offset.

6. The local maximum (peak) of the Y range of potential LCV (i.e., identified in Step 3) was determined using *findpeaks* function of MATLAB. The peak is defined as the starting point of a lane-changing event.
7. Every local peak of the Y range was checked up to 20 s period. Based on the literature, the maximum duration of lane-changing maneuver can vary up to 15 s. Therefore, 20 s was conservatively selected to capture all the lane-changing events. Additionally, a gradual reduction of Y range was considered to identify an end of a lane-changing event. The end of the lane-changing event was determined when the gradual reduction was stopped and Y range was stabilized within ±4 cm. As an example, Figure 3 exhibits the Y range of a lane-changing scenario. The starting point of the lane-changing event is shown in A in the figure. Once the lane-changing event was initiated, the LCV's Y range started to decrease and approached zero. When the LCV's Y range stabilized within ±4 cm, it was considered the completion of lane-changing event. This point is the end of lane-changing and denoted as B in Figure 3. The time required for the LCV to travel from A to B was defined as the lane-changing duration.
8. Once the lane-changing event was identified, lead and lag gaps were calculated. This study defined distance in terms of time rather than space, as time represents better driver behavior compared

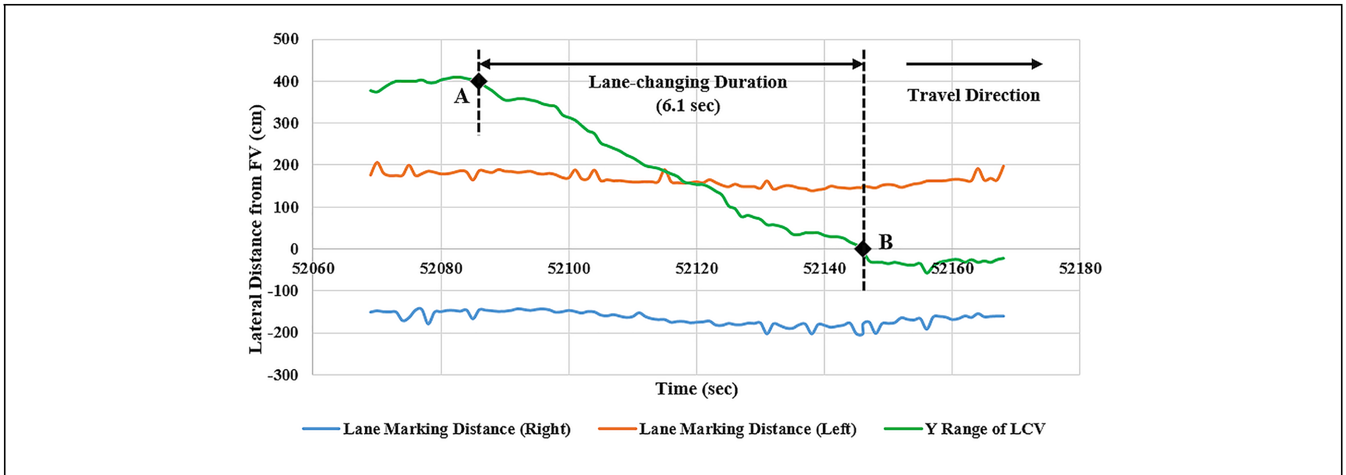

Figure 3. *Y* ranges of a lane-changing event.

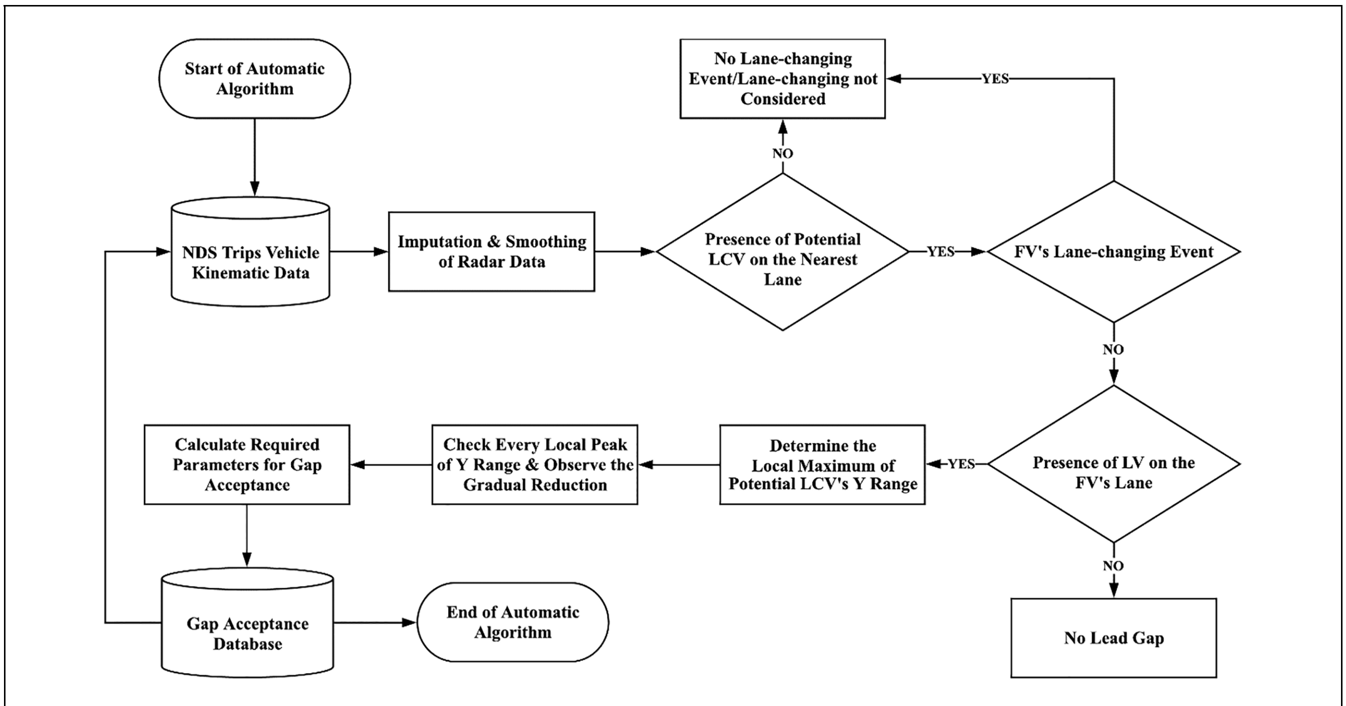

Figure 4. Automatic algorithm for identifying lane-changing events using radar data.

with space. The LCV is concerned with sufficient time for a safe lane-changing maneuver with corresponding travel speed. Therefore, time gaps provide better representation than distance gaps (*39*). The gaps were calculated when the LCV started its movement laterally from the current lane to the desired lane. Lead gap was denoted as the time taken to traverse the longitudinal distance between LV and LCV. On the contrary, lag gap represented the time taken to traverse the longitudinal distance between FV and LCV (*8*).

9. The identification algorithm (Steps 1 to 8) was repeated for the entire received NDS trips automatically in MATLAB environment. In addition, a database was developed after extracting the necessary parameters for gap acceptance. The overall methodology used for automatic identification of lane-changing events has been shown in Figure 4.

The developed algorithm detected 7,031 lane-changing events that correspond to 1,192 NDS trips. As mentioned earlier, front-mounted NDS radar can track up to eight forward vehicles in front of the NDS ego-vehicle; however, the radar cannot continuously track the same vehicles. Radar marks or tags the vehicle number and this number assignment would change if the vehicle disappears/reappears behind traffic or because of roadway curvature. The new tag could be for the same vehicle or a new vehicle. In addition, the lane-changing identification algorithm was developed in a conservative way so that no lane-changing events would be missed. Because of these issues, additional manual verifications were needed. In this study, 599 lane-changing events corresponding to 361 trips were randomly selected, and manually verified and reduced using the Wyoming NDS Visualization and Reduction Tool (*31*, *40*). This step involved manual video observation and annotation of environmental conditions (i.e., weather, surface, and visibility), traffic states (free-flow and mixed-flow), and lane-changing maneuver types (i.e., mandatory and discretionary) for each extracted lane-changing event within a single trip. Note that weather conditions were categorized into clear and adverse (i.e., snow, rain, and fog). In addition, traffic states were classified into free-flow (level of service A and B) and mixed-flow (level of service C to F) following previous studies (*27*, *28*, *41*). The lane-change maneuver types were manually coded considering six categories of maneuvers. The video observation allowed the identification of maneuvers such as passing a slower LV, return to driver's preferred lane, enter the freeway, exit/prepare to exit from the freeway, addition/drop of a lane, among other categories (e.g., merging the freeway, change in visibility condition, etc.) based on a previous study (*11*). Afterward, roadway characteristics provided in the RID database were linked with each lane-changing event to create a final modeling dataset. The overall process from data preparation to analysis is illustrated in Figure 5.

Methodology

Nonparametric techniques have been used in recent years because of their advantages over parametric models in many aspects. The main advantages of using nonparametric techniques include the capability of providing high prediction accuracy, no predetermined assumption related to data distribution, and ability to handle many explanatory variables at the same time (*42*). All of these advantages could be useful for assessing complex driver behavioral performance, such as gap acceptance behavior. This study utilized MARS, a nonparametric modeling technique, to investigate the factors affecting driver gap acceptance behavior. The key advantages of the MARS model are that it can capture nonlinearity in the data; therefore, it can be used to model complex relationships among variables. In addition, MARS can provide an interpretable and transparent model. Moreover, MARS has the ability to handle multicollinearity between variables (*10*). The focus of this paper is to demonstrate how nonparametric MARS approach can be used to model nonlinear relationships and interactions and to provide better insights about the contributing factors, which may affect driver gap acceptance behavior.

MARS is a nonparametric piecewise multivariate regression technique introduced by Friedman (*43*). In the MARS model, the space of independent variables is split into several regions separated by knots. Then it fits a spline function that consists of several polynomial basis functions (BFs) between these knots smoothly. The general form of the MARS model is shown in Equation 1 (*10*, *13*, *14*, *43*).

$$\hat{y} = b_0 + \sum_{m=1}^{M} b_m + \mathrm{BF}_m(x) \tag{1}$$

where, $\hat{y}$ is the predicted response variable (i.e., lead and lag gap in this study), b_0 is the coefficient of constant BF, b_m is the coefficient of the basis function number (m), $\mathrm{BF}_m(x)$ is the basis function number (m) for the independent variable x, and M is the total number of BFs.

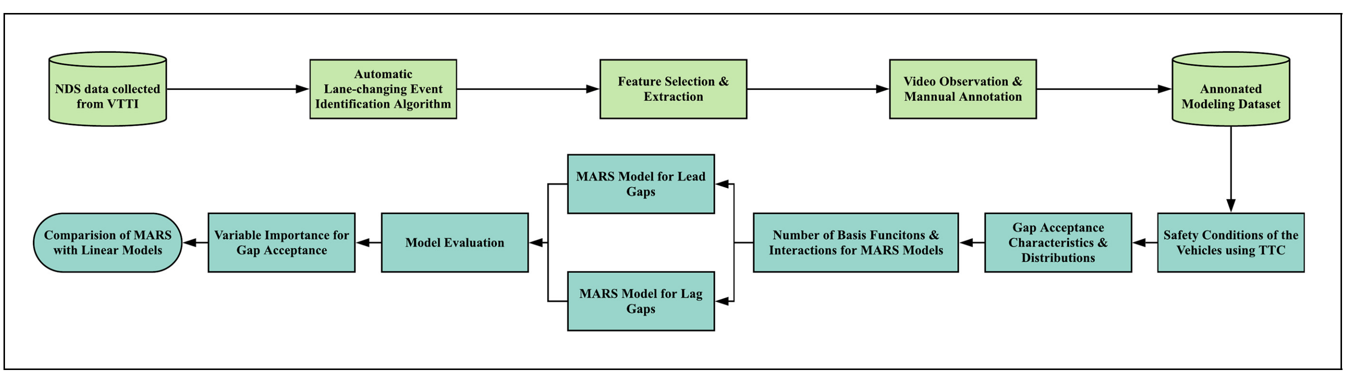

Figure 5. Flow chart summarizing the data preparation and analysis process.

There are two main phases to fit a MARS model. The first phase is the model generation, which is composed of two stages, forward-stepwise regression selection technique, and the backward-stepwise elimination process. The second phase is the model selection. In the forward-stepwise selection stage, an initial model starts with a constant only. BFs are developed using Equation 1 and added to the model in several regions of the independent variables. Afterward, the model searches for significant variable-knot combination, and the model improvement is assessed. The interactions are also introduced in this stage to improve the model fit. The search process is repeated until the best variable-knot combination is identified (*10*, *12*, *13*, *20*).

In the backward backward-stepwise elimination stage, the model identifies a BF with the lowest contribution and eliminate based on the residual sum of squares (RSS) criteria, as provided in Equation 2 (*16*). After refitting the model, the model again identifies a BF to eliminate using the same criteria. The elimination process is repeated until all the BFs have been removed. The final results of the elimination process is a different series of candidate models (*20*).

$$\text{RSS} = \sum\nolimits_{i=1}^{n} (y_i - \hat{y})^2 \tag{2}$$

The second and final phase to fit a MARS model is the model selection phase. The selection of final model is based on the generalized-cross-validation (GCV) criterion or closeness between the training mean square error (MSE) and the test MSE. It is worth mentioning that a penalty for the model complexity is applied for the GCV criterion. The penalty is based on the degrees of freedom charged per each developed knot (*14*, *16*).

$$\text{GCV}(M) = \frac{1}{n} \frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{\left(\frac{1 - C(M)}{n}\right)^2} \tag{3}$$

$$C(M) = M + dM \tag{4}$$

where n is the number of lane-changing events, y_i is the lead/lag gaps for observation i, $C(M)$ is the complexity penalty function, and d is the defined cost for each BF optimization.

Data Analysis

Safety Conditions of the Vehicles

Safety conditions of the vehicles can be investigated through surrogate measures of safety (SMoS). Several SMoS including time-to-collision (TTC), lateral and longitudinal acceleration, and so forth, can be used. Since NDS vehicle was used as a probe vehicle and lane-changing maneuvers of surrounding vehicles (i.e., vehicles in the vicinity of the NDS vehicle's lane) were considered using radar data, accurate measurement of lateral and longitudinal acceleration is computationally expensive. Therefore, this study used TTC as the sole SMoS. TTC is necessary to compute the risk of rear-end collisions and it is extensively used to investigate safety (*44*). It is worth mentioning that there could be three possible TTCs in any lane-changing scenario; 1) TTC of the LCV with the LV in the original lane, 2) TTC of the LCV with the LV in the target lane, and 3) TTC of the FV (i.e., NDS vehicle) with the LCV in the target lane. Note that among the different crash types, rear-end collisions are the most common and hazardous in cut-in maneuvers. Therefore, the third TTC was used to measure the FV response to better understand the impact of lane-changing/cut-in maneuvers on the FV. TTC was defined as in Equation 5.

$$\text{TTC} = \frac{X \text{ Range}}{\text{Relative Speed}} \tag{5}$$

where, X Range is the longitudinal distance between the LCV and FV in target lane and Relative Speed is the speed difference between LCV and FV. Following previous studies, the study calculated TTC to classify the urgency of lane-changing events on a 4-point scale; 1 = non-urgent (TTC $>$ 5.5 s), 2 = urgent (5.5 s $\geq$ TTC $>$ 3 s), 3 = forced (3 s $\geq$ TTC $>$ 1 sec), and critical/near crash (TTC $\leq$ 1 s) (*8*, *44*, *45*). Table 1 demonstrates the results of the TTC classification of lane-changing urgency under different weather and traffic conditions.

As shown in Table 1, the percentages of non-urgent lane-changing events were found to be higher compared with other urgency levels under all weather and traffic conditions. This could be because FV drivers were more cautious by perceiving the LCV's intention of lane-changing maneuvers (*44*). As expected, all urgency levels except 1 (i.e., non-urgent) had higher percentages in mixed-flow than those in free-flow. The findings suggested that LCV drivers were not restricted by surrounding vehicles under free-flow traffic, resulting in higher non-urgent lane-changing maneuvers. In addition, the percentages of non-urgent lane-changing events in all traffic conditions under adverse weather had higher percentages than in clear weather, indicating that drivers remain more cautious while performing lane-changing maneuvers in adverse weather conditions to compensate for the increased risk (*46*). This evidence of human drivers' characteristics to adopt safe lane-changing maneuvers in adverse weather could provide guidance to autonomous vehicles (AV) so that they can learn and perform similar responses to mimic human behavior.

Gap Acceptance Characteristics

This study investigated the characteristics of lead and lag gaps in freeways to determine different features and

Table 1. Lane-Changing Urgency Based on Following Vehicle (FV) Time-to-Collision (TTC) for Different Weather and Traffic Conditions

Weather conditions	Traffic conditions	TTC range	Percentage (%)	Mean (s)	SD (s)	Min. (s)	Max. (s)
Clear	Free-flow	Non-urgent	74.24	47.08	128.82	5.53	645.91
		Urgent	3.94	3.87	0.54	3.04	5.09
		Forced	14.55	1.70	0.53	1.03	2.98
		Critical/near crash	7.27	0.33	0.21	0.07	0.81
	Mixed-flow	Non-urgent	59.09	28.11	22.71	5.72	252.99
		Urgent	6.06	3.93	0.68	3.02	5.46
		Forced	18.18	1.94	0.54	1.14	2.76
		Critical/near crash	16.67	0.52	0.29	0.09	0.97
Adverse	Free-flow	Non-urgent	86.36	24.40	15.23	5.60	42.91
		Urgent	1.14	5.10	0.00	5.10	5.10
		Forced	6.82	1.77	0.53	1.26	2.82
		Critical/near crash	5.68	0.55	0.25	0.18	0.77
	Mixed-flow	Non-urgent	65.31	40.46	48.30	6.06	108.77
		Urgent	14.29	4.39	0.59	3.76	5.47
		Forced	8.16	1.93	0.28	1.56	2.30
		Critical/near crash	12.24	0.63	0.19	0.32	0.96

Note: SD = standard deviation; Min. = minimum; Max. = maximum.

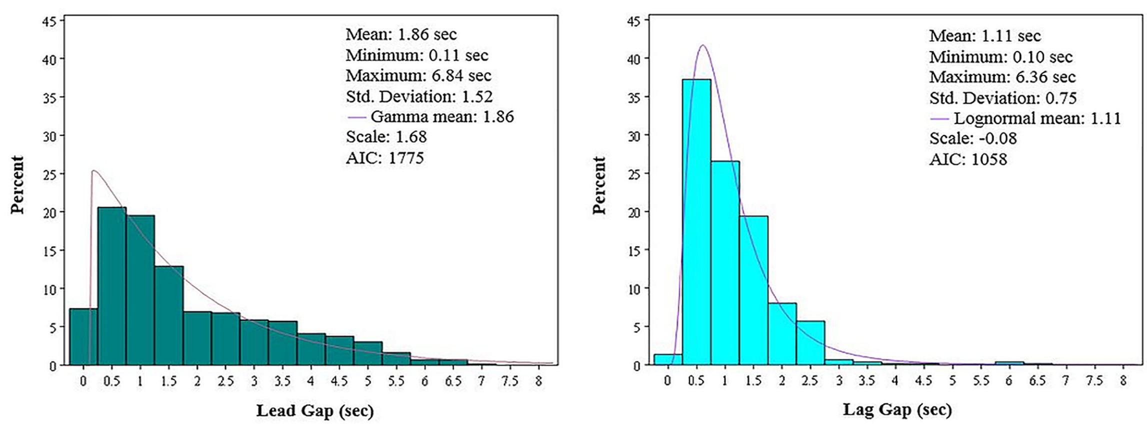

Figure 6. Fitted distribution of lead and lag gaps.

trends of gap acceptance behavior. Although the lead and lag gaps are assumed to follow a lognormal distribution in the literature, naturalistic behavior might change the assumption (*47*). Different distributions of gap acceptance including, lognormal, gamma, Weibull, exponential, and so forth, were examined. The distributions were compared using Akaike information criterion (AIC) for goodness of fit. From the considered lane-changing event dataset, it was revealed that gamma distribution fitted the best for lead gaps. On the contrary, lag gaps followed a lognormal distribution. Figure 6 demonstrates the distributions of lead and lag gaps.

As can be seen in Figure 6, the range of the lead gaps varied between 0.11 and 6.84 s, with a mean of 1.86 s. In contrast, the range of lag gaps varied from 0.10 to 6.36 s. It is worth noting that the corresponding statistics of lead gaps (i.e., mean and maximum) are relatively higher than those of lag gaps. This suggested that LCV drivers accepted shorter headways in front of FV, whereas they maintained larger headway behind the LV. This might be because drivers were more dependent on mirrors while determining lag gap. Therefore, their perception of lag gap might not be reliable in comparison with the lead gap (*44*).

Modeling Gap Acceptance Behavior

As mentioned earlier, 599 lane-changing events were considered for modeling gap acceptance behavior using MARS model. The dependent variables in the models are lead and lag gaps, which were extracted from the developed automatic lane-changing event identification

Table 2. Variables Descriptions for Gap Acceptance Models

Variable	Description	Type	Source
Response variables (unit)			
Lead gap (s)	Time taken to traverse the longitudinal distance between LV and LCV	Continuous	Developed gap acceptance database
Lag gap (s)	Time taken to traverse the longitudinal distance between FV and LCV	Continuous	Developed gap acceptance database
Explanatory variables (unit)			
Traffic flow parameters			
Minimum RH (m)	Minimum distance between LCV and FV that was available after a lane-changing event	Continuous	Developed gap acceptance database
FV acc. (g)	FV's acceleration during lane-changing event	Continuous	Developed gap acceptance database
LCV sp. (km/h)	LCV's speed during lane-changing event	Continuous	Developed gap acceptance database
LCV acc. (g)	LCV's acceleration during lane-changing event	Continuous	Developed gap acceptance database
LV sp. (km/h)	LV's speed during lane-changing event	Continuous	Developed gap acceptance database
LV acc. (g)	LV's acceleration during lane-changing event	Continuous	Developed gap acceptance database
RS ($V_{LCV} - V_{LV}$) (km/h)	Relative speed between LCV and LV during lane-changing event	Continuous	Developed gap acceptance database
RS ($V_{LCV} - V_{FV}$) (km/h)	Relative speed between LCV and FV during lane-changing event	Continuous	Developed gap acceptance database
TC	Traffic conditions during lane-changing event	Categorical (1=free-flow, 2=mixed-flow)	Video observation
Environmental characteristics			
WC	Weather conditions during lane-changing event	Categorical (1=clear, 2=others (rain, snow, and fog)	Video observation
Roadway characteristics			
R (ft)	Curve radius during lane-changing event	Continuous	RID
SL (km/h)	Speed limit during lane-changing event	Continuous	RID
Number of freeway lanes (lanes)	Number of lanes during lane-changing event	Continuous	RID
Motivation parameters			
LC type	Type of lane-changing during lane-changing event	Categorical (1=mandatory, 2=discretionary)	Video observation

Note: FV = following vehicle; LCV = lane-changing vehicle; LV = lead vehicle; RH = rear headway; acc. = acceleration; sp. = speed; RS = relative speed; TC = traffic conditions; WC = weather conditions; R = curve radius; SL = speed limit; RID = Roadway Information Database; LC = lane-changing.

algorithm. The explanatory variables are the potential contributing factors that might have impact on gap acceptance behavior including, traffic flow parameters, environmental characteristics, roadway, and motivation characteristics. The variables were selected based on their importance on gap acceptance behavior and previous studies (*5*, *8*). Table 2 summarizes different variables used to set for lead and lag gap models.

This study utilized a nonparametric MARS model to investigate the factors affecting gap acceptance behavior. To better interpret the obtained model, the maximum number of interactions was set as two, as supported by previous studies (*10*, *16*). Several numbers of BFs were examined to identify the optimum maximum number of BFs for both models. The optimum maximum number of BFs selected was 30 based on the lowest root mean square error (RMSE) value (*16*). Two types of BFs were obtained from the developed MARS model; simple/elementary BFs and complex BFs. Simple BFs include only single variable and no interaction, while complex BFs allow interaction between variables (*20*).

MARS Model for Lead Gaps. MARS model for lead gap acceptance was developed using the explanatory variables provided in Table 2. Table 3 presents developed MARS model for lead gaps. The results of MARS model for lead gaps showed that several variables were involved in simple and complex BFs. As can be seen in Table 3, the explanatory variables did not have a single direction

Table 3. MARS Model for Lead Gap Acceptance

BFs	Basis function	Estimate	Standard error	p-value
BF0	Constant	0.71332	0.24996	0.00452
BF2	[0, 58.3361–(RS ($V_{LCV} - V_{LV}$))]	Not sig.	Not sig.	Not sig.
BF3	(0, RH–76.096)	−2.53186	0.19669	0.00000
BF4	(0, 76.096–RH)	Not sig.	Not sig.	Not sig.
BF5	(0, (RS ($V_{LCV} - V_{LV}$))–8.1244) × BF3	0.05268	0.00290	0.00000
BF6	[0, 8.1244–(RS ($V_{LCV} - V_{LV}$))] × BF3	−0.05042	0.00854	0.00000
BF7	(0, RH–75.872) × BF2	0.04916	0.00438	0.00000
BF8	(0, 75.872–RH) × BF2	0.00059	0.00017	0.00052
BF9	[TC in (1)]	7.00157	0.95816	0.00000
BF12	(0, 45.202–LV sp.) × BF4	0.00341	0.00068	0.00000
BF15	(0, Lanes–2) × BF9	−0.59407	0.12238	0.00000
BF17	(0, LCV SP.–4.4691) × BF9	−0.03607	0.00707	0.00000
BF18	(0, (RS ($V_{LCV} - V_{FV}$))–29.0629) × BF4	−0.00323	0.00059	0.00000
BF19	[0, 29.0629–(RS ($V_{LCV} - V_{FV}$))] × BF4	−0.00092	0.00031	0.00288
BF20	(0, LV sp.–122.005) × BF9	−0.07589	0.02199	0.00061
BF21	(0, 122.005–LV sp.) × BF9	−0.03741	0.00887	0.00003
BF22	(0, (RS ($V_{LCV} - V_{LV}$))–20.636) × BF9	−0.05338	0.02277	0.01948
BF23	[0, 20.636–(RS ($V_{LCV} - V_{LV}$))] × BF9	−0.09273	0.01823	0.00000
BF24	(0, (RS ($V_{LCV} - V_{LV}$))–30.503)	0.17035	0.02964	0.00000
BF25	[0, 30.503–(RS ($V_{LCV} - V_{LV}$))]	Not Sig.	Not Sig.	Not Sig.
BF26	(0, RH–92.48) × BF25	0.00426	0.00175	0.01550
BF29	(0, 83.0836–LV sp.)	−0.03610	0.01224	0.00336

Note: BF = Basic Function; LCV = lane-changing vehicle; LV = lead vehicle; RH = rear headway; RS = relative speed.

of impact on the lead gaps. As mentioned previously, the developed knots split the explanatory variables into several regions meaning that variables might have a different impact on lead gaps. For instance, BF5 and BF6 represent two complex BFs for the (RS ($V_{LCV} - V_{LV}$)), where a knot is developed at a value of 8. Additionally, the two BFs interact with a single BF (BF3), where the knot is developed at a value of 76. The coefficient of BF5 has a positive value and BF6 has a negative value. The estimated parameters of BFs in Table 3 were significant at a 95% confidence level.

According to Table 3, several complex BFs (i.e., interaction) present in the lead gaps model. Therefore, both the main effect variable and the interaction term in the BF should be considered to interpret the interaction term. For example, for the BF15 for the interaction of traffic condition and number of freeway lanes, the equation can be written as:

$$-0.59407 \times (0, \text{Lanes} - 2) \times (\text{Traffic\_Condition in}(1)) \tag{6}$$

Considering the free-flow traffic conditions (i.e., BF9 = 1), the final estimate of the interaction would be −0.59407 representing that if the lane-changing event occurs in more than two lanes in freeway and free-flow traffic conditions, the lead gaps will be lower. The effects of other interactions could be explained in a similar approach.

In MARS model, the effect of a given explanatory variable on the response variable could be easily identified by using the *Rate of Impression Expression*. It is worth mentioning that insignificant BFs needed to be kept in the final model to calculate *Rate of Impression Expression* for an explanatory variable (*10*, *16*). For instance, the effect of LV speed on the lead gaps can be identified considering several steps (*20*). First, all the BFs that involve with LV speed should be selected and combined. For example, the relevant BFs for the LV speed would be BF4, BF9, BF12, BF20, BF21, and BF29. Afterward, the knots of the LV speed (i.e., 45.202, 76.096, 83.0836, and 122.005) and set of intervals/ranges considering the knots should be identified. Subsequently, for each interval, the BFs that are associated with that particular interval should be selected. Next, the *Impact Expression* for each specific interval could be obtained from the MARS equation by selecting the associated BFs for the given interval. The *Rate of Impression Expression* for the LV speed can be determined from the first derivative of the *Impact Expression* with respect to the LV speed. Based on the *Rate of Impression Expression*, the direction of impact could be positive, negative, or no impact. Table 4 shows the *Rate of Impression*

Table 4. Impact of LV Speed on Lead Gaps

Combined impact of LV speed: 0.00341 × (45.202-LV Sp.) × (76.096-RH) – 0.07589 × (LV Sp.-122.005) × (TC in (1)) – 0.03741 × (122.005 - LV Sp.) × (TC in (1)) – 0.03610 (83.0836-LV Sp.)			
Interval/Range	Impact Expression	Rate of Impression Expression	Direction(s) of Impact
≤; 45.202	0.00341× (45.202-LV Sp.) × (76.096-RH) - 0.03741× (122.005 - LV Sp.) × (TC in (1)) − 0.03610 × (83.0836-LV Sp.)	-0.00341(76.096-RH) + 0.03741 × (TC in (1)) + 0.03610	+ve (0.07351) if RH ≥76.096 m in free-flow conditions +ve if 54.539 m<RH<76.096 m in free-flow conditions -ve if RH<54.539 m in free-flow conditions
45.202 < LV Sp.≤ 83.0836	-0.03741 × (122.005 - LV Sp.) × (TC in (1)) − 0.03610 × (83.0836-LV Sp.)	0.03741 × (TC in (1)) + 0.03610	+ve (0.07351) in free-flow conditions
83.0836 < LV Sp.≤ 122.005	-0.03741 × (122.005 - LV Sp.) × (TC in (1))	0.03741 × (TC in (1))	+ve (0.03741) in free-flow conditions
>122.005	-0.07589 × (LV Sp.-122.005) × (TC in (1))	-0.07589 × (TC in (1))	-ve (–0.07589) in free-flow conditions

Note: LV = lead vehicle; RH = rear headway; TC = traffic conditions.

Expression for the LV speed on the lead gaps for each identified interval.

As can be seen from Table 4, LV speed had a different rate of impact on lead gaps in four different intervals. It is also worth noting that the directions of impact of LV speed also depend on the minimum rear headway (RH) for the first interval (i.e., not more than 45 km/h). For instance, the lead gaps would be increased for the minimum RH greater than or equal 76 m under free-flow traffic conditions considering the first interval. This is also true for any minimum RH value from 54.5 to 76 m. However, lead gaps would be decreased with any value of less than 54.5 m. According to Table 4, lead gaps would be increased with the value of LV speed ranging from more than 45 to 122 km/h under free-flow conditions (i.e., second and third interval). Moreover, lead gaps would be decreased in free-flow conditions (fourth interval) for any LV speed of greater than 122 km/h. The developed MARS model for lead gaps can be expressed as Equation 7.

$$
\begin{aligned}
Y = {} & 0.71332 - 2.53186 \times BF3 + 0.05268 \times BF5 \\
& - 0.05042 \times BF6 + 0.04916 \times BF7 + 0.00059 \\
& \times BF8 + 7.00157 \times BF9 + 0.00341 \times BF12 \\
& - 0.59407 \times BF15 - 0.03607 \times BF17 \\
& - 0.00323 \times BF18 - \\
& 0.00092 \times BF19 - 0.07589 \times BF20 \\
& - 0.03741 \times BF21 - 0.05338 \times BF22 \\
& - 0.09273 \times BF23 + 0.17035 \\
& \times BF24 + 0.00426 \times BF26 - 0.03610 \times BF29
\end{aligned} \tag{7}
$$

MARS Model for Lag Gaps. MARS model was developed for the given explanatory variables considering lag gaps as a response variable. The results of the developed MARS model for lag gaps are provided in Table 5. Similar to the lead gaps model, the explanatory variables had no single direction of impact on the lag gaps and nonlinear performance was distinguished for all explanatory variables. The estimated parameters of BFs in Table 5 were significant at a 95% confidence level.

In addition, there are several complex BFs existing in the lag gap model. As discussed earlier, the effect of a particular explanatory variable on the lag gaps could be identified using the *Rate of Impression Expression*. For instance, the *Rate of Impression Expression* for the FV acceleration on the lag gaps is provided in Equations 8–10.

$$
\begin{aligned}
& \text{Impact Expression}: -0.03813 \\
& \times (\text{FV Acc.} + 0.0232) \times (\text{LCVSp.} - 4.46912)
\end{aligned} \tag{8}
$$

$$
\begin{aligned}
& \text{Rate of Impression Expression}: \\
& - 0.03813 \times (\text{LCV Sp.} - 4.46912)
\end{aligned} \tag{9}
$$

$$
\begin{aligned}
& \text{Direction}(s)\text{of Impact}: \\
& + \text{ve if LCV Sp.} > 4.46912, \text{Zero if else}
\end{aligned} \tag{10}
$$

According to Equations 8–10, it is observed that the effect of FV acceleration on lag gaps depends on the LCV speed. For instance, FV acceleration had a positive effect on lag gaps for LCV speed greater than 4.5 km/h, whereas, FV acceleration had no effect on lag gaps for LCV speed less

Table 5. MARS Model for Lag Gap Acceptance

BFs	Basis function	Estimate	Standard error	p-value
BF0	Constant	1.68551	0.13865	0.00000
BF1	(0, LCV sp.–4.46912)	Not sig.	Not sig.	Not sig.
BF2	(0, RH–72.16) × BF1	−0.00139	0.00019	0.00000
BF3	(0, 72.16–RH) × BF1	−0.00195	0.00017	0.00000
BF4	(0, RH–72.16)	0.08951	0.01286	0.00000
BF5	(0, 72.16–RH)	0.12498	0.01053	0.00000
BF7	(0, LCV sp.–64.3712) × BF5	0.00185	0.00022	0.00000
BF11	(0, R–2127)	−0.00005	0.00002	0.00321
BF12	(0, 2127–R)	−0.00007	0.00003	0.02670
BF13	(0, LCV sp.–102.268) × BF11	0.000006	0.00000	0.00156
BF14	(0, 102.268–LCV Sp.) × BF11	0.00001	0.00000	0.00000
BF16	(0, 5–lanes) × BF5	−0.00875	0.00150	0.00000
BF18	(0, 4–lanes) × BF1	0.00342	0.00064	0.00000
BF19	(0, LCV Sp.–84.3765)	−0.01428	0.00457	0.00191
BF20	(0, 84.3765–LCV Sp.)	Not sig.	Not sig.	Not sig.
BF21	(0, lanes–3) × BF20	−0.02436	0.00308	0.00000
BF23	(0, LCV acc.–0.121325)	−0.35226	0.08879	0.00009
BF25	(0, RH–69.76) × BF19	0.00173	0.00036	0.00000
BF27	(0, FV acc. + 0.0232) × BF1	−0.03813	0.01048	0.00031
BF29	(0, RS ($V_{LCV} - V_{FV}$)−3.35113) × BF20	−0.00062	0.00006	0.00000
BF30	(0, 3.35113–RS ($V_{LCV} - V_{FV}$)) × BF20	−0.01089	0.00321	0.00077

Note: FV = following vehicle; LCV = lane-changing vehicle; LV = lead vehicle; RH = rear headway; RS = relative speed; R = curve radius.

than or equal 4.5 km/h. The developed MARS model for lag gaps can be expressed as Equation 11.

$$\begin{aligned} Y = {} & 1.68551 - 0.00139 \times BF2 - 0.00195 \times BF3 \\ & + 0.08951 \times BF4 + 0.12498 \times BF5 + 0.00185 \\ & \times BF7 - 0.00005 \times BF11 - 0.00007 \times BF12 \\ & + 0.000006 \times BF13 + 0.00001 \times BF14 \\ & - 0.00875 \times BF16 \\ & + 0.00342 \times BF18 - 0.01428 \times BF19 \\ & - 0.02436 \times BF21 - 0.35226 \times BF23 \\ & + 0.00173 \times BF25 \\ & - 0.03813 \times BF27 - 0.00062 \times BF29 \\ & - 0.01089 \times BF30 \end{aligned} \tag{11}$$

Variables Importance for Lead and Lag Gap Models. Table 6 provides the relative variable importance for lead and lag gap models, which is one of the most important MARS model outputs. As can be seen in Table 6, relative speed between LCV and LV turned out to be the most important variable affecting lead gap acceptance behavior. This indicates that relative speed between LCV and LV plays a significant role in lead gap acceptance. The finding is consistent with previous studies exhibiting the effect of relative speed between LCV and LV on lead gap acceptance behavior (*5*, *8*). In contrast, LCV speed was found to be the most important variable affecting lag gap acceptance behavior. Minimum RH was the second important variable influencing both gap acceptance behaviors. Additionally, traffic conditions were the third important variable that affects lead gap acceptance, which is also consistent with a previous study (*8*). Table 6 also reveals that LV and LCV speeds, relative speed between LCV and FV, and the freeway number of lanes are the additional contributing factors affecting lead gap acceptance behavior. However, relative speed between LCV and FV was the third important variable affecting lag gap acceptance followed by the freeway number of lanes, curve radius, and LCV and FV accelerations. It is worth noting that the relative speed between LCV and FV was found to be one of the common important factors in both gap acceptance behavior, which is also supported by a previous study (*8*).

Linear Regression Models for Lead and Lag Gaps. The study developed a multiple linear regression model for lead and lag gaps in addition to the MARS model. The same previously used explanatory variables were considered for developing the models. Table 7 summarizes the estimated parameters of lead and lag gap models. Only statistically significant variables at 95% confidence level were retained in the final models. It can be seen from Table 7 that several variables, including minimum RH, LCV speed, LV speed, relative speed between LCV and LV, and some of the interaction terms were found to be significant in the developed lead gap acceptance model. The results showed that lead gaps increased with the increase of relative speed between LCV and LV. The

Table 6. Relative Importance of Variables for Lead and Lag Gap Models

Lead gap		Lag gap	
Variable	Score	Variable	Score
RS ($V_{LCV} - V_{LV}$)	100	LCV sp.	100
RH	78.80	RH	85.58
TC	16.76	RS ($V_{LCV} - V_{FV}$)	55.81
LV sp.	12.83	Lanes	44.83
LCV sp.	10.04	R	34.83
RS ($V_{LCV} - V_{FV}$)	9.81	LCV acc.	18.81
Lanes	9.40	FV acc.	16.21

Note: LCV = lane-changing vehicle; LV = lead vehicle; R = curve radius; RH = rear headway; RS = relative speed; TC = traffic conditions.

interaction between LCV speed and traffic conditions indicated that lead gaps increased with the decrease of LCV speed under mixed-flow traffic. However, several interaction terms were found to be significant in the developed lag gap acceptance model. For instance, the interaction between minimum RH and FV acceleration suggested that the decrease of FV acceleration increased the effect of minimum RH on lag gap acceptance. It was observed that the R^2 values of the linear models were significantly lower than the developed MARS models. Additionally, MSE values of the linear models were found to be significantly higher than the MARS models. Comparing these fit statistics, it can be concluded that MARS models performed better, confirming the non-linear trend in the dataset.

Conclusions

The primary focus of this study was to achieve better insights into lane-changing gap acceptance behavior using trajectory-level data collected in naturalistic settings. The study utilized one of the most comprehensive NDS and RID big datasets collected by the SHRP2. A nonparametric MARS approach was adopted to investigate the causal factors influencing driver gap acceptance behavior. Moreover, this study developed a unique methodology to automatically identify lane-changing events of the vehicles adjacent to the NDS vehicle's lane from the SHRP2 NDS database. In this study, 599 lane-changing events were automatically identified utilizing the developed methodology and manually verified. All 599 identified lane-changing events were used to extract required parameters necessary for analyzing gap acceptance behavior.

Investigation of safety conditions of the vehicles using TTC demonstrated that drivers tend to be more cautious when performing lane-changing maneuvers in adverse weather compared with clear weather in all traffic conditions. Summary statistics of lead and lag gaps showed that several statistics of lead gaps (i.e., mean and maximum) are relatively higher than lag gaps indicating that the perception of lag gaps are inconsistent compared with their lead gaps. Additionally, it was observed that drivers lead and lag gap acceptance followed gamma and lognormal distributions, respectively.

The MARS models of lead and lag gaps were developed to account for complex relationship between variables affecting gap acceptance behavior. Different factors that would affect gap acceptance were identified under several conditions using the developed knots in the MARS model. Considering all the explanatory variables in both models, relative speed between LCV and LV, as well as LCV speed turned out to be the most important variables affecting lead and lag gap acceptance, respectively. In addition, traffic conditions, acceleration of LCV and FV, and roadway geometric characteristics had significant effects on gap acceptance

Table 7. Linear Regression Models for Lead and Lag Gap Acceptance

Lead gap					Lag gap				
Parameter	Estimate	Std. error	*t*-value	*p*-value	Parameter	Estimate	Std. error	*t*-value	*p*-value
Intercept	1.07031	1.67103	0.64051	0.52209	Intercept	2.41033	0.26676	9.03565	0.00000
RH	0.02058	0.00750	2.74258	0.00628	RH	0.02134	0.00399	5.35461	0.00000
LCV sp.	−0.03066	0.01373	−2.23351	0.02589	LCV sp.	−0.02081	0.00181	−11.52837	0.00000
LV sp.	0.04275	0.01132	3.77800	0.00017	FV acc.	1.43348	1.24467	1.15169	0.24991
RS ($V_{LCV} - V_{LV}$)	0.04383	0.01459	3.00353	0.00278	RS ($V_{LCV} - V_{FV}$)	−0.03395	0.00462	−7.35332	0.00000
RS ($V_{LCV} - V_{FV}$)	−0.01161	0.02212	−0.52495	0.59982	R	0.00004	0.00001	2.90348	0.00383
TC	2.98143	2.16669	1.37603	0.16934	Lanes	0.15392	0.06204	2.48095	0.01338
LCV sp. × TC	−0.04990	0.02026	−2.46229	0.01409	RH × FV acc.	−0.07661	0.03141	−2.43857	0.01504
RS ($V_{LCV} - V_{FV}$)× TC	0.17064	0.04003	4.26225	0.00002	RH × R	−0.0000005	0.0000001	−3.21073	0.00140
NA	NA	NA	NA	NA	RH × lanes	−0.00445	0.00142	−3.12752	0.00185
NA	NA	NA	NA	NA	LCV sp. × RS ($V_{LCV} - V_{FV}$)	0.00033	0.00006	5.89433	0.00000

Note: FV = following vehicle; LCV = lane-changing vehicle; LV = lead vehicle; R = curve radius; RH = rear headway; RS = relative speed; TC = traffic conditions; NA = not available.

behavior. Moreover, linear models of lead and lag gaps were developed to investigate the differences in results and it was observed that MARS models performed better confirming the nonlinear trend in the dataset.

The results of this study can help contributing to the calibration of traffic microsimulation and provide guidance to AVs. The fact that human drivers were more cautious in performing lane-changing maneuvers under adverse weather suggests that the findings could be employed as a set of guidelines for AV so that they can learn similar responses consistent with human-driven vehicles. This will be critical to increase AV acceptance. The distributions of lead and lag gaps using naturalistic data could be used as input in microsimulation lane-changing models to generate simulated values of lead and lag gaps (*39*). Using similar data to NDS, the developed automatic identification algorithm could be used to extract necessary gap acceptance parameters in real-time in a connected and autonomous vehicles (CAV) environment. As the developed models provided overall perception into gap acceptance behavior with several variables, the models have important implications for safety improvements in CAV. More specifically, the developed models can help with assessing and ameliorating active safety systems in CAV (*8*).

The study provided significant contributions to the state-of-research in lane-changing gap acceptance behavior; however, it is worth noting some of inherited limitations of the SHRP2 dataset. The headway between the LCV and its LV in their original lane is not available because the LCV is not a SHRP2 NDS-equipped vehicle with a front radar. Moreover, drivers' demographics for surrounding non-SHRP2 NDS vehicles are unknown. These limitations could be mitigated in future research once data from AV could be accessed at a 100% Market Penetration Rate (MPR) in the near future. In addition, future study is needed with more lane-changing events to identify the effect of other additional factors including specific weather conditions, that is, rain, snow, and fog on driver gap acceptance behavior.

Author Contributions

The authors confirm the contribution to the paper as follows: study conception and design: A. Das, M. N. Khan, M. M. Ahmed; data preparation and reduction: A. Das, M. N. Khan, and M. M. Ahmed; analysis and interpretation of results: A. Das, M. N. Khan, M. M. Ahmed; draft manuscript preparation: A. Das, M. N. Khan, and M. M. Ahmed. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was conducted under the Strategic Highway Research Program 2 (SHRP2), which is administrated by the Transportation Research Board of the National Academies of Sciences, Engineering, and Medicine, and it was sponsored by the Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials (AASHTO) and the Wyoming Department of Transportation (WYDOT) (grant number RS08217).

References

1. National Center for Statistics and Analysis. *Traffic Safety Facts 2015: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System*. National Center for Statistics and Analysis, Washington, D.C., 2015.
2. You, F., R. Zhang, G. Lie, H. Wang, H. Wen, and J. Xu. Trajectory Planning and Tracking Control for Autonomous Lane Change Maneuver Based on the Cooperative Vehicle Infrastructure System. *Expert Systems with Applications*, Vol. 42, No. 14, 2015, pp. 5932–5946. https://doi.org/10.1016/j.eswa.2015.03.022.
3. Hou, Y., P. Edara, and C. Sun. Situation Assessment and Decision Making for Lane Change Assistance Using Ensemble Learning Methods. *Expert Systems with Applications*, Vol. 42, No. 8, 2015, pp. 3875–3882. https://doi.org/10.1016/j.eswa.2015.01.029.
4. Ahmed, K. I. *Modeling Drivers' Acceleration and Lane Changing Behavior*. *Doctoral thesis*. Massachusetts Institute of Technology, 1999, pp. 185–189.
5. Toledo, T., H. N. Koutsopoulos, and M. E. Ben-akiva. Modeling Integrated Lane-Changing Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2003. 1857: 30–38.
6. Hill, C., L. Elefteriadou, and A. Kondyli. Exploratory Analysis of Lane Changing on Freeways Based on Driver Behavior. *Journal of Transportation Engineering*, Vol. 141, No. 2006, 2014, pp. 1–11. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000758.
7. Lee, J., M. Park, and H. Yeo. A Probability Model for Discretionary Lane Changes in Highways. *KSCE Journal of Civil Engineering*, Vol. 20, No. 7, 2016, pp. 2938–2946. https://doi.org/10.1007/s12205-016-0382-z.
8. Wang, X., M. Yang, and D. Hurwitz. Analysis of Cut-in Behavior Based on Naturalistic Driving Data. *Accident Analysis and Prevention*, Vol. 124, 2019, pp. 127–137. https://doi.org/10.1016/j.aap.2019.01.006.
9. Das, A., M. M. Ahmed, and A. Ghasemzadeh. Using Trajectory-Level SHRP2 Naturalistic Driving Data for Investigating Driver Lane-Keeping Ability in Fog: An Association Rules Mining Approach. *Accident Analysis and Prevention*, Vol. 129, 2019, pp. 250–262. https://doi.org/10.1016/j.aap.2019.05.024.
10. Ghasemzadeh, A., and M. M. Ahmed. 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*, Vol. 90, 2018, pp. 379–392. https://doi.org/10.1016/j.trc.2018.03.018.

11. Das, A., and M. M. Ahmed. Exploring the Effect of Fog on Lane-Changing Characteristics Utilizing the SHRP2 Naturalistic Driving Study Data. *Journal of Transportation Safety and Security*, 2019, pp. 1–26. https://doi.org/10.1080/19439962.2019.1645777.
12. Haleem, K., A. Gan, and J. Lu. Using Multivariate Adaptive Regression Splines (MARS) to Develop Crash Modification Factors for Urban Freeway Interchange Influence Areas. *Accident Analysis and Prevention*, Vol. 55, 2013, pp. 12–21. https://doi.org/10.1016/j.aap.2013.02.018.
13. Park, J., and M. Abdel-Aty. Assessing the Safety Effects of Multiple Roadside Treatments Using Parametric and Nonparametric Approaches. *Accident Analysis and Prevention*, Vol. 83, 2015, pp. 203–213. https://doi.org/10.1016/j.aap.2015.07.008.
14. Haleem, K., M. Abdel-Aty, and J. Santos. Multiple Applications of Multivariate Adaptive Regression Splines Technique to Predict Rear-End Crashes at Unsignalized Intersections. *Transportation Research Record: Journal of the Transportation Research Board*, 2010. 2165: 33–41. https://doi.org/10.3141/2165-04.
15. Chang, L. Y., H. C. Chu, D. J. Lin, and P. Lui. Analysis of Freeway Accident Frequency Using Multivariate Adaptive Regression Splines. *Procedia Engineering*, Vol. 45, 2012, pp. 824–829. https://doi.org/10.1016/j.proeng.2012.08.245.
16. Gaweesh, S. M., M. M. Ahmed, and A. V. Piccorelli. Developing Crash Prediction Models Using Parametric and Nonparametric Approaches for Rural Mountainous Freeways: A Case Study on Wyoming Interstate 80. *Accident Analysis and Prevention*, Vol. 123, 2019, pp. 176–189. https://doi.org/10.1016/j.aap.2018.10.011.
17. Xu, Y., Q. J. Kong, R. Klette, and Y. Liu. Accurate and Interpretable Bayesian MARS for Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 15, No. 6, 2014, pp. 2457–2469. https://doi.org/10.1109/TITS.2014.2315794.
18. Oduro, S. D., S. Metia, H. Duc, G. Hong, and Q. P. Ha. Multivariate Adaptive Regression Splines Models for Vehicular Emission Prediction. *Visualization in Engineering*, Vol. 3, No. 1, 2015. https://doi.org/10.1186/s40327-015-0024-4.
19. Chen, Y., L. Zhu, J. Gonder, S. Young, and K. Walkowicz. Data-Driven Fuel Consumption Estimation: A Multivariate Adaptive Regression Spline Approach. *Transportation Research Part C: Emerging Technologies*, Vol. 83, 2017, pp. 134–145. https://doi.org/10.1016/j.trc.2017.08.003.
20. Osei-Bryson, K. M., and O. Ngwenyama. *Advances in Research Methods for Information Systems Research*. Springer, New York, 2014.
21. Hutton, J. M., K. M. Bauer, C. A. Fees, and A. Smiley. *Analysis of Naturalistic Driving Study Data: Offset Left-Turn Lanes*. SHRP 2 Report. Transportation Research Board, Washington, D.C., 2014.
22. Campbell, K. L. The SHRP 2 Naturalistic Driving Study. *TR News*, 2012, pp. 30–35.
23. Center for Transportation Research and Education (CTRE). *SHRP2-Roadway Information Database - Background*. http://www.ctre.iastate.edu/shrp2-rid/. Accessed July 31, 2019.
24. Ghasemzadeh, A., and M. M. Ahmed. Quantifying Regional Heterogeneity Effect on Drivers' Speeding Behavior Using SHRP2 Naturalistic Driving Data: A Multilevel Modeling Approach. *Transportation Research Part C: Emerging Technologies*, Vol. 106, 2019, pp. 29–40. https://doi.org/10.1016/j.trc.2019.06.017.
25. Hammit, B. E., A. Ghasemzadeh, R. M. James, M. M. Ahmed, and R. K. Young. Evaluation of Weather-Related Freeway Car-Following Behavior Using the SHRP2 Naturalistic Driving Study Database. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 59, 2018, pp. 244–259. https://doi.org/10.1016/j.trf.2018.08.023.
26. Hammit, B. E., R. James, M. Ahmed, and R. Young. Towards the Development of Weather-Dependent Microsimulation Models. *Transportation Research Record: Journal of the Transportation Research Board*, 2019. 2673: 143–156.
27. Ali, E. M., M. M. Ahmed, and S. S. Wulff. Detection of Critical Safety Events on Freeways in Clear and Rainy Weather Using SHRP2 Naturalistic Driving Data: Parametric and Non-Parametric Techniques. *Safety Science*, Vol. 119, 2019, pp. 1–9. https://doi.org/10.1016/j.ssci.2019.01.007.
28. Ahmed, M. M., and A. Ghasemzadeh. The Impacts of Heavy Rain on Speed and Headway Behaviors: An Investigation Using the SHRP2 Naturalistic Driving Study Data. *Transportation Research Part C: Emerging Technologies*, Vol. 91, 2018, pp. 371–384. https://doi.org/10.1016/j.trc.2018.04.012.
29. Ahmed, M., R. Young, A. Ghasemzadeh, B. Hammit, E. Ali, N. Khan, A. Das, and H. Eldeeb. *Implementation of SHRP2 Results within the Wyoming Connected Vehicle Variable Speed Limit System: Phase 2 Early Findings Report and Phase 3 Proposal*. Department of Civil and Architectural Engineering, University of Wyoming, Laramie, 2017.
30. Khan, M. N., and M. M. Ahmed. Snow Detection Using In-Vehicle Video Camera with Texture-Based Image Features Utilizing K-Nearest Neighbor, Support Vector Machine, and Random Forest. *Transportation Research Record: Journal of the Transportation Research Board*, 2019. 2673: 21–32.
31. Ahmed, M. M., A. Ghasemzadeh, B. Hammit, N. Khan, A. Das, E. Ali, R. K. Young, and H. Eldeeb. *Driver Performance and Behavior in Adverse Weather Conditions: An Investigation Using the SHRP2 Naturalistic Driving Study Data-Phase 2*. Final Report WY-18/05F. Wyoming Department of Transportation, 2018.
32. Das, A., A. Ghasemzadeh, and M. M. Ahmed. Analyzing the Effect of Fog Weather Conditions on Driver Lane-Keeping Performance Using the SHRP2 Naturalistic Driving Study Data. *Journal of Safety Research*, Vol. 68, 2019, pp. 71–80. https://doi.org/10.1016/j.jsr.2018.12.015.
33. Gorman, T., L. Stowe, and J. Hankey. *S31: NDS Data Dissemination Activities, Task 1.6: Radar Post-Processing*.

Transportation Research Board of the National Academies, Washington, D.C., 2015.
34. MathWorks. *Fillmissing*. https://www.mathworks.com/help/matlab/ref/fillmissing.html.
35. Peng, C. Y. J., M. Harwell, S. Liou, and L. H. Ehman. Advances in Missing Data Methods and Implications for Educational Research. In *Real Data Analysis* (S. S. Sawilowsky, ed.), Information Age Publishing, Charlotte, N.C., 2006, pp. 31–78.
36. Peyre, H., A. Leplège, and J. Coste. Missing Data Methods for Dealing with Missing Items in Quality of Life Questionnaires. A Comparison by Simulation of Personal Mean Score, Full Information Maximum Likelihood, Multiple Imputation, and Hot Deck Techniques Applied to the SF-36 in the French. *Quality of Life Research*, Vol. 20, No. 2, 2011, pp. 287–300. https://doi.org/10.1007/s11136-010-9740-3.
37. MathWorks. *Smoothdata*. https://www.mathworks.com/help/matlab/ref/smoothdata.html.
38. Insight Website. *SHRP2 NDS Data Access*. https://insight.shrp2nds.us/. Accessed April 15, 2019.
39. Bham, G. H. Estimating Driver Mandatory Lane Change Behavior on a Multi Lane Freeway. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
40. Ghasemzadeh, A., B. E. Hammit, M. M. Ahmed, and H. Eldeeb. Complementary Methodologies to Identify Weather Conditions in Naturalistic Driving Study Trips: Lessons Learned from the SHRP2 Naturalistic Driving Study & Roadway Information Database. *Safety Science*, Vol. 119, 2019, pp. 21–28. https://doi.org/10.1016/j.ssci.2019.01.006.
41. Khan, M. N., A. Ghasemzadeh, and M. M. Ahmed. 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*, 2018. 2672: 93–104.
42. Ghasemzadeh, A., B. E. Hammit, M. M. Ahmed, and R. K. Young. Parametric Ordinal Logistic Regression and Non-Parametric Decision Tree Approaches for Assessing the Impact of Weather Conditions on Driver Speed Selection Using Naturalistic Driving Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672: 137–147.
43. Friedman, J. H. Multivariate Adaptive Regression Splines. *The Annals of Statistics*, Vol. 11, No. 2, 1991, pp. 416–431.
44. Yang, M., X. Wang, and M. Quddus. Examining Lane Change Gap Acceptance, Duration and Impact Using Naturalistic Driving Data. *Transportation Research Part C: Emerging Technologies*, Vol. 104, 2019, pp. 317–331. https://doi.org/10.1016/j.trc.2019.05.024.
45. Lee, S. E., E. C. B. Olsen, and W. W. Wierwille. *A Comprehensive Examination of Naturalistic Lane-Changes*. Final Report DOT Hs 809 702. National Highway Traffic Safety Administration, Washington, D.C., 2004, pp. 1–236.
46. Van Der Horst, R., and J. Hogema. Time-to-Collision and Collision Avoidance Systems. *Proc., 6th ICTCT workshop: Safety Evaluation of Traffic Systems: Traffic Conflicts and Other Measures*, Salzburg, Austria, 1993, pp. 1–12.
47. Balal, E., R. Cheu, T. Gyan-Sarkodie, and J. Miramontes. Analysis of Discretionary Lane Changing Parameters on Freeways. *International Journal of Transportation Science and Technology*, Vol. 3, No. 3, 2014, pp. 277–296. https://doi.org/10.1260/2046-0430.3.3.277.