



# Utilizing naturalistic driving data for in-depth analysis of driver lane-keeping behavior in rain: Non-parametric MARS and parametric logistic regression modeling approaches



Ali Ghasemzadeh\*, Mohamed M. Ahmed

University of Wyoming, Department of Civil & Architectural Engineering, 1000 E University Ave, Dept. 3295, Laramie, WY 82071, United States

## ARTICLE INFO

### Keywords:

Naturalistic driving study  
Lane keeping  
Lane departure warning system  
Adverse weather conditions  
SHRP2  
Multivariate adaptive regression splines

## ABSTRACT

It is known that adverse weather conditions can affect driver performance due to reduction in visibility and slippery surface conditions. Lane keeping is one of the main factors that might be affected by weather conditions. Most of the previous studies on lane keeping have investigated driver lane-keeping performance from driver inattention perspective. In addition, the majority of previous lane-keeping studies have been conducted in controlled environments such as driving simulators. Therefore, there is a lack of studies that investigate driver lane-keeping ability considering adverse weather conditions in naturalistic settings. In this study, the relationship between weather conditions and driver lane-keeping performance was investigated using the SHRP2 naturalistic driving data for 141 drivers between 19 and 89 years of age. Moreover, a threshold was introduced to differentiate lane keeping and lane changing in naturalistic driving data. Two lane-keeping models were developed using the logistic regression and multivariate adaptive regression splines (MARS) to better understand factors affecting driver lane-keeping ability considering adverse weather conditions. The results revealed that heavy rain can significantly increase the standard deviation of lane position (SDLP), which is a very widely used method for analyzing lane-keeping ability. It was also found that traffic conditions, driver age and experience, and posted speed limits have significant effects on driver lane-keeping ability. An interesting finding of this study is that drivers have a better lane-keeping ability in roadways with higher posted speed limits. The results from this study might provide better insights into understanding the complex effect of adverse weather conditions on driver behavior.

## 1. Background

Driver behavior can be considered as the most unpredictable element in the driver, vehicle, and roadway triangle (Zhou et al., 2008). More specifically, while driving, drivers might change their behavior due to different mental and physical conditions, distractions introduced by in-vehicle technologies, or conversing with passengers, etc. (Suzdaleva and Nagy, 2018; Eboli et al., 2017a, 2017b). Driver lane-keeping ability is one of the main behavioral-performance factors that can be used to assess the broader implications of driving tasks (Ghasemzadeh and Ahmed, 2017). One of the basic operations that needs to be performed by each vehicle in the semi-automated or fully automated highway systems is lane keeping (Suryanarayanan and Tomizuka, 2007). Therefore, better understanding of driver lane-keeping behavior drew researchers' attention in recent years.

Lane-keeping performance has been investigated in previous studies from driver inattention standpoint (Harbluk et al., 2002;

\* Corresponding author.

E-mail addresses: [aghasemz@uwyo.edu](mailto:aghasemz@uwyo.edu) (A. Ghasemzadeh), [mahmed@uwyo.edu](mailto:mahmed@uwyo.edu) (M.M. Ahmed).

Reed-Jones et al., 2008; Chisholm et al., 2008; Collet et al., 2010; Peng et al., 2013). While findings from these studies indicate the effect of different forms of distractions on driver lane-keeping performance, the impact of adverse weather conditions on lane-keeping ability has not been well researched in previous studies.

Adverse weather conditions have negative impacts on both traffic safety and operations. Based on the National Highway Traffic Safety Administration data, weather contributed to over 22% of the total crashes each year (Hamilton, 2015). More specifically, most of the weather-related crashes occur on wet pavement and during rainfall representing 73% and 46% of weather-related crashes, respectively (Hamilton, 2015). Previous studies revealed that the visibility obstruction caused by rainfall increases crashes by 100% or more (Brodsky and Hakkert, 1988; National Transportation Safety Board, 1980), whereas other studies found more moderate but still significant increase (Andreescu and Frost, 1998; Allen et al., 1996).

In addition, previous studies utilized survey questionnaires and driving simulators to better understand drivers' behavioral and performance factors such as speed and headway selection and lane keeping performance in different adverse weather conditions (Wu et al., 2018; Hamdar et al., 2016; Hassan and Abdel-Aty, 2011; Brooks et al., 2011; Ni et al., 2010). Wu et al. (2018) investigated driver speed adjustments when using real-time fog warning systems in different fog levels in a driving simulator study. While their study concluded that adverse weather and limited visibility have a significant impact on driver performance, they found that warning systems do not impact drivers final speed choice in fog. Hamdar et al., (2016) analyzed the impact of weather conditions and geometric characteristics on driver longitudinal behavior. They claimed that drivers' speed, time/distance headways and time to collision can be affected by both weather conditions and roadway characteristics. In addition, they found that drivers select longer headways in low visibility conditions. Another survey-based study assessed factors affecting drivers' compliance with variable speed limits (VSL) and dynamic message signs (DMS) in various traffic and environmental conditions and found that drivers' satisfaction with VSL/DMS is the most determining factor affecting drivers' compliance with VSL/DMS under reduced visibility conditions (Hassan and Abdel-Aty, 2011). Drivers' speeding behavior and lane-keeping performance under foggy weather conditions were studied by Brooks et al. (2011) using a driving simulator. The results indicated that drivers were not able to reduce their speed efficiently and had poor lane keeping in thick fog conditions.

One of the main contributing factors in run-off-the-road crashes and near-crashes is driver lane-keeping performance. In fact, almost 20% of total crashes are single vehicle roadway departure crashes in the US (Dagan et al., 2004; Pomerleau and Everson, 1999). Lane Departure Warning (LDW) systems can reduce 30–70% of lane departure crashes by alerting driver in different dangerous departure situations, which could be caused by alcohol, drowsiness, and distraction (Xu et al., 2015). Adverse weather conditions might exacerbate poor lane keeping due to poor visibility and slippery surface conditions. Understanding driver behavior and response in adverse weather conditions, specifically when visibility is reduced below a certain level may be helpful not only in mitigating lane-departure crashes, but also in finding a more efficient threshold for LDW systems in adverse weather conditions.

Even though there are many studies that investigated the impact of driver behavior in adverse weather conditions, the studies that investigated the impact of heavy rain on driver behavior and performance on a microscopic scale are few (Yu et al., 2014; Ghasemzadeh and Ahmed, 2017). Therefore, there is a need to analyze driver performance using microscopic data. In recent years, Naturalistic Driving studies (NDS) worldwide are becoming more popular in addressing traffic safety and operation challenges. Utilizing NDS Data has made it possible to study driver behavior and performance at a microscopic level. To be specific, NDS will allow for better understanding of how drivers adjust their behaviors to compensate for increased risk due to reduction in visibility. The main goal of this study is to investigate driver lane-keeping ability using the NDS data obtained from the second Strategic Highway Research Program (SHRP2) in heavy rain and slippery road conditions.

Provided contributions in this paper are emphasized as follows. First, considering the fact that weather data is not readily available in the SHRP2 NDS data, three complementary methodologies were developed to identify weather-related trips from the massive SHRP2 NDS database. Second, drivers speed and lane position offset were compared under the free flow conditions and using matched trips (same driver, vehicle, and traversed routes) in heavy rain and clear weather conditions to provide a better insight into driver behavioral-factors when they are not affected by traffic conditions. Third, a threshold was proposed to eliminate lane-changing events from the dataset using a detailed trajectory analysis at 10 Hz (10 observations per second), considering that the focus of this study is to better understand driver lane-keeping ability not lane-changing. Forth, parametric logistic regression and non-parametric multivariate adaptive regression splines (MARS) models were developed to identify factors affecting drivers' lane-keeping performance.

## 2. Data source

This study utilized data from the SHRP2 safety data, which is comprised of two main datasets including Naturalistic Driving Study (NDS) data and the Roadway Information Database (RID). NDS refers to the real-time behavioral data that was collected from more than 3400 volunteer drivers in six site locations including New York, Florida, Washington, Indiana, Pennsylvania, and North Carolina. Each driver received an annual incentive of \$500 and was asked to provide access to his/her vehicle in order to remove and replace the hard drive with the collected data. Some participants remained in the study for several months and then replaced with new drivers and vehicles, while others remained for the full duration of the study. The project duration was about 39 months (October 2010–December 2013) and during this time drivers' behavior and their vehicle-use were recorded continuously, which made this project one of the largest naturalistic driving studies worldwide.

The Data Acquisition System (DAS) was developed by VTTI (Virginia Tech Transportation Institute), which includes four cameras, forward radar, GPS, vehicle network information, on-board computer vision lane tracking system, as well as other computer vision systems and capability of storing data. The collected NDS data includes but not limited to vehicle kinematics and network

information recorded at 10 Hz frequency (10 observations per second) such as speed, acceleration, and steering wheel position; video views including forward and rear views, and driver's face and hands view, in addition to information obtained from a variety of sensors such as forward radar, alcohol sensor, and accelerometers (Hankey et al., 2013).

As mentioned earlier, RID was also utilized in this study, which was developed by the Center for Transportation Research and Education (CTRE) of Iowa State University (Smadi et al., 2014). The RID contains a comprehensive description of roadway characteristics for the six NDS states (McLaughlin and Hankey, 2015). The roadway data was collected not only from the state highway department, but also from other sources including equipped vans that collected roadway characteristics while traveling at posted speed limits on most traversed routes by SHRP2 vehicles. The NDS and RID data were linked in this study to associate driver lane-keeping behavior with roadway characteristics.

Weather data extracted from the National Climate Data Center (NCDC) were also utilized in this study to initially identify NDS trips occurred in adverse weather conditions. It is worth noting that SHRP2 do not have weather conditions reported for the NDS normal driving trips, weather conditions are only provided for SHRP2 event data (i.e., crashes and near crashes). The NCDC archives weather information from different weather stations at a national level and provide public access to the historical weather data. Data extracted from the NCDC has been used in previous studies to determine real-time roadway weather conditions (Wu et al., 2017; Wang et al., 2015; Ahmed et al., 2014; Hassan and Abdel-Aty, 2013; Abdel-Aty et al., 2012).

### 3. Data preparation

The initial acquisition of data was crucial to the success of this study, and it presented a unique challenge for researchers to develop creative and unique methods for leveraging the full extent of the provided NDS and RID data. A Previous study utilized vehicle's wiper settings to identify weather-related trips in the NDS (Ahmed et al., 2015). While vehicle's wiper settings give an indication of precipitation intensity, results from previous studies indicated that wiper settings are not consistent between vehicles (Ghasemzadeh and Ahmed, 2017; Ahmed et al., 2017). Utilizing wiper settings reported by the NDS vehicle Data Acquisition System (DAS) for identifying weather-related trips has some limitations. First, provided wiper status indicates the position of the wiper switch, rather than wiper swipe rate; second, different drivers have different tolerances to precipitation/visibility, and splashes from neighboring vehicles may affect driver choice of the appropriate wiper speed. Third, the wiper setting was not provided for many trips – likely due to the high percentage of older vehicles (due to higher participation of younger drivers) or errors in the DAS recording. To compensate for these challenges, three complementary approaches to effectively extract weather-related trips were introduced in this study.

#### 3.1. Wiper status

The first approach used the windshield wiper method developed by the authors in a previous study (Ghasemzadeh and Ahmed, 2017). The process relies on the time series wiper status to identify trips where wipers were active at a high speed for an extended length of time along freeway segments.

#### 3.2. External weather data (weather stations and weather-related crashes)

In efforts to identify more NDS trips that overlap with particular adverse weather events, two innovative methodologies were developed using the weather data extracted from the National Climate Data Center (NCDC) and weather-related crashes from the RID. These data sources allowed researchers to enact a type of perimeter around the weather station or crash location and search for trips that occurred during the same period.

Ahmed et al. concluded that airport weather stations provided 60% to 80% accurate spatial-temporal weather conditions for adjacent roadways within five nautical miles and within a two-hour time period (Ahmed et al., 2014). In this study, daily weather data were acquired from the NCDC and NDS trips were requested based on the daily weather information to identify all trips impacted by adverse weather events (such as those occurred on icy or slushy road surfaces), not only those occurring during active precipitation. Therefore, the date and time for every weather event was superimposed on the NDS trips for freeways within 5 nautical miles. As an example, Fig. 1 shows weather stations used to identify the weather-related trips in Washington State and the 5 nautical miles coverage area used in the NDS trip data acquisition process. In addition, weather-related crashes reported in the RID were utilized, and crash locations with weather conditions were essentially considered “weather stations”. The same process, as described for weather stations was used to identify NDS trips within a five- nautical mile radius occurring within the period of weather-related crash.

Using mentioned three complementary data extraction processes, the research team received more than 11,000 potential weather-related trips and their matched trips in clear weather conditions. A matched trip in clear weather was defined as a trip in clear weather but for the same driver, vehicle, and traversed routes. The identified NDS trips represent 1523 drivers between 16 and 99 years of age with the majority of the drivers in the young age group, 16–29 years old. Gender representation was balanced in most age groups, with the exception of a slight overrepresentation of female drivers between 20 and 24 years old. The total duration of adverse weather trips plus two matched trips in clear conditions represents over 11,205 h of driving.

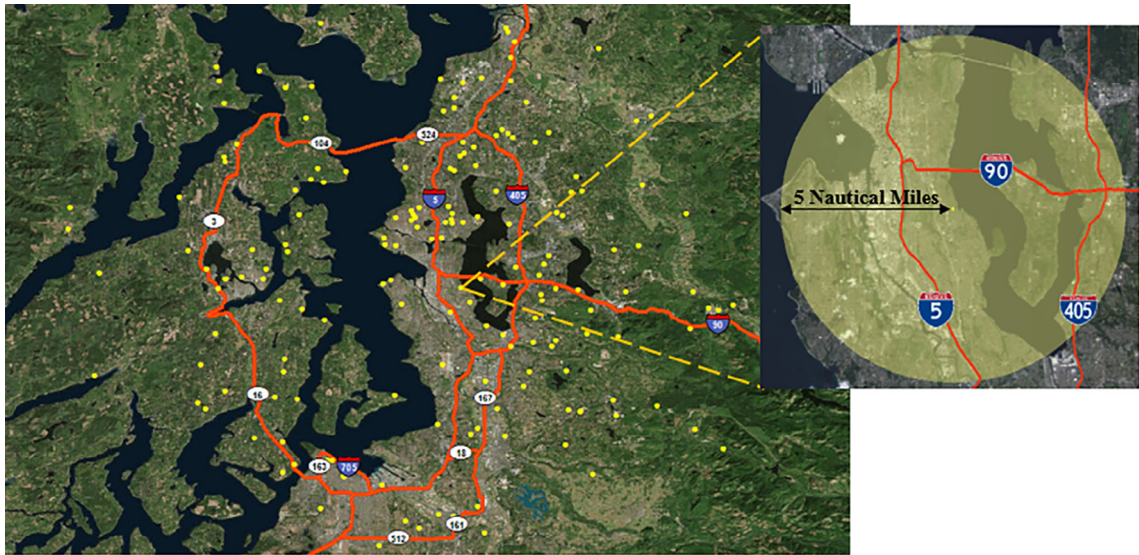


Fig. 1. Weather stations considered in Washington state and a five nautical miles coverage area for a sample weather station.

#### 4. Data reduction

The preliminary screening process was required because of the wide projection of adverse weather events on an entire day, but allowed for collection of trips that would not have been otherwise identified. In order to screen and eliminate trips in clear weather or dry surface conditions, extensive manual video observation and preliminary data reduction was carried out. Preliminary investigation of the received data revealed that 37% of the received freeway trips were confirmed to have adverse weather and/or surface conditions.

A previous study revealed high variability in weather conditions within a single trip (Ahmed et al., 2015). Therefore, each trip was divided into one-minute chunks to create homogeneous segments with identical environmental and traffic conditions. This process of splitting trips into segments is a novel analysis protocol introduced in this study to expand the capability of evaluating a significant number of trips, drivers, and conditions in the most efficient and effective manner possible.

While weather conditions can be obtained from weather stations, reporting frequency of weather stations is relatively long. Moreover, weather stations are mostly mounted at higher elevations and do not necessarily represent roadway surface weather/ and visibility limit (i.e., what the driver is actually seeing on the road). Therefore, identifying trajectory-based light rain and heavy rain from weather stations is not feasible. NDS video data on the other hand can provide accurate real-time weather and surface conditions from the driver's perspective. Therefore, NCDL weather data were utilized only as a starting point to identify potential locations/times where weather-related NDS took place. Then, manual video reduction and annotation was conducted to verify the roadway type, weather and surface conditions, visibility, and traffic conditions. A data reduction template for each trip was created with corresponding timestamps at one-minute sampling rate, to facilitate this process.

It is also worth mentioning that analysis of wipers status as well as visual inspections of all NDS videos were utilized to identify heavy/light rain and clear weather condition trips. More specifically, the research team received three files including forward and rear view cameras as well as an excel sheet collected time series data at 10 Hz for each trip. There is a variable in the time series data that represents wiper setting. The process started and relied on the analysis of wiper status variable. This variable represents settings of windshield wipers and can take following values: 0 = not active, 1 = intermittent, 2 = low, 3 = high, 254 = invalid data, and 255 = data not available. Those cases where wiper status were active at a high speed could be considered as heavy rain. However, as mentioned earlier, there are several issues with this approach. First, preliminary investigation of acquired trips revealed that some trips do not have the full spectrum of values for wiper setting. In fact, wiper setting was defined as only "on" and "off" for those trips. Second, it was also found that the wiper setting variable is missing for most of the acquired trips. Among the randomly selected trips in this study, almost 92% of the acquired trips were lack of wiper setting variable. Third, wiper settings are not consistent among different vehicles and the fact that vehicle wiper setting variable represents the position of wiper switch instead of wiper swipe rate. Therefore, to get the accurate real-time trajectory-level information about precipitation intensity, visibility, surface, and traffic conditions, manual observations of the trips were inevitable. Each trip was divided into one-minute segments to provide homogeneous segments with similar traffic and environmental conditions. An observation template for each trip was created with corresponding timestamps in one-minute time intervals, to facilitate this process. To eliminate subjectivity and any potential bias in identifying weather, traffic, or roadway conditions, comprehensive training and a detailed description of each condition, as described and shown in the Figs. 2–4, were provided to video observers before the manual data observation and reduction began. The video observers then leveraged the benefits of another developed software called the Wyoming NDS Visualization and Visibility Identification Tool to review the video and time series data for each trip. For more information about the developed software please see



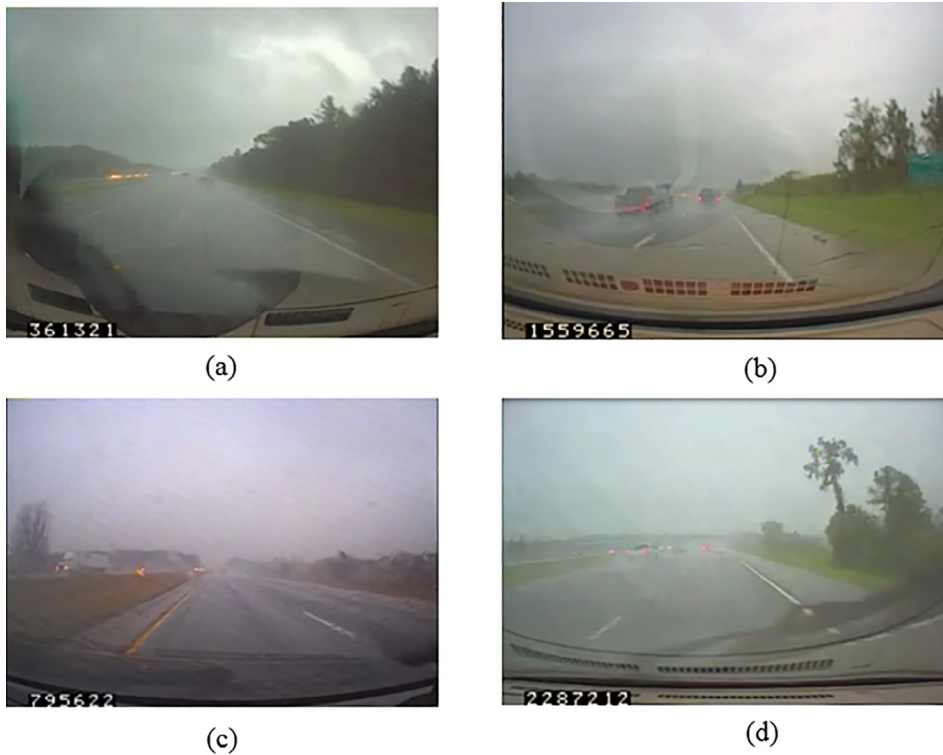


Fig. 2. Heavy Rain: Few road markings are visible, surroundings, traffic and the horizon cannot be seen properly (a, b, c, d), the sign is unreadable (b), wipers high speed.

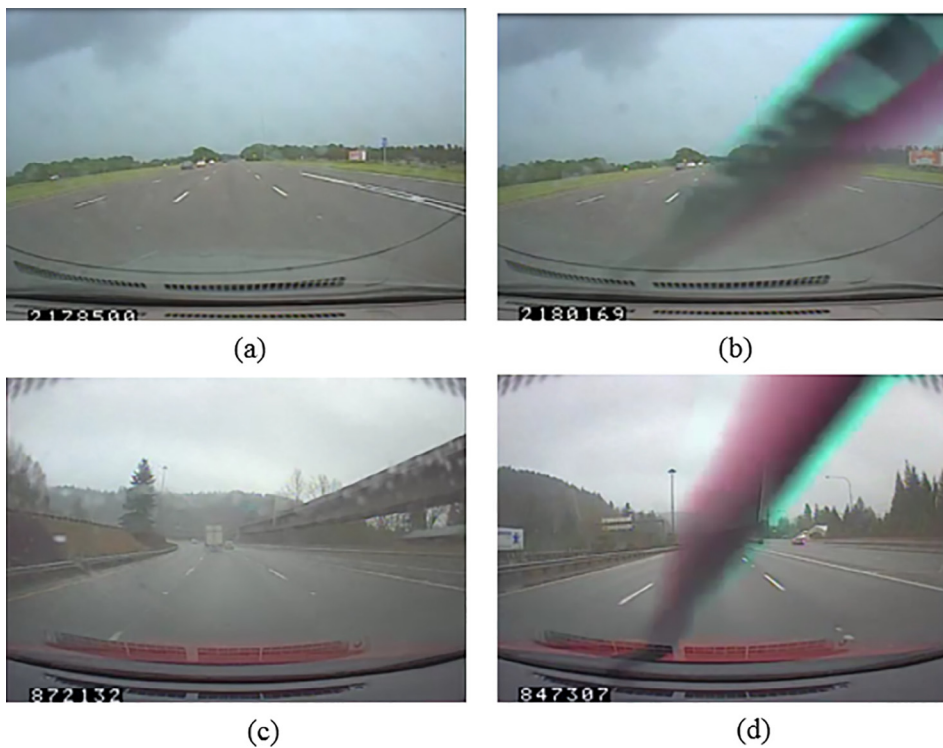


Fig. 3. Light Rain: Road markings are visible, signs are readable, surroundings and traffic can be seen properly to some extent, raindrops on front windshield, slower sweeping wipers.



Fig. 4. Clear Weather: Wipers off, traffic and surroundings are clear and signs are readable.

(Ahmed et al., 2015; Ahmed et al., 2017).

In this study, we classified rain into two categories including heavy rain and light rain, using qualitative-based measures extracted from the NDS videos. Rain was classified based on the visibility of road markings, readability of road signs, roadside surroundings (delineators, guardrail, New Jersey barriers, etc.). The rain was reported as heavy rain if the video observers:

- See high wiper swipe rate recorded by the front view camera.
- Can only see few road markings in front of the NDS vehicle.
- Cannot read the information on the road signs.
- Cannot see the roadside surroundings and traffic ahead clearly.
- Cannot see the horizon.

On the other hand, the rain was classified as a light rain if the video observers:

- See low wiper swipe rate recorded by the front view camera.
- See raindrops on front windshield.
- Can see the road markings and read the information on the road signs.
- Can see the roadside surroundings and traffic ahead.

Figs. 2–4 show some sample pictures of heavy rain and light rain as well as clear weather conditions.

Of the received 2881 trips in rain, 196 trips were randomly selected for further analysis in this study. In addition, 392 matching trips in clear weather conditions (2:1 matching ratio) have been fully processed in this study. It is worth mentioning that the research team requested four matching trips in clear weather for each identified trip in rain to cover different traffic states (level of service); however, the matching of 2:1 (two matching clear for each trip in rain) was the maximum ratio we could get. It is worth mentioning that identifying matching trips was not an easy task considering the fact that matching was defined as the same driver, vehicle, and traversed route. The selected NDS trips involved 141 drivers between 19 and 89 years of age with the majority of the drivers in the young age group (19–29 years old). Gender was mainly balanced among age groups, except for a slight overrepresentation of female drivers between 20 and 24, which follows the same distribution reported by VTTI for all SHRP2 NDS trips.

A total of 4884 one-minute segments – equivalent to nearly 82 h of driving time and 6305 km plus their matching trips in clear weather conditions – have been processed. Once non-freeway segments were removed, 4753 one-minute segments were remained for further analysis. The roadway information provided in the RID and driver demographics provided in the Driver Demographics Questionnaire were used to merge geometric characteristics and driver demographics with each one-minute segment.

## 5. Preliminary analysis

It is of utmost importance to exclude lane change maneuvers from the final dataset, as the focus of this study is to assess driver lane-keeping ability. A criterion for identifying lane changes was proposed considering continuous time series data at 10 HZ. Twenty lane changes were randomly selected in both clear and rainy weather conditions to identify a threshold that differentiates lane changes from lane wandering. The analysis of lane change trajectories revealed that a  $\pm 70$  cm lateral shift in the position of a vehicle can trigger lane changes. However, the research team considered a more conservative threshold ( $\pm 100$  cm) to capture poor lane-keeping behavior with higher accuracy. In other words, lane offset values above 100 cm and below  $-100$  cm were set to flag lane changing events, especially when these events varied to the right and left over a short duration of time. Fig. 5 shows samples of lane changes in clear and rainy weather conditions considering 10 s before and after each lane change. Those segments that had one or more lane changes during a one-minute segment were excluded from the final dataset. More specifically, Fig. 5 represents lane change events and the spikes in this Figure show the occurrences of lane changes. As an example, Fig. 5a represents a lane change

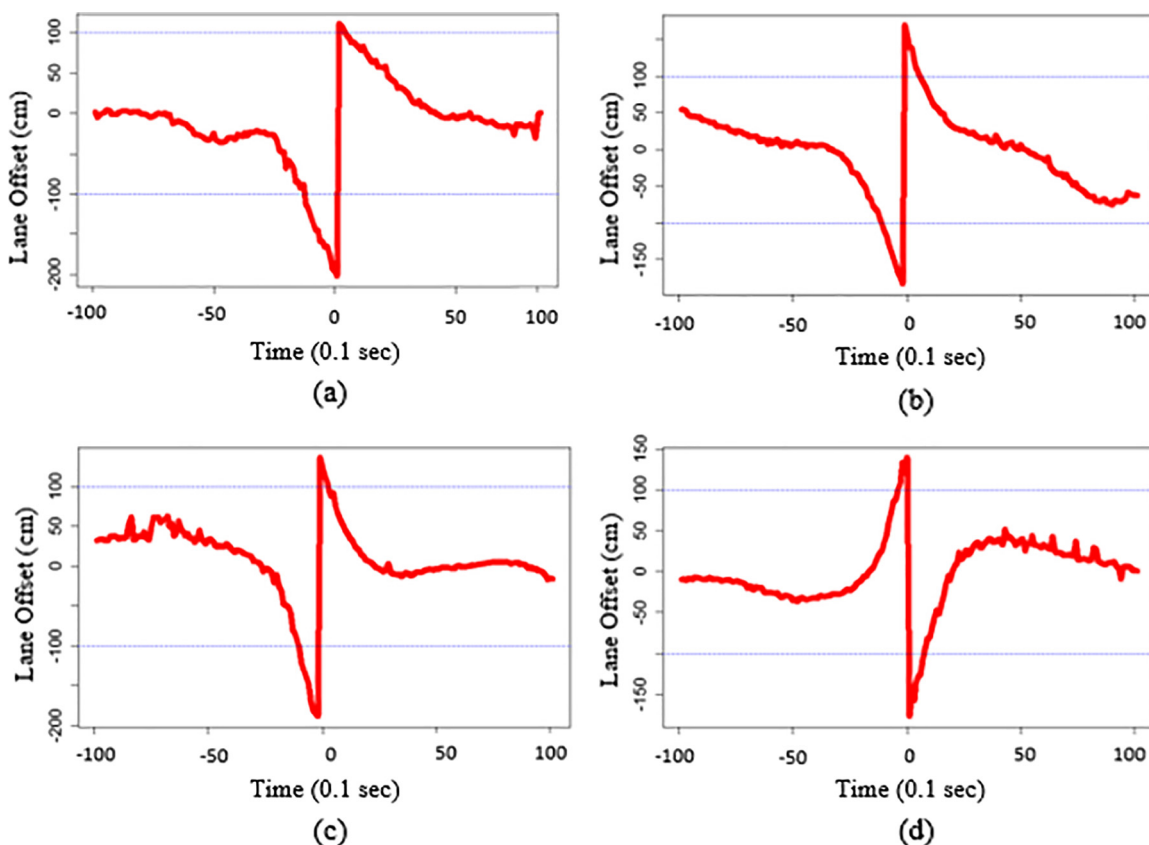


Fig. 5. Identifying a criterion for excluding lane changes (a and b: lane change in clear weather conditions; c and d: lane change in rainy weather conditions).

event 10-s before and 10-s after the lane change occurrence. Therefore, Fig. 5a shows that driver changed his/her lane to the adjacent lane on the left (the negative value of lane offset represents offset to the left of lane center). As can be seen, first, the value of lane offset to the left have been increased (increase in negative value of lane position offset), which shows vehicle movement to the left of lane center. This negative value (vehicle movement to the left of the lane center) increases until we have a jump, which represents that the vehicle is on the far right of the adjacent lane. Then the decrease on lane offset is expected as the vehicle moves toward the center of the target lane.

Table 1 shows preliminary analysis including average, standard deviation, minimum, maximum, and median as well as various statistical tests for two of the most important driver behavioral factors, which are speed and lane offset in heavy rain/clear weather in the free-flow conditions. Cohen's d effect size, which is an indication of the magnitude of the difference between heavy rain and clear weather, is also provided in Table 1. Cohen's d effect size can be interpreted as d = 0.2 small size effect; d = 0.50 medium size effect; and d = 0.80 large size effect (Cohen, 1977).

Both speed and speed reduction percentage ( $Speed\ Reduction\ \% = \frac{Speed - Speed\ Limit}{Speed\ Limit} * 100$ ) have been calculated to provide more insight about driver speed selection behavior in rainy weather conditions. It is worth mentioning that vehicle speed indicated on speedometer collected from vehicle network.

As shown in Table 1, t-test indicated that the average speed in heavy rain under the free-flow traffic conditions was significantly (7.84 km/h) lower than in clear weather and free-flow traffic conditions. In case of light rain, the difference between average speed in light rain and clear weather conditions was smaller than the heavy rain but still the average speed in light rain was lower than the matched clear trips. The variability of speed in both heavy rain and light rain is higher in comparison with clear conditions under free-flow traffic, which could be an indication of increased safety risk (Ahmed and Abdel-Aty, 2013). It is important to consider the free flow speed for variable speed limit (VSL) application because the speed choice here is not affected by the interaction with traffic.

The analysis of lane position offset was also assessed as an indication of driver lane keeping ability. As mentioned earlier, lane position offset can be defined as the distance to the right or left of the center of the lane based on machine vision. The analysis results showed that average and variability of lane offset to the right is significantly higher in clear weather under free flow speed conditions.

**Table 1**  
Preliminary analysis for the NDS instrumented vehicles.

Statistical tests	Heavy rain				Light rain			
	Free-flow traffic				Free-flow traffic			
	Heavy rain		Clear		Light rain		Clear	
	Speed	% Speed reduction	Speed	% Speed reduction	Speed	% Speed reduction	Speed	% Speed reduction
Speed (km/h)								
Average	93.711	-10.588	101.553	-8.368	98.199	-10.412	102.881	-7.992
SD	11.023	8.815	9.186	7.498	13.322	10.515	12.271	8.639
Min.	51.092	-49.458	72.384	-35.171	40.653	-55.646	50.27	-47.595
Max.	128.542	-0.023	130.786	-0.004	130.744	-0.108	137.448	-0.006
Median	93.668	-8.770	100.966	-6.724	99.302	-7.159	103.077	-5.008
t-test	Avg. speed is significantly lower in heavy rain. t(899) = -12.99, P < 0.05; Effect size (Cohen's d) = 0.79				Avg. speed is significantly lower in heavy rain. t(1964) = -9.55, P < 0.05; Effect size (Cohen's d) = 0.37			
F-test	Speed variability is significantly higher in heavy rain. F <sub>483,745</sub> = 1.44, P < 0.05				Speed variability is significantly higher in light rain. F <sub>1053,2094</sub> = 1.18, P < 0.05			
Lane Offset (cm)								
	Right	Left	Right	Left	Right	Left	Right	Left
Average	17.437	-14.124	19.184	-14.889	15.559	-17.950	17.201	-17.595
SD	12.254	11.475	10.956	11.847	11.686	16.856	13.041	15.933
Min.	0.182	-60.166	0.105	-61.924	0.137	-88.470	0.119	-95.098
Max.	47.495	-0.218	58.063	-0.027	57.716	-0.102	61.619	-0.042
Median	16.925	-10.653	18.328	-10.437	16.299	-10.411	14.787	-9.700
t-test	Average lane offset to the right is significantly higher in clear weather. t(365) = 1.7, P < 0.05 Effect size (Cohen's d) = 0.15 No sig. difference in average lane offset to the left t(399) = -0.72, P > 0.05 Effect size (Cohen's d) = 0.065				Average lane offset to the right is significantly higher in clear weather. t(753) = 2.46, P < 0.05 Effect size (Cohen's d) = 0.16 Average lane offset to the left is significantly higher in light rain. t(1914) = 0.51, P < 0.05 Effect size (Cohen's d) = 0.022			
F-test	Lane offset to the right variability is significantly higher in clear weather. F <sub>417,365</sub> = 1.25, P < 0.05 No sig. difference in Lane offset to the left variability F <sub>398,342</sub> = 1.06, P > 0.05				Lane offset to the right variability is significantly higher in clear weather. F <sub>598,335</sub> = 1.22, P < 0.05 Lane offset to the left variability is significantly higher in light rain. F <sub>1475,939</sub> = 0.89, P < 0.05			

## 6. Methodology

Two lane-keeping models were developed using the logistic regression and Multivariate Adaptive Regression Splines (MARS) to better understand factors affecting driver lane-keeping ability in clear and rainy weather conditions. Each model has its own advantages. The main advantages of the MARS model are that MARS is a non-parametric modeling technique without any prior assumption related to the distribution of the data. In addition, MARS is not severely affected by collinearity between variables. Moreover, unlike many other non-parametric models, MARS functions are interpretable and transparent; therefore, it does not have the “black box” problem known for other machine learning techniques (Quirós et al. 2009). Finally, MARS model has the ability to perform analysis on parameter relative importance (Zhang and Goh, 2016). On the other hand, despite the limitations of parametric models such as logistic regression, they are effective in interpreting the marginal effects of various risk factors. The focus of this paper is not to show one model is superior to the other, but it attempts to demonstrate how the two proposed complementary parametric and non-parametric approaches can help researchers to provide better insights into the factors, which may affect driver lane-keeping performance in rainy weather conditions.

### 6.1. Multivariate Adaptive Regression Splines (MARS)

MARS can be defined as a piecewise, multivariate regression that can consider the complex relationships among variables. This model was introduced by (Friedman, 1991). In the MARS model, the space of predictors is divided into multiple knots, and a spline function, which is comprised of several Basis Functions (BFs), is fitted between those knots (Abraham et al., 2001). More clearly, Basis functions are those elements that can be used to fit a MARS model. Each basis function can be a main function or an interaction of different variables. Eq. (1) shows a general form of the MARS model (Haleem et al., 2013; Park and Abdel-Aty, 2015).



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

where  $\hat{y}$  can be defined as the predicted response variable, which is standard deviation of lane position offset (an index to show driver lane-keeping ability) in this study,  $\alpha_0$  is the coefficient of the constant basis function,  $\alpha_m$  is the coefficient of the  $m$ th basis function,  $\beta_m(x)$  is the  $m$ th basis function, and  $M$  is the total number of basis functions in the developed MARS model.

There are two main steps to fit a MARS model. These two steps can be summarized as first, forward-stepwise regression selection and second, backward-stepwise elimination procedure (Osei-Bryson and Ngwenyama, 2014). In the constructive phase, the initial model starts with just a constant, then the model searches for a possible variable-knot combination, and the improvement of the model is measured. The process would be repeated to identify the best variable-knot combination. The search process will continue until reaching the maximum number of basis functions. In the elimination phase, MARS identifies a BF to drop based on residual sum of squares criteria. After refitting the model, another BF is selected to drop based on the same criteria. The process is repeated until all the BFs have been deleted. Finally, the result of the backward-stepwise elimination procedure is a distinctive series of candidate models (Osei-Bryson and Ngwenyama, 2014). The final selection of the model would be based on the generalized cross-validation (GCV) criterion as shown in Eq. (2) (Haleem et al., 2013; Osei-Bryson and Ngwenyama, 2014; Park and Abdel-Aty, 2015):

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

where  $N$  is the number of one-minute segments;  $y_i$  is driver lane-keeping ability for observation  $i$ ; and  $C(M)$  is a complexity penalty function. To be specific, in Eq. (2), the numerator represents the lack of fit on the model with  $M$  basis function and the denominator contains a term to consider a penalty for model complexity, which is  $C(M)$ . This term is related to the number of estimated parameters in the model (Osei-Bryson and Ngwenyama, 2014).

### 6.2. Logistic regression

Logistic regression is a commonly used model in traffic safety and operation studies. Therefore, for brevity, logistic regression modeling is not provided in this paper. For more information please see Agresti (2007).

## 7. Lane keeping model

Both Logistic regression and MARS model have been used to develop the lane-keeping model and investigate factors that might have impact on driver lane-keeping ability. The dependent variable in the model is Standard Deviation of Lane Position Offset (SDLP). Lane Position Offset represents the distance to the left or right of the center of the lane based on machine vision procedure. The explanatory variables are the factors which may have significant influence on the lane-keeping ability, including driver demographics; geometric characteristics; and environmental and traffic characteristics. The SDLP is defined as a binary variable with two levels including SDLP less than 20 cm and SDLP greater than 20 cm. Previous studies used wide range of SDLP (18–50 cm) to investigate the lane-keeping ability (Zhou et al., 2008; Peng et al., 2013; Ghasemzadeh and Ahmed, 2017). However, it was found that the typical value of SDLP is almost 20 cm for normal driving (Zhou et al., 2008). Table 2 is a summary of the different variables used to set lane-keeping models.

## 8. Results and discussion of key factors

### 8.1. Lane keeping model: MARS

Previous studies showed that increasing the number of interactions in MARS may increase the model complexity; therefore, the applicability of the model and interpretability of the results might be decreased (Haleem et al., 2013; Park and Abdel-Aty, 2015; Osei-Bryson and Ngwenyama, 2014). Hence, the maximum order of interactions was defined as two in this study. Table 3 presents the developed MARS model for driver lane-keeping ability.

As can be seen in Table 3, the truncated basis function BF1 indicate that weather condition is not linearly related to the driver lane keeping ability. As an example, there is a knot “3” in the truncated function BF1, which can be simplified as Eq. (3):

$$\text{(Weather in (3))}: \begin{cases} \text{(Weather in (3))}: & \text{if weather cond. is heavy rain} \\ 0: & \text{Otherwise} \end{cases} \tag{3}$$

In order to interpret the interaction terms, both main effect variables and the interaction term should be considered together in the basis function. As an example, the BF13 for the interaction of heavy rain and speed limit, the equation can be written as Eq. (4):

$$0.132 * (\text{Speed limit in (below 60 mph)}) * (\text{Weather in (3 = heavy rain)}) \tag{4}$$

Considering the heavy rain conditions (BF1 = 1), the final coefficient of interaction would be +0.132, which represents worse lane keeping in segments with speed limit below 60 mph and under heavy rain weather conditions. The effects of other interactions can be interpreted in the same manner. The developed MARS model can be expressed as Eq. (5).

**Table 2**  
Data description.

Variable	Description	Type	Source	Definition	Assigned code	Reference level
<b>Response variable:</b>						
SDLP	Standard deviation of lane position offset Distance to the left or right of the center of the lane based on machine vision	Categorical	Naturalistic driving time series data	SDLPO ≤ 20 cm	1	*
				SDLPO > 20 cm	2	
<b>Explanatory variables:</b>						
<i>Environmental factors</i>						
Weather conditions	Predominant weather conditions in 1-min video observation	Categorical	Video Observation	Clear	1	*
				Rain	2	
				Heavy Rain	3	
Speed limit	Predominant posted speed limit in 1-min video observation	Binary	Roadway Information Database (RID)	< =60 mph (median of speed limit: 60) > 60 mph	0 1	
Traffic conditions	Predominant Traffic conditions (LOS) in 1-min video observation	Binary	Video observation	A and B	1	*
				C-F	2	
<i>Demographics</i>						
Gender	The gender the participant identifies with	Binary	Electronic online questionnaire administered during participant in-processing	Male	1	*
				Female	2	
Age	The age group corresponding to the driver's birthdate.	Categorical	Electronic online questionnaire administered during participant in-processing	Young < 25	1	*
				Middle (25–44)	2	
				Old > 44	3	
Education	The participant highest completed level of education	Categorical	Electronic online questionnaire administered during participant in-processing	Below high school diploma	1	*
				Above high school diploma	2	
Driver mileage last year details	The approximate number of miles the participant drove last year	Categorical	Electronic online questionnaire administered during participant in-processing	Less than < =12,000	1	*
				> 12,000	2	
Driving experience	Number of years driving experience	Categorical	Electronic online questionnaire administered during participant in-processing	< 3 years	1	*
				> 3 years	2	
<i>Roadway characteristics</i>						
Curve	Whether the majority of 1-min driving was driven on tangent or curve	Binary	RID	Tangent	1	*
				Curve	2	
Number of lanes	Number of lanes that the majority of the 1-min driving was travelled on	Count	RID	–	–	–

$$Y = 0.103 + 0.028*BF1 + 0.157*BF3 + 0.064*BF5 - 0.073*BF7 + 0.146*BF9 + 0.063*BF11 + 0.132*BF13 - 0.032*BF15 - 0.096*BF21 + 0.045*BF27 \tag{5}$$

In order to understand the total effect of a given explanatory variable on the response variable the following steps should be considered (Osei-Bryson and Ngwenyama, 2014):

- (1) All basis functions that are associated with that specific explanatory variable should be identified. For example, in this study weather condition was the variable of interest, then the relevant BFs would be BF1, BF2, BF3, BF5, BF7, BF9, BF11, BF13, and BF15 (see Table 3).
- (2) The knots of the given explanatory variable should be identified considering the selected BFs (e.g., Weather in [3], which represents heavy rain).
- (3) Considering the knots, set of intervals should be identified. If there is a single knot (e.g. for weather conditions), there are 2 intervals (weather in [3] and weather in [1, 2]).

**Table 3**  
Driver lane keeping model Using MARS.

BF	Basis function	Basis function description	Coefficient	Standard error	P-value
Intercept	Intercept	Constant	0.103	0.017	0.00000
BF1	(Weather in (3))	Main effect	0.028	0.039	0.0001
BF2	(Weather in (2, 1))	Not sig.	Not sig.	Not sig.	Not sig.
BF3	max(0, Number of lanes – 2) * BF1	Interaction	0.157	0.022	0.00000
BF5	(TRAFFIC_CAT in (1)) * BF2	Interaction	0.064	0.013	0.00000
BF7	(Age in (1)) * BF2	Interaction	–0.073	0.014	0.00000
BF9	(Education level in (2)) * BF1	Interaction	0.146	0.033	0.00001
BF11	max(0, Number of lanes – 2) * BF2	Interaction	0.063	0.015	0.00001
BF13	(Speed limit in (below 60 mph)) * BF1	Interaction	0.132	0.033	0.00005
BF15	(Education level in (2)) * BF2	Interaction	–0.032	0.013	0.01401
BF19	max(0, Number of lanes – 3)	Not sig.	Not sig.	Not sig.	Not sig.
BF20	max(0, 3 – Number of lanes)	Not sig.	Not sig.	Not sig.	Not sig.
BF21	(Driving experience in (2)) * BF19	Interaction	–0.096	0.020	0.00000
BF27	(Speed limit in (below 60 mph)) * BF20	Interaction	0.045	0.011	0.00006

- (4) For each specific interval, those basis functions that are associated with that interval should be selected (e.g., BF1, BF3, BF9 and BF13 for weather in [3]).
- (5) The “*impact impression*” for each interval can be calculated by selecting the associated BFs for the given interval from the MARS equation.
- (6) “*Rate of Impression Expression*” can be obtained from the first derivative of the “*impact expression*”
- (7) Considering the “*Rate of Impression Expression*”, the direction of effect can be positive, negative or no effect.

For the (Weather in [3]), the impact impression can be written as Eq. (6):

$$0.028*(\text{Weather in (3)}) + 0.156*\max(0, \text{Number of lanes}-2)*(\text{Weather in (3)}) + 0.146*(\text{Education level in (2)})*(\text{Weather in (3)}) + 0.132*(\text{Speed limit in (below 60 mph)})*(\text{Weather in (3)}) \tag{6}$$

Rate of impact expression can be obtained from the first derivative of the Eq. (6) based on Weather (3):

$$0.028 + 0.156*\max(0, \text{Number of lanes}-2) + 0.146*(\text{Education level in (2)}) + 0.132*(\text{Speed limit in (below 60 mph)}) \tag{7}$$

Given the number of lanes more than 2; education level = above high school diploma; and speed limit = below 60 mph; we have:  $0.028 + 0.156 + 0.146 + 0.132 = 0.462$ . Therefore, as the total effect of heavy rain has a positive sign (0.462), heavy rain increases the SDLP. In other words, drivers are expected to have worse lane keeping in heavy rain conditions. The impact impression of other explanatory variables can be obtained in the same manner.

Table 4 indicates the variable importance in the developed MARS model. As can be seen, weather conditions turned out to be the most important variable affecting lane-keeping ability, such that it is 3 times more important than the second variable. It indicates that weather conditions play a key role in driver lane-keeping ability. This finding is consistent with a previous study that demonstrated the effect of weather conditions on driver behavior in general and lane maintaining in specific (Ghasemzadeh and Ahmed, 2017). The second important variable is speed limit and the third one is traffic conditions. In addition, Table 4 reveals that age, driving experience, and number of lanes are other important factors affecting drive lane-keeping ability.

8.2. Lane keeping model: logistic regression

To confirm the suitability and fitness of the model, the Hosmer–Lemeshow Goodness-of-Fit was used (Hosmer and Lemeshow, 2000). A non-significant Chi square supports that the model adequately fit the data. Table 5 shows the results of the model. The Multi-collinearity was assessed by calculating the Variance Inflation Factor (VIF) for each predictor, which indicates how much the variance of an estimated regression coefficient increases if the predictors are correlated. A VIF between 5 and 10 shows high correlation between predictors and VIF greater than 10 indicates that the regression coefficients are poorly estimated due to multi-collinearity

**Table 4**  
Relative importance of variables.

Variable	Score
Weather cond.	100
Speed limit	28.23
Traffic cond.	26.29
Age	23.59
Driving experience	11.38
Number of lanes	9.32

**Table 5**  
Estimation of logistic regression model for lane keeping behavior.

Parameter	Description		Estimate	Standard error	Wald chi-square	Pr > ChiSq	Odds ratio
Intercept	–	–	–1.58	0.31	25.88	< 0.0001	0.206
Weather	Rain	–	–0.86	0.22	14.53	0.0001	0.425
Weather	Heavy rain	–	0.79	0.25	9.82	0.0017	2.203
Speed limit	> 60 mph	–	–0.27	0.11	5.62	0.0177	0.764
Age	Middle age	–	0.31	0.17	3.58	0.0584	1.37
Gender	Female	–	0.58	0.27	4.72	0.0298	1.785
Traffic cond.	LOS A-B	–	0.25	0.06	20.95	< 0.0001	1.288
Driver mileage last year	> 12,000 miles	–	–0.12	0.06	4.40	0.0358	0.885
Interaction between weather cond. and curve	Light rain	Curves	0.26	0.15	3.00	0.083	1.298
Interaction between weather cond. and traffic cond.	Light rain	LOS A-B	0.32	0.08	15.96	< 0.0001	1.373
Interaction between weather cond. and traffic cond.	Heavy rain	LOS A-B	–0.29	0.09	11.15	0.0008	0.747
Interaction between weather cond. and age	Heavy rain	Older drivers	–0.29	0.16	3.42	0.0645	0.747
Interaction between weather cond. and driver mileage last year	Light rain	> 12,000 miles	–0.19	0.08	6.63	0.01	0.824
Interaction between weather cond. and gender	Heavy rain	Female	0.13	0.08	2.74	0.098	1.141
Interaction between weather cond. and number of lanes	Heavy rain	–	0.21	0.07	10.51	0.0012	1.238
Interaction between weather cond. and speed limits	Light rain	> 60 mph	0.31	0.17	3.40	0.065	1.367
Interaction between weather cond. and speed limits	Heavy rain	> 60 mph	–0.29	0.16	3.10	0.0783	0.749
Interaction between gender and driving experience	Female	< 3 years	0.60	0.27	5.12	0.0236	1.827
Interaction between gender and driver mileage last year	Female	> 12,000 miles	0.20	0.05	18.64	< 0.0001	1.219
Model fit statistics:							
AIC						3905.016	
SC						4118.405	
Log-likelihood at convergence						–4236.771	
Number of observations						4753	

Hosmer and Lemeshow goodness-of-fit test chi square = 10.10678, p = 0.258.

(Marquardt, 1970). The explanatory variables introduced to the model produced VIF values between 1.01 and 1.23, excluding any concerning multi-collinearity. Only statistically significant variables were retained in the final model.

As can be seen in Table 5, heavy rain has a significant effect on driver lane-keeping ability. This means that standard deviation of lane position is more likely to be higher in heavy rain conditions. This may be attributed to the shorter sight distance and low visibility of lane marking in heavy rain conditions. This finding is in agreement with previous studies, showing the negative effect of adverse weather on driver performance in general and lane-keeping ability in particular (Brooks et al., 2011; Hamdar et al., 2016; Ghasemzadeh and Ahmed, 2017; Ahmed and Ghasemzadeh, 2017). To be specific, drivers in heavy rain conditions were 2.2 times more likely than clear weather to have higher SDLP (OR = 2.203).

The results also showed that, maximum-posted speed limits had a significant effect on driver lane-keeping ability. More specifically, the negative coefficient of maximum-posted speed limit represents better lane keeping in higher speed. Better lane keeping in higher speed limits could be because of the fact that drivers pay more attention to the road ahead considering the higher speed. In addition, segments with higher speed limits might have better geometry design and sight distance compared to segments with lower speed limit. The results showed that drivers who were driving on a roadway segment with a posted speed limit greater than 60 mph were almost 24% more likely to have better lane keeping compared to those who were driving on road segments with posted speed limits below 60 mph (OR = 0.784).

In addition, driver age was a significant factor in the lane-keeping model. To be specific, middle-aged drivers were 1.4 times more likely to have worse lane-keeping ability in comparison with young drivers.

Traffic conditions were found to have a significant effect on lane-keeping ability. This is not surprising as by increasing traffic congestion, drivers are limited and do not have enough space for maneuvers including swerve and lane change. In other words, drivers are forced to have a better lane keeping in congested traffic. The results provided in Table 5 reveals that drivers who drove in a free flow condition were 1.3 times more likely to have higher SDLP (worse lane keeping) in comparison with those who were driving in congested traffic.

Driver mileage last year was found to be a statistically significant factor that affects driver lane-keeping ability. More specifically, those drivers that drove more than 12,000 miles last year were more likely to have a better lane-keeping in comparison with drivers that drove less than < 12,000 miles. This variable can be an index for driving experience; therefore, the result shows the significant effect of driving experience on lane-keeping ability.

Interaction between variables was also examined in the logistic regression model. The interaction between weather, curve, traffic conditions, age, driver millage last year, number of lanes, and speed limits came out to be significant interactions in the final model. Considering the interaction terms, the results indicated that older drivers at an affected visibility and wet surface conditions caused by heavy rain were 25% more likely to have better lane keeping compared to young drivers. In addition, the interaction between weather conditions and curve revealed that the drivers were 1.3 times more likely to have worse lane keeping on curves and under rainy weather conditions.



## 9. Conclusions

This study provided more insight into driver lane keeping performance under rainy weather conditions by considering 196 trips in rain and additional 392 matching trips in clear weather conditions, representing 141 drivers in a naturalistic driving setting and using parametric and non-parametric modeling techniques. This paper overcomes a significant limitation highlighted in a previous study of using wiper status in the NDS data to identify trips in adverse weather conditions (Ghasemzadeh and Ahmed, 2017). Wiper status was found to be missing in the majority of the SHRP2 NDS time-series data. Therefore, in this paper, additional two complementary methods were developed to effectively identify more trips in adverse weather conditions utilizing external data sources such as NCDL weather database and police crash reports. In addition, this study sampled the data at 1-min chunks to ensure homogeneity compared to the whole trip used as the analysis unit in a previous study (Ghasemzadeh and Ahmed, 2017). Due to the high variability in environmental and traffic conditions found in the NDS trips, this approach deemed to enhance the accuracy of the analyses provided in this paper.

Preliminary descriptive analysis showed that both heavy rain and light rain had a higher standard deviation and a lower average of speed compared to the matched trips in clear weather. The number of lane changes is higher in clear weather; however, lane wandering is higher in heavy rain conditions.

The results from the MARS model revealed that weather condition was not linearly associated with the SDLP. In addition, among all the explanatory variables considered in the model, weather conditions turned out to be the most important variable affecting lane-keeping performance. On the other hand, results from the logistic regression indicated that drivers in heavy rain conditions were 2.2 times more likely than clear weather to have higher SDLP (i.e., poor lane keeping).

As the importance of analyzing driver behavior in real-time at a trajectory level is becoming more important for various tasks in transportation engineering, the NDS data may help in not only providing a reliable source of trajectory-level driver-behavioral and vehicle information, but also in developing driving models that could be applied to different areas, including but not limited to safety analysis considering microscopic individual driver data; calibrating driver behavior models, specifically in different weather and traffic conditions; and developing control logics for Advanced Driving Assistance Systems (ADAS), and Connected and Automated Vehicles (CAV).

The results from this paper may shed some light on driver lane keeping behavior at a trajectory level. Better understanding of driver lane keeping behavior might help in developing better Lane Departure Warning (LDW) systems. More specifically, the limitations of these systems during inclement weather conditions may be addressed using the NDS data, which can provide more accurate and practical threshold for LDW systems to reduce false warnings. Finally, using the microscopic driver behavior and performance data might be more accurate and practical compared to commonly used aggregated traffic data (i.e., speed, volume and occupancy) in updating micro-simulation models.

## Acknowledgment

This work was conducted through the second Strategic Highway Research Program (SHRP2), which is administrated by the Transportation Research Board (TRB) of the National Academies of Sciences, Engineering, and Medicine, and it was sponsored by the Federal Highway Administration (FHWA) in cooperation with the American Association of State Highway and Transportation Officials (AASHTO).

## References

- Abdel-Aty, M.A., et al., 2012. Real-time prediction of visibility related crashes. *Transport. Res. Part C: Emerg. Technol.* 24, 288–298.
- Abraham, A., Steinberg, D., Philip, N.S., 2001. Rainfall forecasting using soft computing models and multivariate adaptive regression splines. *IEEE SMC Trans., Spec. Iss. Fusion Soft Comput. Hard Comput. Indust. Appl.* 1, 1–6.
- Agresti, A., 2007. *An Introduction to Categorical Data Analysis*, second ed. John Wiley & Sons Inc., Publication.
- Ahmed, M., Abdel-Aty, M., 2013. A data fusion framework for real-time risk assessment on freeways. *Transport. Res. Part C: Emerg. Technol.* 26, 203–213. <http://dx.doi.org/10.1016/j.trc.2012.09.002>.
- Ahmed, M.M., et al., 2015. Driver Performance and Behavior in Adverse Weather Conditions: An Investigation Using the SHRP2 Naturalistic Driving Study Data—Phase 1, doi:10.13140/RG.2.2.24061.05602 < [http://ntl.bts.gov/lib/59000/59200/59242/1608F\\_RS04215\\_Driver\\_Performance\\_and\\_Behavior\\_in\\_Adverse\\_Weather\\_Conditions.pdf](http://ntl.bts.gov/lib/59000/59200/59242/1608F_RS04215_Driver_Performance_and_Behavior_in_Adverse_Weather_Conditions.pdf) > .
- Ahmed, M.M., et al., 2014. Real-time assessment of fog-related crashes using airport weather data: a feasibility analysis. *Acc. Anal. Prevent.* 72, 309–317. <http://dx.doi.org/10.1016/j.aap.2014.07.004>.
- Ahmed, Mohamed, Young, Rhonda K., Ghasemzadeh Ali, 2017. Implementation of SHRP2 Results within the Wyoming Connected Vehicle Variable Speed Limit System: Phase 2 Early Findings Report and Phase 3 Proposal. FHWA report. <http://shrp2.transportation.org/Documents/Safety/05-Phase2EarlyResearchFindings-WY-Weather.pdf>.
- Ahmed, M.M., Ghasemzadeh, A., 2017. Exploring the impacts of adverse weather conditions on speed and headway behaviors using the SHRP2 naturalistic driving study data. In: 96th Transportation Research Board Annual Meeting, 2017. Washington, DC.
- Allen, R.W., Parseghian, Z., Stein, A.C., 1996. A driving simulator study of the performance effects of low blood alcohol concentration. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 40(18), pp. 943–946.
- Andrescu, M.P., Frost, D.B., 1998. Weather and traffic accidents in Montreal, Canada. *Clim. Res.* 9 (3), 225–230.
- Brodsky, H., Hakkert, A.S., 1988. Risk of a road accident in rainy weather. *Acc. Anal. Prevent.* 20 (3), 161–176.
- Brooks, J.O., et al., 2011. Speed choice and driving performance in simulated foggy conditions. *Acc. Anal. Prevent.* 43 (3), 698–705.
- Chisholm, S.L., Caird, J.K., Lockhart, J., 2008. The effects of practice with MP3 players on driving performance. *Acc. Anal. Prevent.* 40 (2), 704–713.
- Cohen, J., 1988. *Statistical power analysis for the behavioral sciences*, second ed. Erlbaum Associates, Hillsdale. N.J.978-0-12-179060-8.
- Collet, C., Guillot, A., Petit, C., 2010. Phoning while driving I: a review of epidemiological, psychological, behavioural and physiological studies. *Ergonomics* 53 (5), 589–601.

- Dagan, E., et al., 2004. Forward collision warning with a single camera. In: Intelligent Vehicles Symposium, 2004 IEEE. IEEE, Berlin, pp. 37–42.
- Eboli, L., Mazzulla, G., Pungillo, G., 2017a. How to define the accident risk level of car drivers by combining objective and subjective measures of driving style. *Transport. Res. Part F: Traff. Psychol. Behav.* 49, 29–38.
- Eboli, L., Mazzulla, G., Pungillo, G., 2017b. The influence of physical and emotional factors on driving style of car drivers: a survey design. *Travel Behav. Soc.* 7, 43–51.
- Friedman, J.H., 1991. Multivariate adaptive regression splines. *Ann. Statist.* 1–67.
- Ghasemzadeh, A., Ahmed, M.M., 2017. Drivers' lane-keeping ability in heavy rain: preliminary investigation using SHRP 2 naturalistic driving study data. *Transport. Res. Record: J. Transport. Res. Board* 2663, 99–108.
- Haleem, K., Gan, A., Lu, J., 2013. Using multivariate adaptive regression splines (MARS) to develop crash modification factors for urban freeway interchange influence areas. *Acc. Anal. Prevent.* 55, 12–21.
- Hamdar, S.H., Qin, L., Talebpour, A., 2016. Weather and road geometry impact on longitudinal driving behavior: exploratory analysis using an empirically supported acceleration modeling framework. *Transport. Res. Part C: Emerg. Technol.* 67, 193–213.
- Hamilton, B.A., 2016. Ten-Year Averages from 2005 to 2014. Retrieved from Road Weather Management Program - NHTSA: [http://www.ops.fhwa.dot.gov/weather/q1\\_roadimpact.htm](http://www.ops.fhwa.dot.gov/weather/q1_roadimpact.htm).
- Hankey, J.M., Perez, M.A., McClafferty, J.A., 2013. Description of the SHRP2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets. Task Report, The Strategic Highway Research Program 2 Transportation Research Board of The National Academies, 53(9), 1689–1699.
- Harbluk, J.L., Noy, Y.I., Eizenman, M., 2002. The Impact of Cognitive Distraction on Driver Visual Behavior and Vehicle Control. Transport Canada, Ottawa.
- Hassan, H.M., Abdel-Aty, M.A., 2011. Analysis of drivers' behavior under reduced visibility conditions using a structural equation modeling approach. *Transport. Res. Part F: Traff. Psychol. Behav.* 14 (6), 614–625.
- Hassan, H.M., Abdel-Aty, M.A., 2013. Predicting reduced visibility related crashes on freeways using real-time traffic flow data. *J. Safety Res.* 45.
- Hosmer, D.W., Lemeshow, S., 2000. *Applied Logistic Regression* (2nd ed. Wiley-Interscience Publication, New York W. S. in P. and Statistics).
- Marquardt, D.W., 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics* 12 (3), 591–612.
- McLaughlin, S.B., Hankey, J.M., 2015. Naturalistic Driving Study: Linking the Study Data to the Roadway Information Database. SHRP 2 Report S2-S31-RW-3. Transportation Research Board of the National Academies, Washington, D.C.
- National Transportation Safety Board, 1980. *Fatal Highway Accidents on Wet Pavement—The Magnitude, Location, and Characteristics*. National Technical Information Service, Springfield, HTSB-HSS-8.
- Ni, R., Kang, J.J., Andersen, G.J., 2010. Age-related declines in car following performance under simulated fog conditions. *Acc. Anal. Prevent.* 42 (3), 818–826.
- Osei-Bryson, K.-M., Ngwenyama, O., 2014. *Advances in Research Methods for Information Systems Research*. Springer.
- Park, J., Abdel-Aty, M., 2015. Assessing the safety effects of multiple roadside treatments using parametric and nonparametric approaches. *Acc. Anal. Prevent.* 83, 203–213.
- Peng, Y., Boyle, L.N., Hallmark, S.L., 2013. Driver's lane keeping ability with eyes off road: Insights from a naturalistic study (November 2015). *Acc. Anal. Prevent.* 50, 628–634. <http://dx.doi.org/10.1016/j.aap.2012.06.013>.
- Pomerleau, D., Everson, J., 1999. Run-Off-Road Collision Avoidance Using IVHS Countermeasures. National Highway Traffic Safety Administration Washington, DC.
- Quirós, E., Felicísimo, Á.M., Cuartero, A., 2009. Testing multivariate adaptive regression splines (MARS) as a method of land cover classification of TERRA-ASTER satellite images. *Sensors* 9 (11), 9011–9028.
- Reed-Jones, J., Trick, L.M., Matthews, M., 2008. Testing assumptions implicit in the use of the 15-second rule as an early predictor of whether an in-vehicle device produces unacceptable levels of distraction. *Acc. Anal. Prevent.* 40 (2), 628–634.
- Smadi, O., et al., 2014. Naturalistic Driving Study: Development of the Roadway Information Database, SHRP 2 Report S2-S04A-RW-1. Transportation Research Board of the National Academies, Washington, D.C.
- Suryanarayanan, S., Tomizuka, M., 2007. Appropriate sensor placement for fault-tolerant lane-keeping control of automated vehicles. *IEEE/ASME Trans. Mechatron.* 12 (4), 465–471.
- Suzdaleva, E., Nagy, I., 2018. An online estimation of driving style using data-dependent pointer model. *Transport. Res. Part C: Emerg. Technol.* 86, 23–36.
- Wang, L., et al., 2015. Real-time crash prediction for expressway weaving segments. *Transport. Res. Part C: Emerg. Technol.* 61, 1–10.
- Wu, Y., et al., 2018. Effects of real-time warning systems on driving under fog conditions using an empirically supported speed choice modeling framework. *Transport. Res. Part C: Emerg. Technol.* 86, 97–110.
- Wu, Y., Abdel-Aty, M., Lee, J., 2017. Crash risk analysis during fog conditions using realtime traffic data. *Acci. Anal.* 114, 4–11. <http://dx.doi.org/10.1016/j.aap.2017.05.004>.
- Xu, L.H., Hu, S.G., Luo, Q., 2015. A new lane departure warning algorithm considering the driver's behavior characteristics. *Math. Prob. Eng.* 2015.
- Yu, R., et al., 2014. Utilizing microscopic traffic and weather data to analyze real-time crash patterns in the context of active traffic management. *IEEE Trans. Intell. Transport. Syst.* 15 (1), 205–213.
- Zhang, W., Goh, A.T.C., 2016. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geosci. Front.* 7 (1), 45–52.
- Zhou, J., Peng, H., Gordon, T.J., 2008. Characterization of the lateral control performance by human drivers on highways. *SAE Int. J. Pass. Cars-Mech. Syst.* 1, 450–458 (2008-01-0561).