5-2018

Understanding Micro-Level Lane Change and Lane Keeping Driving Decisions: Harnessing Big Data Streams from Instrumented Vehicles

Meng Zhang
University of Tennessee, mzhang27@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Meng Zhang entitled "Understanding Micro-Level Lane Change and Lane Keeping Driving Decisions: Harnessing Big Data Streams from Instrumented Vehicles." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Civil Engineering.

Asad J. Khattak, Major Professor

We have read this dissertation and recommend its acceptance:

Candace E. Brakewood, Subhadeep Chakraborty, Wenjun Zhou

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Understanding Micro-Level Lane Change and Lane Keeping Driving Decisions: Harnessing Big Data Streams from Instrumented Vehicles

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Meng Zhang

May 2018

This dissertation is dedicated to my parents, Gongzhi Zhang and Rongmei Zhou.

ACKNOWLEDGEMENTS

I would like to express my special thanks to the many people who generously contributed to the work presented in this dissertation during my graduation education.

First and foremost, I would like to thank my major advisor Dr. Asad J. Khattak who guided, supported and encouraged me during my doctoral study. The Ph.D. study is a wonderful experience and I thank Dr. Khattak heartily. He not only provides me sufficient guidance but also shares the tremendous academic experience and knowledge with me which will be useful in my whole life.

Second, I would like to thank my committee members, Dr. Candace Brakewood, Dr. Subhadeep Chakraborty and Dr. Wenjun Zhou. Dr. Wenjun Zhou also served as my statistics advisor. They provide sufficient instruction on my dissertation. In addition, my work also gets support from other faculty members at University of Tennessee, Dr. Han Lee, Dr. Christopher Cherry, and Dr. David Clarke. I am deeply grateful to them all.

Last but not least, I want to thank all the other transportation students who have shared in my experiences: Ziwen Ling, Xiaobing li, Yuandong Liu, Zhihua Zhang, Behram Wali, Abdul Mussah, Ramin Arvin, Mohsen Kamrani, Ali Boggs, Yang Zhang, Bumjoon Bae, Kay Boakye, and many others. They are great friends and I have very good memories of my time with them.

ABSTRACT

It is important to get a deeper understanding of instantaneous driving behaviors, especially aggressive and extreme driving behaviors such as hard acceleration, as they endanger traffic efficiency and safety by creating unstable flows and dangerous situations. The aim of the dissertation is to understand micro-level instantaneous driving decisions related to lateral movements such as lane change or lane keeping events on various roadway types. The impacts of these movements are fundamental to microscopic traffic flow and safety. Sufficient geo-referenced data collected from connected vehicles enables analysis of these driving decisions. The "Big Data" cover vehicle trajectories, reported at 10 Hz frequency, and driving situations, which make it possible to establish a framework.

The dissertation conducts several key analyses by applying advanced statistical modeling and data mining techniques. First, the dissertation proposes an innovative methodology for identifying normal and extreme lane change events by analyzing the lane-based vehicle positions, e.g., sharp changes in distance of vehicle centerline relative to the lane boundaries, and vehicle motions captured by the distributions of instantaneous lateral acceleration and speed. Second, since surrounding driving behavior influences instantaneous lane keeping behaviors, the dissertation investigates correlations between different driving situations and lateral shifting volatility, which quantifies the variability in instantaneous lateral displacements. Third, the dissertation analyzes the "Gossip effect" which captures the peer influence of surrounding vehicles on the instantaneous driving decisions of subject vehicles at micro-level. Lastly, the dissertation explores correlations between lane change crash propensity or injury severity and driving volatility, which quantifies the fluctuation variability in instantaneous driving decisions.

The research findings contribute to the ongoing theoretical and policy debates regarding the effects of instantaneous driving movements. The main contributions of this dissertation are: 1) Quantification of instantaneous driving decisions with regard to two aspects: vehicle motions (e.g., lateral and longitudinal acceleration, and vehicle speed) and lateral displacement; 2) Extraction of critical information embedded in large-scale trajectory data; and 3) An understanding of the correlations between lane change outcomes and instantaneous lateral driving decisions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

In 2015, the lane change crashes account for 4.6% (451,000) of all reported single and two-vehicle crashes that occurred in the United States. Resulting from these crashes were 678 deaths, representing 1.6% of the fatalities in 2015. Although such crashes do not account for a sizable portion of all roadway crashes, the decrease in such crashes can still have substantial benefits regarding social cost. Figure 1.1 shows the examples of lane change crashes.

Previous studies have shown evidence that a lane change crash is correlated with various factors, such as driving and vehicle factors [1-8]. Variability in instantaneous driving decisions could be the contributor to unsafe events. Since a lane change or lane keeping event is an operation that a driver may show high variation in instantaneous driving decisions, i.e., abrupt acceleration or hard braking, it is very important to get an in-depth understanding of instantaneous lateral driving behaviors, especially aggressive or extreme driving behaviors. Sufficient geo-referenced data embedded in connected vehicles enable the analysis.

Figure 1. 1 Examples of lane change related crashes

The dissertation aims to establish a framework to get an in-depth understanding of instantaneous lateral driving decisions using sufficient geo-referenced trajectories data collected from connected vehicles. The dissertation proposes a way to extract key information from public data set for conducting driving behavior analysis. Six main research questions are explored in this dissertation are:

1) How to take advantage of massive transportation data?
2) How to understand and measure instantaneous driving decisions from two aspects: vehicle motion and lateral displacement?
3) How to identify normal and extreme lane change events using massively connected vehicle data?
4) How the surrounding vehicles influence the instantaneous driving decisions of the subject vehicle?
5) What are the correlates of lateral shifting volatility which quantifies the variability in instantaneous lateral displacement?
6) What are the correlates of lane change crash propensity with driving volatility which quantifies the fluctuations in instantaneous driving decisions?

The results indicate different potential applications, including adding driving assistance functions to current onboard driving assistance system to help drivers to make informed driving decisions, updating current traveler information system, helping the vehicle and accessory design, and providing insights to transportation managers and policy makers regarding safety outcome.

Two major data sources are used for analysis: 1) Safety Pilot Model Deployment Data (SPMD), and 2) SHRP 2 Naturalistic Driving Study (NDS) Data. While the dissertation focuses on micro-level instantaneous driving decisions, the key extracted variables will be vehicle speed, lateral displacement, longitudinal and lateral acceleration.

This dissertation contains six parts. Following this chapter, the second chapter quantifies driving volatility in instantaneous lateral driving decisions and proposes an innovative methodology to identify extreme lane change maneuvers. The third chapter proposes a measurement called lateral shifting volatility to quantify the variability in instantaneous lateral displacement and the correlates of shifting volatility are explored. The fourth chapter analyzes the “Gossip effect” which captures the peer influence of surrounding vehicles on the instantaneous driving decisions of subject vehicles at micro-level. The fifth chapter continues to investigate the effects of instantaneous driving decisions on the occurrence of a lane change crash, which is under-explored in previous studies. With the micro changes of the instantaneous driving decision, the dissertation examines relations between safety outcome with driving volatility which quantifies variability in instantaneous driving decisions. The last chapter summarizes the key conclusions of the dissertation. A wide conceptual framework is developed. Figure 1.2 shows the detailed information of conceptual framework. The framework emphasizes the analysis of lane change identifications and distributions of instantaneous lateral driving decisions. The main contributions of the dissertation are: 1) Quantification of instantaneous driving decisions with regard to two aspects: vehicle motions (e.g., lateral acceleration and vehicle speed) and lateral displacement; 2) Extraction of critical information embedded in large-scale trajectory data; and 3) An understanding of the correlations between lane change outcomes and instantaneous lateral driving decisions.

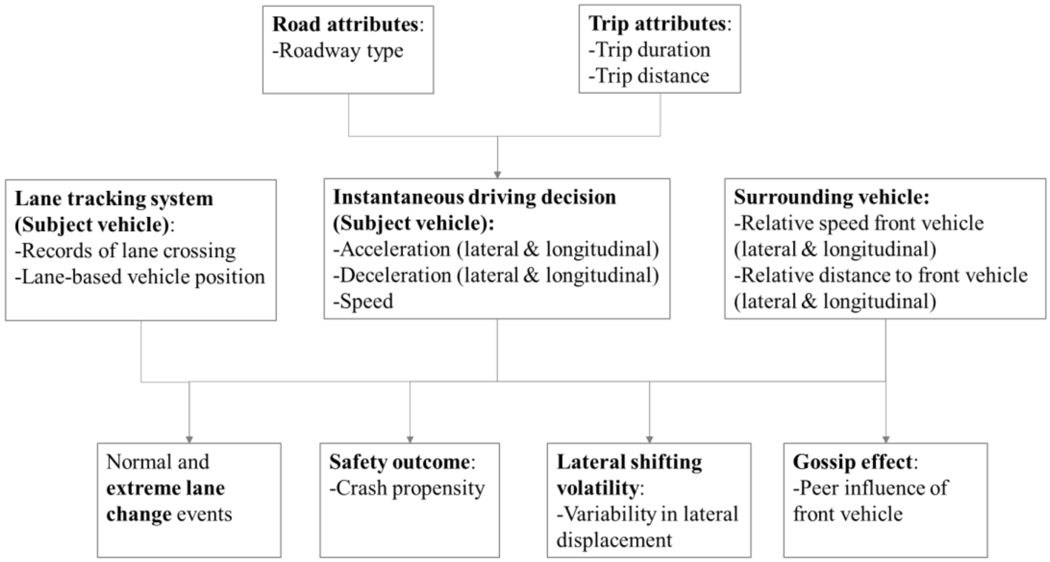


Figure 1. 2 Conceptual framework

CHAPTER 2 IDENTIFYING AND ANALYZING EXTREME LANE CHANGE EVENTS USING BASIC SAFETY MESSAGES IN A CONNECTED VEHICLE ENVIRONMENT

This chapter presents a modified version of a research paper by Meng Zhang and Asad J. Khattak. The paper was presented (TRB 18-04734) at the 97th Annual Meeting of Transportation Research Board in Washington, D.C., in January 2018. And this chapter was submitted to publication review at Journal of Transportation Research Part C: Emerging Technologies.

ABSTRACT

Traffic congestion and safety are challenging problems in the United States and cost an estimated one trillion dollars annually. The United States can potentially reduce dangerous situations and unstable flows caused by aggressive or extreme behaviors through a deeper understanding of driving behaviors and extracting useful information from emerging connected vehicle data. Because lane changes are fundamental maneuvers for traffic flow and safety, this study focuses on microscopic instantaneous driver-level decisions in situations where drivers make lane change maneuvers on various roadway types, especially extreme lane change events. The study analyzes a sub-sample of 1,940,678 Basic Safety Messages (BSMs) recorded from 192 randomly-selected trips (10 minutes or longer) from 64 drivers. The BSMs come from connected vehicles participating in the Safety Pilot Model Deployment program in Michigan. Since BSMs describe vehicle operation and performance, lane changes are identified from multiple criteria including vehicle position (i.e., a sharp change in distance between a vehicle's centerline and the lane boundaries) and lane crossings recorded by onboard units (i.e., when a vehicle crosses a lane marker). Extreme lane change events were then identified as those where lateral acceleration exceeds the 95th percentile threshold between the initiation and the end of the lane change maneuver. A total of 654 lane changes and 128 extreme lane changes were identified in the data. On average, the test vehicles generated 3.4 lane changes (0.67 extreme lane changes) every 20

minutes. Modeling results show that subject drivers are likely to make more lane changes if an object is present in the travel path or the relative speed vis-a-vis the front vehicle is low. Based on the analysis of data, connected vehicle technologies can generate early warnings to help drivers make more informed driving decisions that avoid potential risks in extreme lane changes.

2.1 INTRODUCTION

Traffic congestion and safety are social concerns as they result in enormous economic and social costs annually [9]. A deeper understanding of instantaneous driving behaviors, especially aggressive or extreme driving behaviors (e.g., hard accelerations or fast lane changes), is critical as they endanger occupants of vehicles by creating dangerous situations and unstable flows. Sufficient geo-referenced data embedded in connected vehicles enable the analysis. As the impact of the lane change is fundamental to microscopic traffic flow and safety, the aim of this study is to understand and model normal and extreme lane change behaviors, which can form the basis for generating alerts and warnings that can reduce the impacts of such behaviors. Specifically, this study focuses on microscopic driver-level instantaneous decisions regarding situations where drivers make extreme lane change maneuvers on various roadway types.

This study proposes an innovative methodology to identify extreme lane change events using Basic Safety Messages (BSMs) data sent, at a frequency of 10 Hz, by participating vehicles and received by roadside equipment in the Safety Pilot Model Deployment (SPMD) program in Ann Arbor, Michigan. As BSMs provide sufficient temporal and spatial resolution of lane-based vehicle position (e.g., distance of vehicle centerline relative to left and right boundary of travel lane), onboard device records of lane crossing (e.g., a vehicle is meeting and crossing the lane marker) and motion (e.g., speed and acceleration), it is possible to identify lane change maneuvers and harness useful information about extreme lane change events. Since some lane change maneuvers are relative safe and which might not need additional warning or control assistance, this study is trying to extract critical information of extreme lane change maneuvers embedded in BSMs. Therefore, in real driving environments, alerts and advanced warnings of extreme lane change events could help drivers make informed driving decisions to avoid hazards generated by vehicles or driving environments [10-12], through the applications of vehicles-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) technologies. In summary, the objectives of this study are to:

1) Identify lane change maneuvers based on multiple criteria, such as sharp change in vehicle distance (e.g., from zero to lane width) relative to the boundary of travel lane might be an indicator of a lane change.

2) Quantify extreme lateral driving behaviors (e.g., hard accelerations) by establishing varying thresholds of lateral acceleration under different speed ranges.

3) Recognize extreme lane change events, which are those where lateral acceleration between consecutive 0.5 time stamps exceeds the 95th percentile threshold at the initiation and the end of the lane change maneuver. These extreme events form the basis of generating warnings or control assists provided to drivers achieving safer lane change under connected vehicles; and

4) Explore the correlates of lane change events. For this purpose, information of driving environment, such as relative distance or speed to front vehicles, is extracted from the data for modeling relationships.

2.2 LITERATURE REVIEW

Previous studies have developed and implemented algorithms to identify lane changes based on different data sources, including traffic simulation, driving simulator, instrumented vehicles, and naturalistic traffic recordings. As different input variables reflecting patterns of lane change were available in diverse data sources, the methods of lane change identification vary widely [13]. These key input variables include heading angle [14, 15], path-curvature [14], yaw rate [16, 17], lane index [18, 19], vehicle lateral position [15, 20-22], steering wheel angle [21, 23], image processing technologies [24], and onboard device records of lane crossing [25].

Bogard and Fancher proposed two methods to identify lane change events using GPS data and path-curvature data [14]. They noticed heading angles collected from GPS data can be one indicator of the lane change event. They proposed that sharp changes in angles are due to lane change while smooth changes in heading angles are due to curvatures. Besides heading angle, path-curvature data also reports vehicle yaw acceleration, which can be used for lane change identification. They calculated the heading corners and fitted reference line between heading corners and calculating the difference between the heading angle peak and the reference. A lane change event is identified if the calculated values exceed the defined thresholds. Notice the noisy-sine-wave-like yaw rate signal during a lane change, Miller and Srinivasan identified lane change events of heavy trucks based on yaw rate [16].

A lane change event can be regarded as a function of the characteristics of origin and target lane. Knoop et al. identified lane change events based on the loop detectors placed on each lane of a three-lane freeway about 100 meters apart [19]. Since a vehicle can be identified repeatedly from one detector to the next detector, a lane change event will be recognized if a vehicle is re-identified at a downstream detector in another lane. But this method is useful under uncongested traffic conditions where the vehicle speed is high.

Vehicle trajectory data obtained from naturalistic traffic recording can be used for lane change identification. Thiemann at al. proposed a smooth algorithm to identify lane change events using NGSIM data [18]. The critical variable used in the analysis is the lane index that the vehicle is currently occupying. A lane change event is identified when the lane index is found to change between two consecutive time points. Similarly, R Chen at al. identified lane change events based on the lane change signal recorded by the onboard lane tracking system [25]. A lane change event is triggered when the vehicle center line meets and crosses the lane boundary. The onboard device also reports the confidence level of the lane tracking system for correct distance evaluation.

If road geometry information is readily available, one can easily identify lane change events by comparing a single trajectory with the existing road geometry. Xuan and Coifman established a reference trajectory to present roadway geometry using vehicle trajectory information collected from DGPS (Different Global Positioning System) [20, 22]. They proposed that a sinusoidal wave showed in the mean of lateral distance to reference trajectory indicating a lane change. Table 2.1 summarizes key input variables and identification methods used for the lane change.

While previous studies have developed methodologies to identify lane changes, the value of data embedded in the connected vehicle has not been fully harnessed, especially for extreme lane change identification and analysis. Although roadside-based warnings, such as warnings of lane merge or lane division at a fixed point (e.g., ½ mile before an Exit), can be given to drivers for safer driving, the fixed warning points cannot capture the complexity of drivers' lane change behaviors during an entire trip. Given sufficient geo-referenced data collected from connected vehicles, it is possible to identify and analyze extreme lane change events in real-life driving environments and develop the basis for providing instantaneous feedback about extreme lane change behaviors, so they can avoid future high-risk lane change situations.

Table 2. 1 Key input variables for lane change identification used in selected studies

Author	Data source	Key input variables and Identification methods
Bogard and Fancher [14] / 1999	Instrumented vehicle	1) GPS data: analyzing figure of heading angle vs. time → sharp changes in heading angle due to lane change; 2) Path-curvature: heading angle, yaw acceleration
Miller and Srinivasan [16] / 2005	Instrumented vehicle	Yaw rate → a sine-wave in yaw rate indicating a lane change
Thiemann et al. [18] / 2008	Naturalistic driving recording	Vehicle width, lane index and vehicle position → lane index is found to change between two continuous time stamps
Knoop et al. [19] / 2012	Naturalistic driving recording	Loop detectors record time, lane index, vehicle speed, vehicle length → a vehicle was re-identified at a downstream detector in another lane, indicating a lane change
Xuan and Coifman [20, 22]/ 2006,2012	Instrumented vehicle	Vehicle lateral position → mean of lateral distance to established reference trajectory shows a sinusoidal wave
Salvucci et al. [26] / 2002	Driving simulator	Participants' verbal protocol and experimenter's judgment
R Chen at al. [25] / 2015	Naturalistic driving recording	Records of lane crossing → vehicle centerline meets lane marker as vehicle crosses the lane
Wang and Coifman [24] / 2007	Naturalistic driving recording	Employing Vehicle Re-identification (VRI) image processing technologies

2.3 METHODOLOGY

2.3.1 Data source

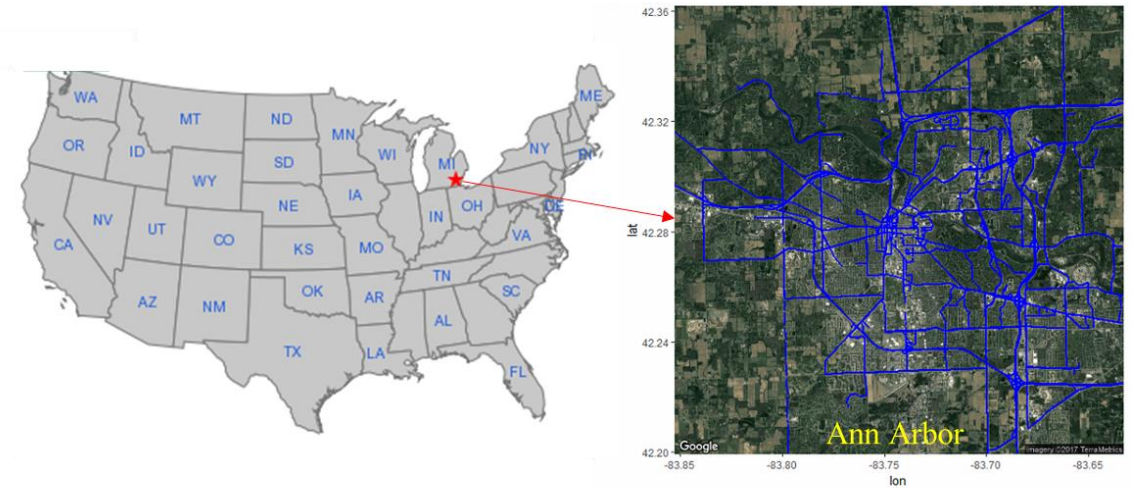
The data used in this study are BSMs sent by participating vehicles and received by roadside equipment in the Safety Pilot Model Deployment (SPMD) program in Ann Arbor, Michigan. The field test contains 75 miles of instrumented roadway installed with approximately 26 roadside equipment [27], which enables the communication with appropriately equipped vehicles. This study uses BSMs archived in Driving Dataset for analysis, which is available to the public through the Research Data Exchange website (RDE, available from: https://www.its-rde.net/) managed by the U.S. Department of Transportation (USDOT). This study uses Driving

Dataset catalogs BSMs data obtained from 64 vehicles equipped with Data Acquisition Systems [28] – developed by Virginia Tech Transportation Institute (VTTI).

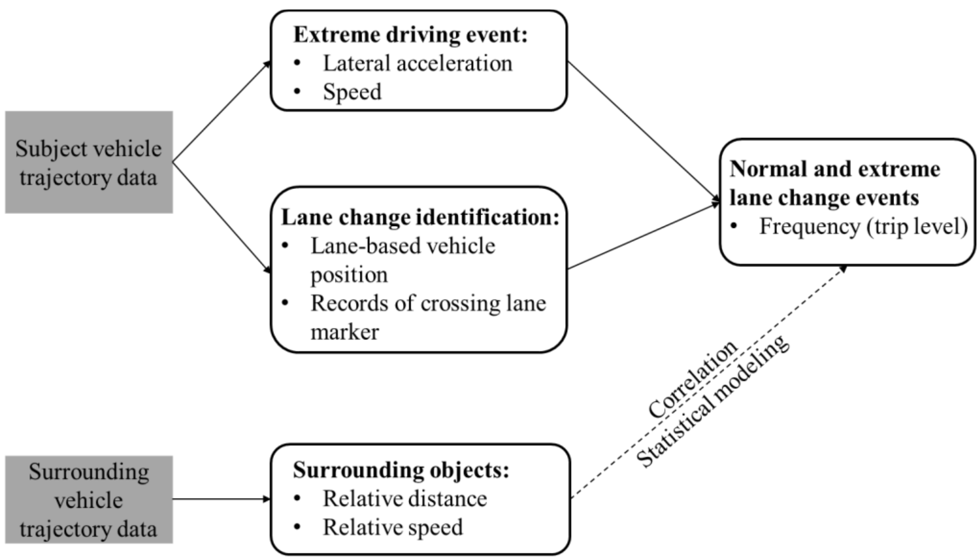
Three files are involved in the Driving Dataset: 1) The HV Primary file, which describes the subject vehicle's operation and performance, including geographic coordinates based on position (e.g., latitude and longitude), lane-based vehicle position (e.g., distance of vehicle centerline to the left or right boundary of travel lane), motion (e.g., heading, speed, and acceleration), status of a vehicle's components (e.g., lights, wipers, brakes, and turn signals), driving contexts (e.g., time and lane width), onboard device records of lane crossings (e.g., lane cross aborted, and a vehicle meets and crosses the boundary of travel lane), and fidelity of tracking lane boundary correctly; 2) The HV Radar file, which describes the objects in front of the subject vehicle, including type of front surrounding objects, and relative distance or speed to front surrounding objects; and 3) The DAS2 Trip Summary file, which provides a list of summary measures for each trip, such as trip duration and average speed. The data elements were collected at a frequency of 10 Hz. More information about other variables in driving data is available in SPMD Sample Data Handbook [29].

The whole data set contains two months (April 2012 and October 2013) of subject vehicle operations data with 83,384,195 records generated from 14,315 trips by 64 vehicles. Since a frequency of 10 Hz results in the data set being very large, this study randomly selected three trips (minimum trip duration is longer than 10 minutes) from each vehicle for analysis due to computational limitations. The final data contains 1,940,678 BSMs records from 192 trips by 64 vehicles. To investigate the influence of driving environment on lane change events, this study links the information of surrounding vehicles, e.g., relative distance or speed, to the subject vehicle trajectory data for final analysis. Figure 2.1 (a) shows the spatial distribution of vehicle trajectories for 192 trips. These trips cover major road networks in Ann Arbor, Michigan. Most trips were generated in Ann Arbor, and some long trips reached Chelsea, Farmington Hills, Canton, and Toledo. The data was verified and error-checked for outliers using descriptive statistics. Note, there are reported errors of GPS data; as the rule used for identifying lane change event is based on the relative distance to the lane boundary, the measurement errors can be eliminated.

Figure 2.1 (b) presents the conceptual framework of this study, which indicates the input variables for each step. The major objective is to identify lane change maneuvers and quantify extreme lateral driving behaviors to recognize extreme lane change events embedded in BSMs in a connected vehicle environment. The relationship between speed and lateral acceleration is investigated to establish a varying threshold of extreme lateral driving behavior at various speeds [30]. By identifying extreme lane change behaviors in real-time, the risks posed to other drivers can be identified and communicated. Also, the driver can be provided instantaneous feedback (warnings or control assists), through applications of V2V and V2I. Such information can help them make more informed decisions regarding avoiding high-risk lane change situations.



a) Spatial distribution of vehicle trajectories (192 trips)



b) Conceptual framework

Figure 2. 1 Distribution of vehicle trajectories and cconceptual framework

2.4 IDENTIFICATION OF EXTREME LANE CHANGE EVENTS

2.4.1 Identification of lane change events

The identification of lane changes in this study is based on onboard tracking systems recording a vehicle's crossing lane marker information (shown in Figure 2.2) and patterns of changes in the vehicle lateral displacement embedded in lane-based vehicle position, as a lane change is triggered when the vehicle centerline meets and crosses the lane boundary. Key variables used to determine a lane change includes:

1) Records of the vehicle meeting and crossing the lane boundary,

2) Lane-based vehicle position: distance of vehicle centerline to the left or right boundary of travel lane,

3) Tracking fidelity, i.e., that the vehicle-based vision is providing correct data for tracking lane markers, values from 0-1024 (thus the fidelity increases 100/1024 = 0.0977% with a unit increase in its value),

4) Records that a lane crossing was aborted (shown in Figure 2.2 (c)), and

5) Records that a vehicle crosses a lane successfully (shown in Figure 2.2 (a)).

The proposed algorithm contains two parts to identify lane change events. In part 1, when the onboard device provides records that a vehicle crosses a lane successfully (shown in Figure 2.2 (a)), a lane change maneuver is easy to be identified. An acceptable valid lane change is triggered when: no records of lane cross aborted, records of vehicle meets and crosses the boundary of the travel lane, the fidelity of tracking lane marker is larger than 30% [25], and records of the vehicle crossing the lane successfully.

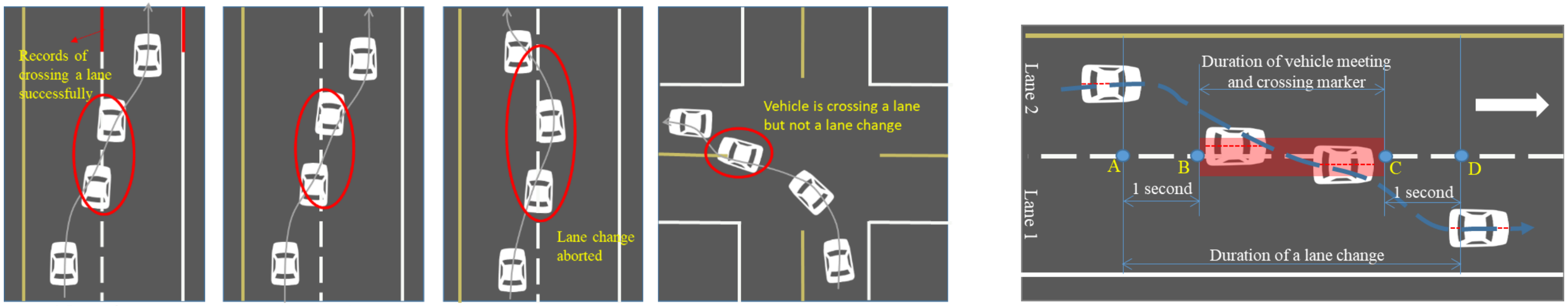
In part 2, when a driver has made a lane change but the onboard device does not provide records that the vehicle crossed a lane successfully, as shown in Figure 2.2 (b), this study captures these lane change maneuvers based on patterns of change in vehicle distance relative to lane boundary, such as sharp change in distance, from the minimum (approximated to zero) to the maximum (approximated to lane width), is an indicator of a lane change maneuver. Figure 2.2 (e) and (f) show detail patterns of the real-time vehicle distance relative to lane boundary for left and right lane change, separately.

A left lane change is coded to have occurred when the distance of the vehicle centerline to the left boundary of travel lane *decreases to a minimum* (approximately equal to 0 - distance to dash marker of lane 1) just before the vehicle centerline meets the left side marker, and then *suddenly increases to a maximum* (approximately equal to the lane width - distance to yellow marker of lane 2) just after the vehicle centerline crosses the left-side marker. Also, this left-side marker of the old lane (lane 1) becomes the right-side marker of the new lane (lane 2). The change in distance relative to the right boundary is opposite to the procedure described above.

Therefore, an acceptable valid lane change event based on lane-based vehicle position is triggered when: no records of lane cross aborted, records of vehicle meeting and crossing the boundary of travel lane, the fidelity of tracking lane marker to be larger than 30%, and the vehicle follows the lane-based vehicle position rules shown in Figure 2.2 (e) and (f). Similarly, a right lane change can also be identified.

Note that, a lane change is triggered when the vehicle centerline meets and crosses the lane marker. Although the lane change maneuver can be identified, it is hard to get the exact initial and end points of a lane change. As shown in Figure 2.2 (g), the data set provides the initial (point B) and end (point C) time points representing the time stamps that the vehicle is occupying the lane boundary. However, a real lane change maneuver should start earlier than time point B and end later than time point C. Since this study is only interested in identifying normal and extreme lane change events from thousands of lane crossing records, the identification of the real initial and end point of a lane change will not be involved. This study assumes the lane change maneuver starts one second earlier (point A) before the vehicle meets lane boundary and ends one second later (point D) after the vehicle departs from the lane boundary. Therefore, time point A and D is recognized as the initial and end point of a lane change, separately. All the analysis in this study is based on patterns of lane change recognized between the defined initial (point A) and end (point D) point.

In addition, the method is relying highly on the onboard device tracking of lane makers, so these identified lane changes are limited to specific locations with relatively clear lane markers, thus this study did not account for lane changes occurring in the intersection. However, the onboard tracking system may record boundary crossing when a vehicle makes a turn (e.g., at intersection or junction), shown in Figure 2.2 (d). A sharp change in heading will occur when a vehicle makes a turn; this study eliminated such situations based on the vehicle heading information. While other studies recommend that the intersection angle should not be skewed from 90 degrees by more than 15 to 20 degrees [31, 32], this study excluded the turning behavior if the change in vehicle heading is larger than 70 degrees during a turning maneuver. In addition, not all boundary crossings will result in lane change events. As shown in Figure 2.2 (c), a vehicle can abort a lane change by crossing back over, which is also excluded in this study. Therefore, the lane change is clearly identified on relatively straight roadways (when the angle of a curve is larger than 70 degrees) and where the lane markers are clear in this study. Figure 2.3 shows a flow chart for the onboard tracking system based on an identification algorithm.



a) Lane change with success signal

b) Lane change without success signal

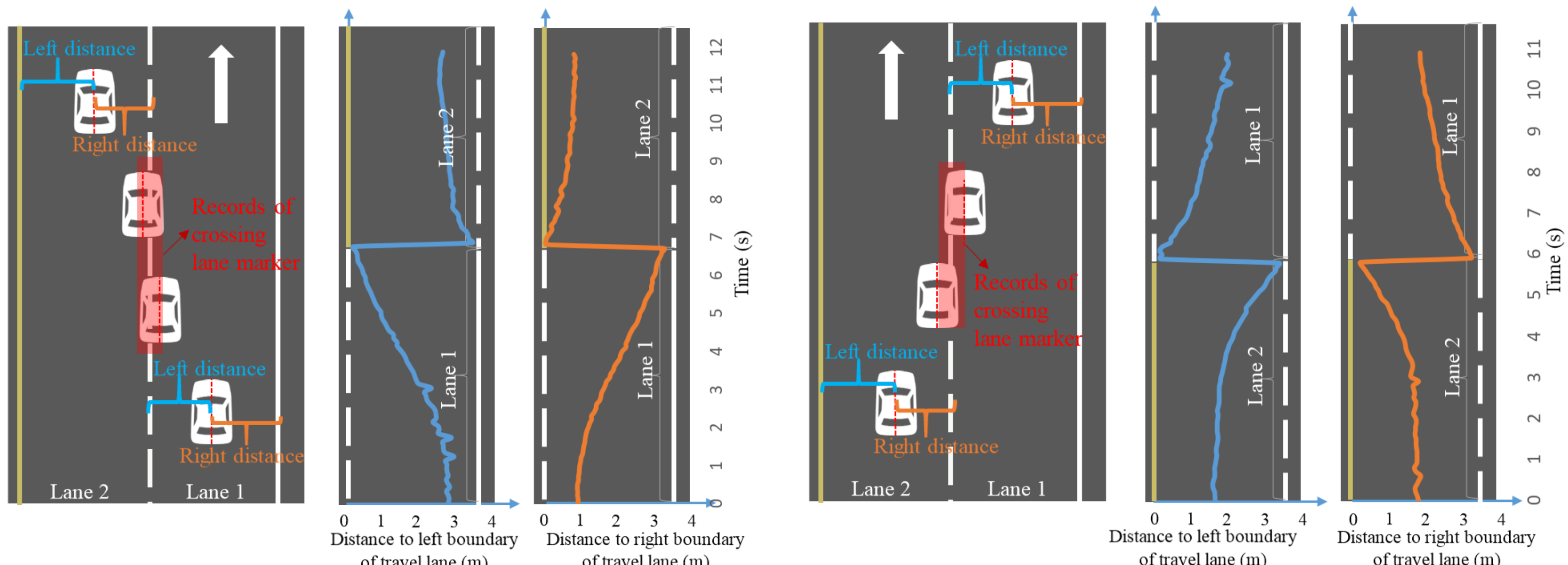
c) Lane change is aborted

d) Turning behavior at intersection

(g) The initial and end point of a lane change

1) Onboard tracking system records of a vehicle's crossing lane marker information

3) The initial and end point of a lane change



(e) Distance of vehicle centerline to boundary of travel lane in left lane change

(f) Distance of vehicle centerline to boundary of travel lane in right lane change

2) Different patterns of lane change events

Figure 2. 2 Lane change identification criteria

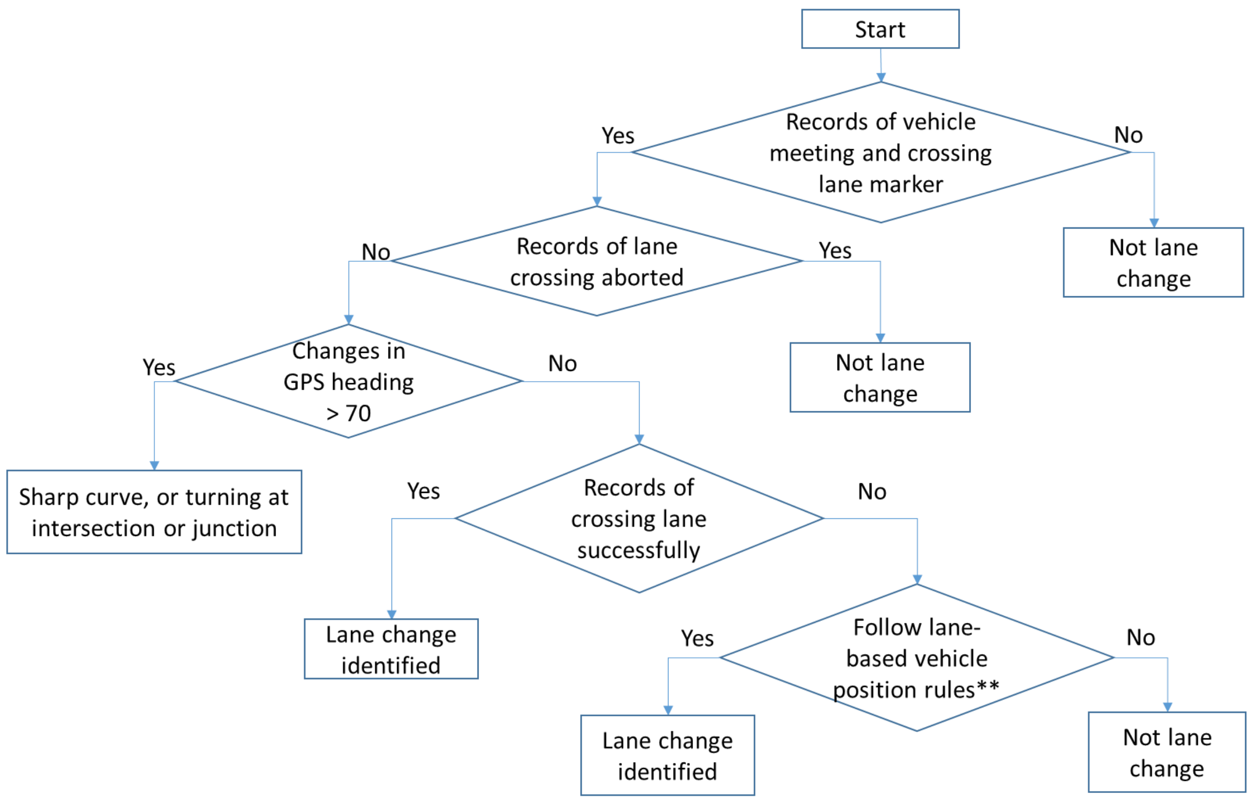


Figure 2. 3 Flow chart of identifying lane change events.
**: Figure 2.2 (e) and (f) showed detail patterns of lane-based vehicle positions

2.4.2 Extreme lane change events

Calculation of lateral acceleration

To identify extreme lateral driving, the relationship between speed and lateral acceleration was visualized. A vehicle's lateral acceleration based on lateral displacement (lane-based vehicle distance of the vehicle centerline to the boundary of travel lane) needs to be calculated. Note that a vehicle's lateral acceleration is unavailable in the data set; also, the calculated value only captures a vehicle's lateral acceleration on relatively straight roadways, which is acceptable given the lane change focus of this study. Since the onboard device records the distance of vehicle centerline to the lane boundary at a rate of 10 Hz (0.1 second), the lateral displacement of vehicle centerline from $(i-1)^{th}$ 0.1 second to i^{th} 0.1 second can be calculated. The equations used to calculate lateral speed and acceleration are as follows:

$$V_i^{Lateral} = \frac{\Delta D_{i-1,i}}{\Delta T_{i-1,i}} = \frac{Abs(Abs(D_i^{left}) - Abs(D_{i-1}^{left}))}{T_i - T_{i-1}} \tag{1}$$

$$A_{i+1}^{Lateral} = \frac{\Delta V_{i+1,i}}{\Delta T_{i+1,i}} = \frac{V_{i+1}^{Lateral} - V_i^{Lateral}}{T_{i+1} - T_i} \tag{2}$$

Where:

V_i = Lateral speed at the i^{th} 0.1 second;

T = Time stamp of 0.1 second, T = 0, 0.1, 0.2, 0.3,

i = Index for time stamp, i = 2,3,4,5,

$Abs(D_i)$ = Distance of vehicle centerline to the left boundary of travel lane at the i^{th} 0.1 second. As the BSM dataset reports D_i in negative values (e.g., -1.711 m), the absolute values of D_i were taken for the calculations;

$\Delta D_{i,i-1}$ = Absolute value in lateral displacement of vehicle centerline during $(i-1)^{th}$ 0.1 second to i^{th} 0.1 second;

A_{i+1} = Lateral acceleration when lateral speed changes from V_i toV_{i+1};

Figure 2.4 (left side) presents time series examples of lateral speed, and acceleration calculated based on Equations 1 and 2. There are clear fluctuations in lateral speed and lateral acceleration. To smooth out some fluctuations (remove noise), this study applies a 10-point moving average (a time window of 10 data points, representing one second) to calculate lateral acceleration, shown in Figure 2.4 (right side). The fluctuations in lateral speed and acceleration are reduced after smoothing the data.

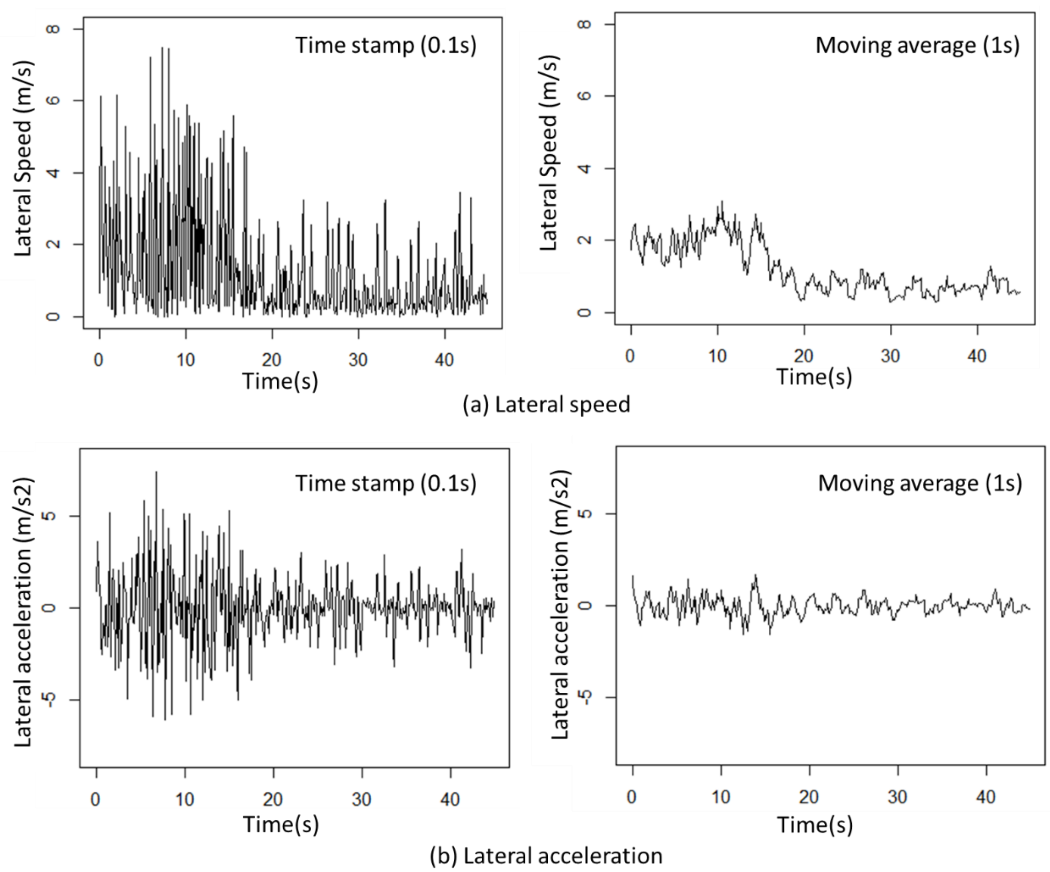


Figure 2. 4 Time series of lateral speed and lateral acceleration for a sample trip

Extreme lateral driving events

In order to understand patterns of instantaneous lateral acceleration decisions, this study visualizes the distribution of lateral acceleration across different speed ranges, shown in Figure 2.5. The figure shows that high speed (> 55 mph) is associated with relatively small lateral acceleration, indicating that lateral acceleration decreases when speed is high. As vehicles with high speed should overcome high air resistance [33], the maneuverability of vehicles would be low. Figure 2.5 also indicates a nonlinear relationship between speed and lateral acceleration.

Previous studies have proposed methods, such as giving a cut-off value of acceleration as a threshold to distinguish extreme driving and calm driving [34-37]. However, Figure 2.5 reveals obvious variations of lateral acceleration across different speed ranges in the real driving environment, this study uses an innovative method to quantify extreme lateral driving events [30, 33, 38]. A speed-based method was used. Instead of using a given cut-off value of acceleration as a threshold, the new cut-off value of acceleration changes along with speed. The detail steps of identifying extreme lateral acceleration events are given below:

- In order to show the magnitude of lateral acceleration under different speed situations, this study first splits speed into different bins with a 0.5 mph of bandwidth. For example, “bin=1” refers to BSMs records whose speeds were reported between 0 and 0.5 mph. The maximum speed of 192 trips was about 96 mph, so more than 182 speed groups (>192 bins) are generated.
- Each speed bin would generate a corresponding distribution of lateral acceleration. This study used the 95th percentile value of lateral acceleration in each bin as the threshold [38]. Specifically, within one speed bin, if the lateral acceleration of one BSM (0.1 seconds) is higher than the 95th percentile value of acceleration, this BSM will be identified as an extreme lateral acceleration event.

Figure 2.5 also presents thresholds (edge of the band) for identifying extreme lateral driving patterns for all speed ranges. The thresholds vary across the different speed ranges. The red points present extreme lateral acceleration events, which indicates the subject vehicle is volatile at these timestamps. Notably, the quantification of the extreme instantaneous driving behavior is defined in a broad relative level, that is the volatile behaviors are these timestamps where the accelerations are much higher or lower than the normal situations within each speed group, as shown in Figure 2.6. Therefore, these extreme lane change maneuvers identified in this study are relative aggressive compared to normal lane change maneuvers. Warnings can be generated if there are more than five continuous BSMs (> 0.5 seconds) that have lateral accelerations larger than the 95th percentile threshold, indicating an extreme lateral driving event.

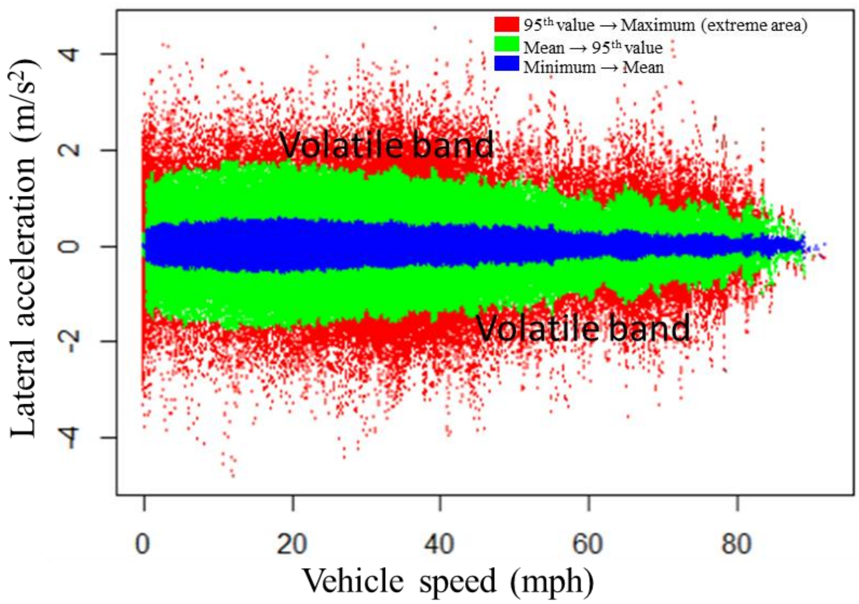


Figure 2. 5 Distribution of vehicle speed and lateral acceleration

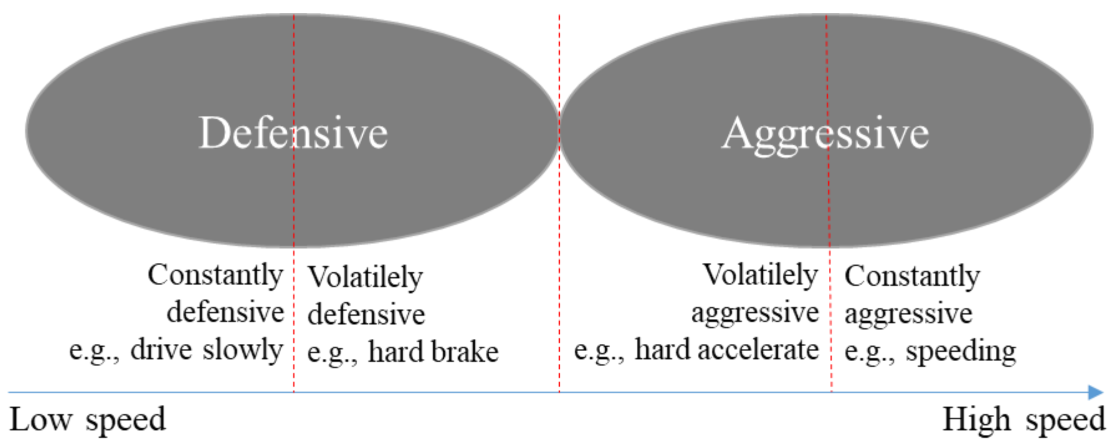


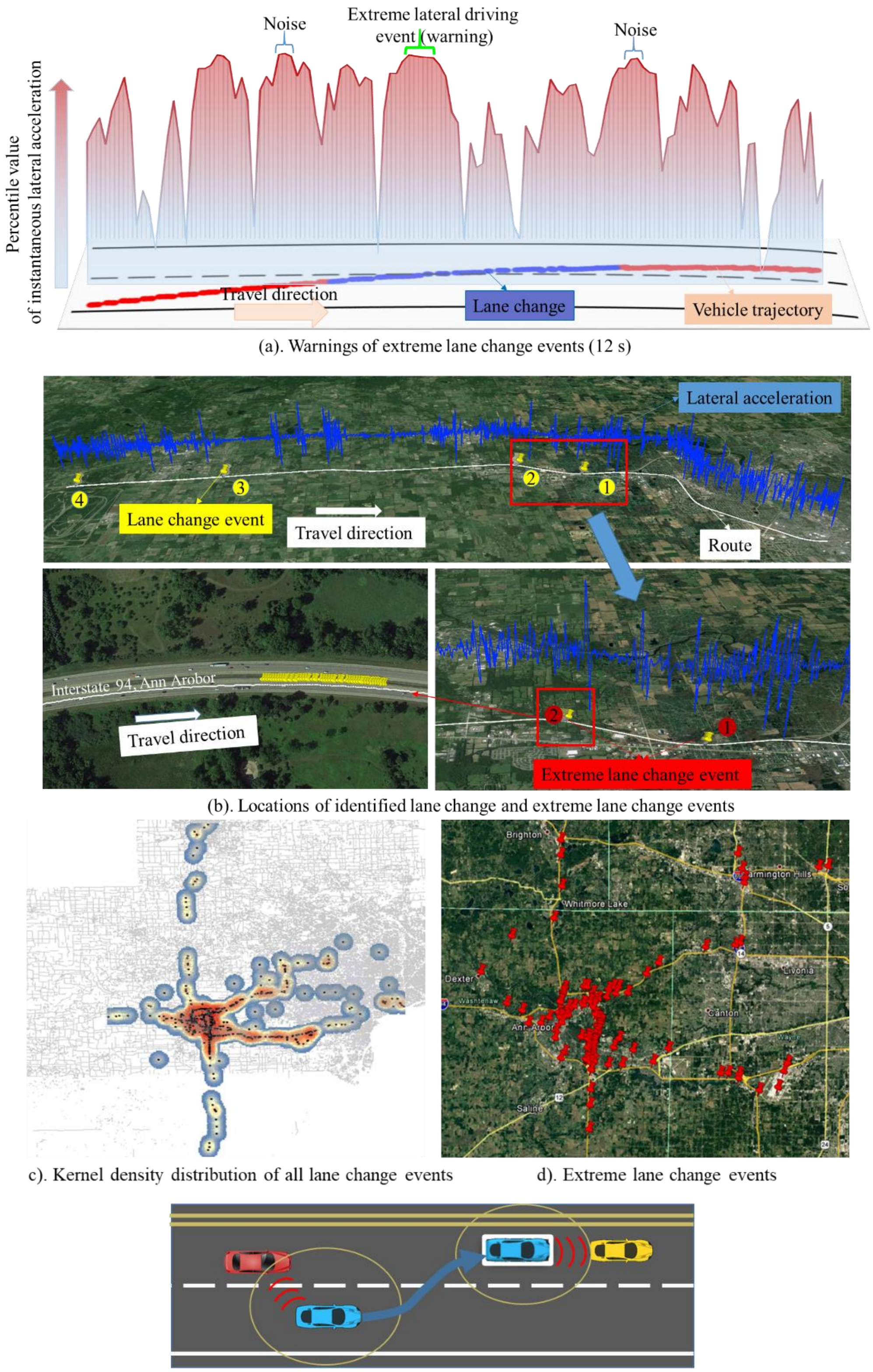
Figure 2. 6 Volatile driving behavior

Extreme lane change events

Based on the above discussion, extreme lane change events are identified, which are those where lateral acceleration between consecutive 0.5 time stamps exceeds the 95^{th} percentile threshold at the initiation and before the end of the lane change maneuver. Figure 2.7 (a) presents a sample of identified extreme left lane change events. During the left lane change (blue color), an extreme lateral driving event (five continuous BSMs that the lateral acceleration exceeds the thresholds) is identified. Note, the lateral acceleration of some time stamps also exceed the 95^{th} percentile threshold but not continued to 0.5 seconds, these will not be recognized as extreme driving events ("noise" shown in Figure 2.7 (a)).

Figure 2.7 (b) visualizes a trip with the patterns of lateral acceleration, locations of identified lane change events (1 and 2), and extreme lane change events (3 and 4). As expected, driving near city areas is more volatile than driving near rural areas based on magnitudes of lateral accelerations.

Figure 2.7 (c & d) visualizes distributions of total lane change and extreme lane change events in ArcGIS and Google Earth, respectively. The identified "hot spot" locations of extreme lane change events have the potential applications to improve the traffic safety through proper roadway design, since the subject vehicle might make an extreme lane change event due to the improper roadway design. Figure 2.7 (e) also presents an example of specific warnings or control assists that could be applied in real driving environments when extreme left lane change event is recognized. If the host vehicle (blue car) makes an extreme left lane change with hard braking at the curve, a sideswipe crash warning or control assist can be provided to the red car. After the host vehicle (blue car) makes a successful left lane change and continues to accelerate hard, a warning to the yellow car can be provided.



(a). Warnings of extreme lane change events (12 s)

(b). Locations of identified lane change and extreme lane change events

c). Kernel density distribution of all lane change events

d). Extreme lane change events

e). Applications in real driving environment

Figure 2. 7 Visualization of lane change and extreme lane change events in space and applications of warnings and control assists

2.4.3 Results of identified lane change events

Figure 2.8 presents the distribution of identified lane change events. A total of 654 lane change events were identified from 1,557 meeting and crossing lane marker events generated from 192 trips by 64 vehicles. Not all drivers provided turn lights to inform their lane change behaviors (424 out of 654). Notably, 128 extreme lane change events were identified. As the trip duration of many trips were less than 15 minutes, the majority of lane change frequencies are less than 3 per trip. High frequencies of lane change events are found in high average travel speed range. Drivers might expect to achieve high speed through lane change maneuvers, especially when there are vehicles with low speeds in front in their travel lane.

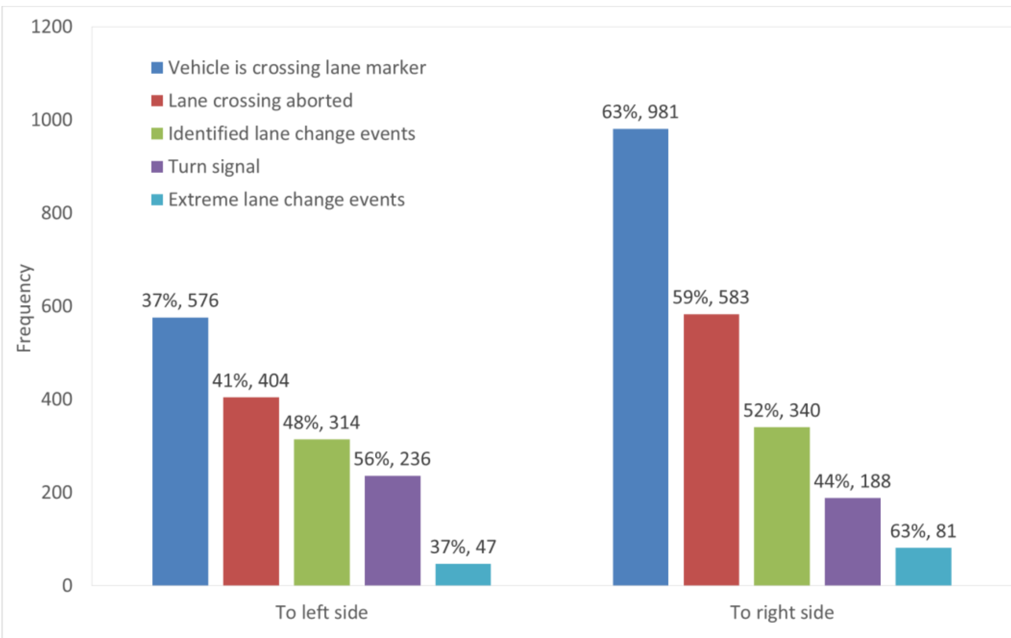


Figure 2. 8 Distribution of lane change frequency

This study uses a confusion matrix to validate the performance of lane change identification algorithms. Four trips were randomly selected whose number of lane change events were larger than 5 for validation. These trips were visualized on Google Earth to compare the actual number of lane change events and algorithm-based identified number of lane change events. The sensitivity (true positive rate, the proportion of lane change events that are correctly identified) and specificity (true negative rate, the proportion of non-lane change events that are correctly identified) were calculated for evaluation. The higher the sensitivity and specificity, the better the performance [39]. Given the value of sensitivity (0.889) and specificity (0.909), it seems the identification method performed well for a lane change with sufficient lane change signals occurring on a relatively straight roadway (curve angle < 20 degrees), where the lane markers are clear. The lane changes identified incorrectly, were due to the unclear lane markers, low quality of data, and were near intersections.

This study also calculates the average distance and duration for normal and extreme lane change events. As expected, the average distance and duration of extreme lane change events are higher than normal lane change events, however, the average speed of extreme lane change events is lower than the normal lane change events, which indicates the subject vehicle might make an extreme lane change with higher acceleration in short distance and duration, as a result, it might be more dangerous than the normal lane change event.

2.5 CORRELATES OF LANE CHANGE EVENTS

After identifying lane change events, it is important to understand these events. Considering the count nature of lane change event frequency, a Poisson regression model is estimated. The probability of trip *i* having y_i lane change or extreme lane change events is written as:

$$P(y_i) = \frac{exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}$$

Where:

$P(y_i)$ is the probability of trip i having y_i lane change or extreme lane change events, $i = 1, 2, \dots, n$;

λ_i is the expected number of lane change or extreme lane change events.

In Poisson regression models, the relationship between frequency of lane change or extreme lane change events generated by trip i and explanatory variables is assumed to be given by:

$$\lambda_i = EXP(\beta X_i)$$

Where:

X_i are explanatory variables (e.g., driving speed);

β are estimated coefficients of explanatory variables.

Descriptive statistics

Table 2.2 shows the statistical description of variables at the trip level. Since the study explores the relationship between lane change events and surrounding objects, there exist four trips without surrounding objects. Therefore, they are exclusive in the final analysis. Finally, 188 trips with the influence of surrounding objects are used for analysis. On average, each trip generates 3.5 lane change events (1.7 left and 1.8 right). Of these, there are 0.68 extreme lane change events (0.25 left and 0.43 right) per trip. The average trip duration is 20.5 minutes. Note, that in one trip (maximum travel speed 81.7 mph) generated 28 lane change events for 47 minutes duration while 53 (out of 188) trips did not generate any lane change events.

Modeling results - trip level

Table 2.3 shows the final modeling results for correlates of the number of lane change events. Variables in the model specification were eliminated using backward stepwise variables selection method (at 10% level), as they only explain little variations in the data [40]. Both models have shown the reasonable goodness of fit. Note that these models were limited to a lower sample size and related explanatory variables, the estimation results might change when more data is used. Notably, the results revealed that the maximum speed during a trip and long trip duration are associated with more lane change events. The results of surrounding objects show interesting results. The number of right/left side objects are associated with less lane change events, but the number of lane change events is high when there are front objects in the travel path. The subject vehicle makes less lane change events along with the increase in relative speed to front object, indicating the subject does not need to make a lane change to achieve the satisfied speed. For extreme lane change events, only maximum speed, season and trip duration have shown statistically significant correlations. Similarly, the subject vehicle makes more extreme lane changes along with the increases in the maximum speed during a trip. Note that these models were limited to a lower sample size and related explanatory variables, the estimation results might change when more data is used.

Table 2. 2 Data descriptive of variables at trip-level (N=188)

Variables		N	Mean	Std. Dev.	Min	Max
Lane change behavior	Total number of identified lane change	188	3.468	4.622	0	28
	Number of identified left lane change	188	1.670	2.403	0	13
	Number of identified right lane change	188	1.798	2.531	0	16
	Total number of identified aggressive lane change	188	0.681	1.154	0	8
	Number of identified aggressive left lane change	188	0.250	0.553	0	3
	Number of identified aggressive right lane change	188	0.431	0.859	0	7
	Total number of aborted line crossing	188	5.261	6.886	0	37
Trip attributes	Trip duration (min)	188	20.522	11.409	10.833	56.413
	Average speed (mph)	188	42.743	16.279	6.338	75.739
	Maximum speed (mph)	188	64.950	14.356	45.012	96.109
Vehicle maneuvering	ABS state	188	0.657	0.471	0.000	1
	Brake (engaged) (%)	188	18%	0.129	0%	59%
	Headlight (engaged) (%)	188	17%	0.336	0%	100%
	Stable control (engaged) (%)	188	18%	0.386	0%	100%
	Vehicle wiper (engaged) (%)	188	3%	0.176	0%	100%
	Total Number of turn signal	188	2.245	3.682	0	23
	Number of left turn signal	188	1.255	2.018	0	12
	Number of right turn signal	188	0.989	1.873	0	12
Contextual factors	An exit on the left side (engaged) (% *1000)	188	9%	0.242	0%	127%
	An exit on the right side (engaged) (% *1000)	188	36%	0.466	0%	237%
	Season (1-spring, 0-autumn)	188	0.487	0.437	0	1
	Darkness	188	0.080	0.253	0	1
	Rush hour	188	0.540	0.460	0	1
	Average Lane width (m)	188	3.391	0.654	0	4.845
	Average distance to left lane marking (m)	188	-1.848	0.486	-3.364	0
	Average distance to right lane marking (m)	188	1.698	0.453	0	2.555
Surrounding objects	Percentage of time with surrounding objects (%)	188	19%	0.108	1%	70%
	Average number of front objects	188	0.704	0.473	0.016	2.984
	Percentage of time with front vehicle in path (%)	188	52%	0.205	2%	100%
	Average of surrounding object on right side	188	1.567	0.248	1.000	2.649
	Average of surrounding object on left side	188	1.506	0.256	1.107	2.622
	Average relative speed to front object(m/s)	188	-0.152	0.601	-2.108	2.557
	Average relative distance to front object (m)	188	36.300	16.362	9.226	78.910
	Percentage of time in freeway (%)	188	37%	0.332	0%	100%

Table 2. 3 Poisson model for frequency of normal and extreme lane change events

Variables (Dependent variable = Number of lane change events at trip level)		Normal lane change model: Poisson model		Normal lane change model: Poisson model - stepwise		Extreme lane change model: Poisson model		Extreme lane change model: Poisson model – stepwise	
		β	e^{β}	β	e^{β}	β	e^{β}	β	e^{β}
Trip attributes	Trip duration (min)	0.033***	1.034	0.033***	1.034	0.032***	1.032	0.029***	1.029
	Average speed (mph)	0.014	1.014	-	-	0.008	1.008	-	-
	Maximum speed (mph)	0.027***	1.028	0.027***	1.028	0.033***	1.033	0.030***	1.031
Vehicle maneuvering	Brake (engaged) (%)	1.479**	4.386			0.271	1.311	-	-
	Vehicle wiper (engaged) (%)	-0.129	0.879	-	-	0.171	1.187	-	-
Contextual factors	An exit on the left side (engaged) (% *1000)	0.736***	2.087	0.715***	2.044	0.513	1.670	-	-
	An exit on the right side (engaged) (% *1000)	0.012	1.012	-	-	0.309	1.361	-	-
	Season (1-spring, 0-autumn)	-0.089	0.915	-	-	-0.597**	0.550	-0.591***	0.554
	Darkness	0.641***	1.898	0.605***	1.831	0.580	1.786	-	-
	Rush hour	0.151	1.163	-	-	0.023	1.023	-	-
	Average Lane width (m)	-0.270**	0.763	-0.195*	0.823	-0.511*	0.600	-	-
	Average distance to left lane marking (m)	-0.496***	0.609	-0.411***	0.663	-0.598*	0.550	-	-
Surrounding objects	Average of front object	0.340***	1.405	0.312***	1.366	0.340	1.405	-	-
	Percentage of time with front vehicle in path (%)	0.548**	1.729	0.454*	1.575	0.802	2.231	-	-
	Average of surrounding object on right side	-0.404*	0.667	-0.442**	0.643	-0.669	0.512	-	-
	Average of surrounding object on left side	-0.502**	0.605	-0.413**	0.662	-0.646	0.524	-	-
	Average relative speed to front object(m)	-0.166**	0.847	-0.156**	0.856	0.066	1.068	-	-
	Average relative distance to front object (m/s)	0.005	1.005	-	-	-0.001	0.999	-	-
	Percentage of time in freeway (%)	-0.367	0.693	-	-	-0.143	0.867	-	-
Constant		-1.712**	0.181	-0.858*	0.424	-1.734	0.177	-2.930***	0.053
Summary statistics	Sample size	188		188		188		188	
	Adjusted R^2	0.269		0.264		0.180		0.152	
	Log likelihood at β	-491.747		-495.008		-190.408		-196.946	
	Prob. > χ^2	0.000		0.000		0.000		0.000	

Notes: "***"means statistical significant associations were found (at 1% level); "**"means statistical significant associations were found (at 5% level); "*"means statistical significant associations were found (at 10% level).
Adjusted R^2 refers to 1 – (Log Likelihood at β/Log Likelihood at 0);

2.6 LIMITATIONS

The data used for analysis are BSMs data collected from vehicles by roadside equipment participating in SPMD. Data acquisition system (DAS) devices are used to collect data from connected vehicles. Although these devices are expected to provide highly accurate data, there still might be some unknown measurement errors in the data set. The methods of identifying lane change events and the calculation of lateral acceleration are highly reliant on the quality of reported lane marker tracking data. The accuracy of the lane-based position will influence the results directly as errors existed in GPS data. To eliminate the influence of GPS errors, this study removes cases with low tracking fidelity. In addition, since the rule used for identifying lane change events is based on the relative distance to the lane boundary, therefore, the GPS errors can be eliminated. In sum, the influence is minor based on the validation results on Google Earth.

Another limitation is that some high influencing factors, such as traffic density, are not involved in the analysis. For example, more abrupt lane changes might result from high traffic density. An additional limitation is the selected data used for analysis. Due to computational limitations, only a sub-sample data from 192 trips are used for analysis.

2.7 CONCLUSIONS

This study contributes to understandings of normal and extreme lane change behaviors by interpreting connected and automated vehicle data. A deeper understanding of these behaviors can form the basis for generating alerts and warnings that can reduce the impacts of extreme lane change events. The proposed lane change methodology uses multiple indicators that include:

1) Vehicle position, i.e., a sharp change in the distance of vehicle’s centerline relative to lane boundaries.

2) Lane crossings recorded by an onboard unit, i.e., when a vehicle occupies then crosses a lane marker. Complementing this are data recorded by the onboard unit when a vehicle aborted a lane change.
3) The lane marker tracking data quality, as indicated by a fidelity variable.

The methodology and connected vehicle data used identified 654 lane changes for all 192 trips, and showed an average of 3.5 lane changes per trip. Plotted trajectories of selected trips on Google Earth validated these lane changes.

Analysis of the data showed that lateral accelerations are higher at lower speeds, but get lower as speeds increase. This relationship formed the basis for the investigation of extreme lane changes. In this study, lateral accelerations in the 95^{th} percentile at the initiation and before the end of a lane change maneuver were considered extreme lane changes. The data showed 128 extreme lane changes (0.68 extreme lane changes per trip). Poisson regressions identified the key causes of lane changes. These causes included existing objects in the travel path, small speed differences with the front object, higher maximum speed during the trip, darkness, and exiting on the left side of the travel direction.

Based on analysis of this data, warnings that help surrounding drivers adjust their behaviors in order to accommodate extreme behavior by the host vehicle driver can be generated. The application of connected vehicle technologies will help proximate vehicle drivers make more informed decisions and avoid drivers who are undertaking high-risk lane changes. Connected vehicle technology can warn the host vehicle driver if their frequency of extreme lane change behaviors during a trip is relatively high so that they are encouraged to make smoother lane changes during the remainder of their trip. Predicting extreme lane change behaviors in real-time for the host driver is challenging and needs further research. In addition, researchers can visualize the “hot spot” locations of extreme lane change events in Google Earth using connected vehicle data, which may indicate when the subject vehicle might make an extreme lane change event due to improper roadway design. We can consider improved roadway design or proper warnings at these “hot spots” locations.

CHAPTER 3 WHAT IS THE LATERAL SHIFTING VOLATILITY OF LANE KEEPING BEHAVIORS

This chapter is a revised version to be submitted by Meng Zhang, Asad Khattak, and Zachary Roberts

ABSTRACT

Roadway and lane departure crashes represent a sizable portion of all roadway crashes, which results in large portion of social cost. Advanced technology such as onboard lane keeping warning systems is developed to prevent these crashes. To get an in-depth understanding of lane keeping behaviors, this study explores the relationship between driving situations and lateral shifting volatility, which quantifies the fluctuation in instantaneous lateral displacement, by analyzing a sub-sample of 1,550,107 Basic Safety Messages (BSMs) records sent by vehicles, at a 10 Hz frequency, and received by roadside equipment. There were 192 randomly selected trips (10 minutes or longer) from 64 drivers. The trajectories' data come from connected vehicles participating in the Safety Pilot Model Deployment (SPMD) program in Michigan. The BSMs describe vehicle operation performance measures, e.g., relative distance from vehicle centerline to lane boundaries; thus, a measure called lateral shifting volatility, which quantifies fluctuation in lateral displacement, is developed. The study uses the coefficient of variation (COV), defined as the ratio of standard deviation to mean, to quantify shifting volatility. To explore the correlation between shifting volatility and different driving situations, a linear regression model is estimated in this study. The modeling results show that the subject vehicle is more volatile when traveling at high speeds and when the vehicle keeps a low space gap with the vehicle in front of it. These results provide insights on how lane departure warning systems can help drivers make informed lane departure decisions in a connected vehicle environment.

3.1 INTRODUCTION

Lane departure crashes, including single-vehicle, head-on, and sideswipe crashes, provide a tremendous opportunity to increase roadway safety through intelligent transportation systems technology. According to the statistics from Federal Highway Administration (FHWA), there are 18,275 fatalities, which represents 54% of traffic fatalities in the United States, resulted from lane departure annually between 2013 and 2015 [41].

Until lane keep assist and full autonomy become commonplace on roadways, the best solution for reduction of crashes and crash severity is to provide drivers with lane departure and blind-spot warnings. A 2016 Insurance Institute for Highway Safety (IIHS) study reveals that lane departure warning can reduce the rate of lane departure crashes by 11% and lower injury rates by 21%. Unfortunately, many drivers still see these warnings as an annoyance and deactivate them [42]. This highlights the importance of being able to predict driver behavior and deploy targeted warning systems that can keep drivers alert and responsive, without excessive or unnecessary activation frequency. Additionally, once connected and automated vehicles (CAVs) begin to share the roadways with conventional, human-driven vehicles, it will be helpful for these vehicles to be able to better predict the likelihood of another vehicles' failure to maintain their lane.

This study aims to develop a measure called shifting volatility to quantify the variability in instantaneous lateral displacement, which is the unique aspect of this study. Previous studies applied different measurements to describe driving behaviors. Liu and Khattak proposed a new measurement named "driving volatility" to quantify the extreme driving decision at micro-level based on vehicle motion, e.g., the distribution of vehicle acceleration and speed [10]. To explore the volatile driving decision, this study proposes the shifting volatility measured by coefficient of variations (COV), defined as the ratio of standard deviation to mean, to quantify fluctuations in the instantaneous lateral displacement [43]. The sufficient geo-referenced trajectories data collected from connected vehicles enable the analysis. These data are Basic Safety Messages (BSMs) sent by vehicles (reported at 10 Hz) and received by roadside equipment participating in the Safety Pilot Model Deployment (SPMD) program in Michigan. These BSMs describe a vehicle's performance, e.g., relative distance from vehicle centerline to lane boundary, which makes it is possible to measure the shifting volatility.

In summary, the key objectives of this study are: 1) develop a measure called shifting volatility to quantify the variability in the instantaneous lateral displacement; and 2) explore the correlates of shifting volatility with different driving situations.

3.2 LITERATURE REVIEW

Previous studies identified three primary factors, including trajectory based, driver based sensors and external sensors, for lane departure prediction. A trajectory based system would model lane-keeping ability based on attributes embedded in current vehicle trajectory, such as speed, acceleration (lateral and longitudinal), steering angle, yaw, etc. [44-46]. Driver based factors use sensors to determine a driver's attentiveness based on eye tracking, biometrics, facial emotion or reaction, etc [45, 47-49]. External sensors contain environmental conditions such as weather, lane geometry, vehicle targets, pedestrian targets, and other features that could serve as distractions or otherwise affect a driver's ability to maintain their lane [45, 47]. These factors show potential for recognizing the likelihood of a lane departure event.

The most traditional method of predicting lane departures is to look at the trajectory of the vehicle relative to the boundary and model the likelihood of a lane departure. Lee, et. al.

studied lane change characteristics as a baseline to determine what a typical to severe lane change looks like. The authors concluded that turn signal use represented only about 44% of intentional lane changes and that the mean duration is 9.61 seconds from tangent to tangent, with some variation depending on the type of roadway. The steering angle is also reasonably predictable based on the situation, which can be determined by surrounding vehicles, but averages a peak of 8.11 degrees. These characteristics can be used to discern when a lane change that has been initiated is intentional [46]. McCall also looked at lane position prior to a lane change event. In this study, a time from initiation to crossing the boundary represented approximately 2 seconds. However this lane change is measured to a different end point [44].

Roadway departure crashes are most frequently a result either directly or indirectly of human error, including driving too fast under different weather conditions, inattention, impairment, or other means of failing to maintain control of the vehicle. Based on path alone, it is difficult to determine the intention of a driver being approaching the boundary of a lane or roadway. Driver intentions have been measured by several studies. McCall used driver facial analysis to model driver intent. A relationship was established between head motion and lane change intention using Bayesian learning. The author was able to observe that lane change intent could be identified 0.5s earlier when using data from head motion versus vehicle path alone [50]. Distraction is another predictor in lane departures due to human behavior. In a 2011 study of roadway departure crashes, Lord et. al. found that 92 of the 394 roadway departure crashes (23%) were the result of a distracted driver [47]. Edwards, et. al. also looked extensively at driver behavior and determined that among behavioral factors considered, an overlapping secondary task was the single highest predictor of maximum lane deviation variance in test cases [45]. Hallmark, et. al. could show that the more time drivers spent looking ahead at the roadway, the less likely a roadway departure event was to be captured in the data [51]. However, Sayer, et. al. collected random samples of drivers during warning and no-warning time periods and concluded that warnings were no more likely to be issued when engaged in a secondary task than when not [49].

Nodine, et. al. analyzed naturalistic data for various near-crash factors, and found that secondary tasks were distracting drivers during 52% of sensor alerts. This same study also found that the application of sensor based warning systems could reduce the rate of lane-change and road departure risk events by 33% and 19%, respectively [48]. Navarro, et. al. echoed this finding in a 2016 study, showing that a lane departure warning device significantly improved steering reaction time during a distraction task by approximately 0.3 seconds [52]. Although more difficult to detect with non-intrusive measures, driver fatigue could also be representative of inattention. Moller, et. al. identified “microsleep” events were a high predictor of lane departure risk. These events were significantly more likely to occur in the afternoon, versus morning or mid-day [53].

Driver reaction to lane-keeping warning systems is also an important consideration. Sayer, et. al. found that the presence of warning systems cut the number of lane departures in half, from 14.6 departures per 100 vehicle miles to 7.6. The duration of lane departure also dropped from a mean of 1.98 seconds to 1.66 seconds. Additionally, a 12.6% increase in the number of lane changes made indicates that these systems empower drivers with an added feeling of security [49]. To the contrary, Nodine, et. al. found that the presence of lane assist warning devices had no effect on drivers’ attention to the roadway, noting that drivers had their eyes focused on something besides the roadway immediately prior to 6% and 7% of alerts given with alerts un-equipped and equipped, respectively [48]. The downsides to installing a lane departure warning system may be miniscule, as Navarro, et. al. indicated in their 2016 study that the existence of a lane departure warning system did not negatively affect driver behavior in the instance of a missed warning [52].

Driver behavior is also affected by external agents. Roadway characteristics have been shown to be predictors of road departure crashes. Lord, et. al. found that shoulder type was correlated with run-off-the-road crashes in Texas. 52% of road departure crashes were found to have occurred on surfaced shoulders. This characteristic is overrepresented in the crash data, as only 43% of vehicle miles occurred on surfaced shoulder roadways. Nodine, et. al. found that 64% of near-miss road departures occurred to the left of the traveled way [48]. Sayer, et. al. similarly found that when testing response to lane departure warnings, 69% of these warnings were issued to the left side of the road [49].

Some data exists to indicate that location of other vehicles on the roadway also plays a role in driver awareness of lane position. Sayer, et. al. concluded that the average duration of a lane departure in the opposite direction of an adjacent vehicle increased due to the presence of the vehicle. The average duration with no vehicle present was 1.80 seconds and was 2.28 seconds with a vehicle present. The authors went on to find that when an adjacent lane was occupied, drivers moved away from the vehicle on average 27 cm (10.6 in) to the left or 10.7 cm (4.2 in) to the right, in the opposite direction from the adjacent vehicle [49]. Drivers may treat adjacent vehicles similarly to roadside obstacles. When a potential conflict is known, it could increase driver attentiveness, as Hallmark, et. al. showed that roadside barriers reduced the likelihood of a roadway departure to the right, as did chevrons, raised pavement markers and other forms of curve delineation [51].

Potential for bias in these studies could arise from the fact that the simulation and naturalistic data all came from participants that knew they were being observed. Additionally, studies that identify characteristics that do not play a role in lane departures are not likely to be published, unless they are isolated characteristics in a larger study with more attention-grabbing results. Gaps in the research include a lack of focus on how lane departure probabilities can be affected by target vehicles in the front and rear. Data also appears to be limited with respect to driving situation and lane departure.

3.3 METHOD

3.3.1 Data and conceptual framework

This study creates a unique data set by combing multiple data sources: 1) Basic Safety Messages (BSMs) collected from Safety Pilot Model Deployment (SPMD) in Ann Arbor, Michigan, and 2) Roadway information extracted from OpenStreetMap.

Basic Safety Messages (BSMs)

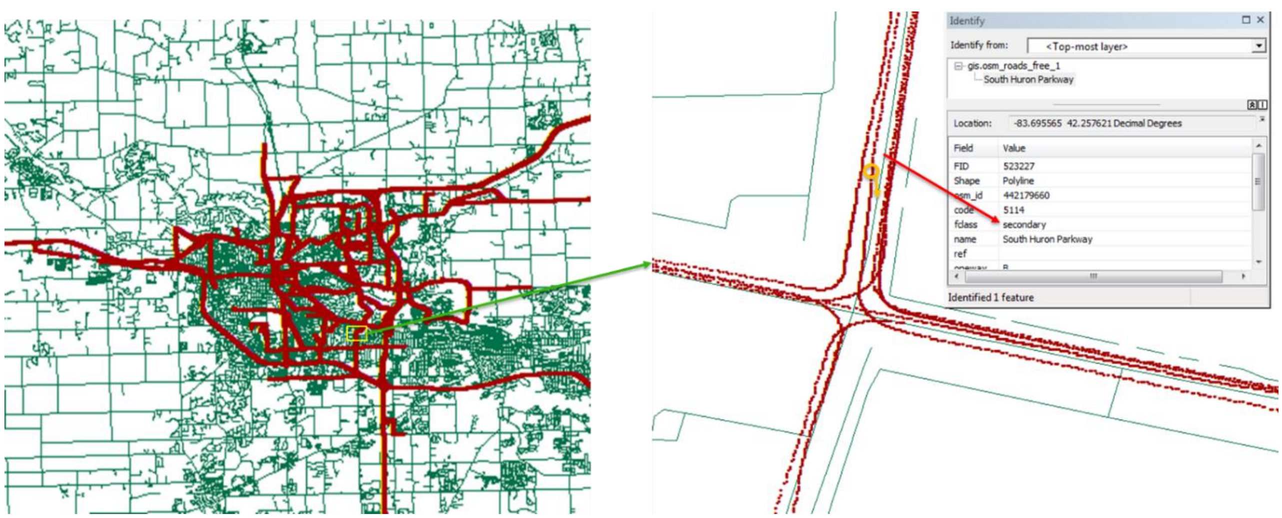
The data used for analysis are BSMs, reported at a 10Hz frequency, sent by vehicles and received by roadside equipment participating in the Safety Pilot Model Deployment (SPMD) in Ann Arbor, Michigan. These BSMs data are obtained from Research Data Exchange (RDE, available from: https://www.its-rde.net/), maintained by the US Department of Transportation. This program provides different types of data, including contextual data and vehicle-based data. The vehicle operation data archived in the Driving Dataset are used for analysis, which is collected from vehicles equipped with Data Acquisition System (DAS) – developed by Virginia Tech Transportation Institute (VTTI).

The Driving Dataset contains three sub-files: 1) trajectory data of subject vehicle (reported at 10 Hz frequency), which describes the subject vehicle's operation and performance, including lane-based vehicle position (e.g., the distance of vehicle centerline to the boundary of travel lane), geographic position (e.g., latitude and longitude), vehicle motion (e.g., speed and acceleration), onboard device records of lane tracking information (e.g., vehicle meets and crosses the lane boundary), driving context (e.g., time stamp), and vehicle performance information (e.g., lights, wipers and brakes). Given the high-resolution of lane-based vehicle position, it is possible to capture the vehicle shift displacement from the lane center; 2) trajectory data of surrounding vehicles (reported at 10 Hz frequency), which describes the relative distance and speed to surrounding vehicles; and 3) trip summary of subject vehicle (aggregated trip level), which contains the trip-level information, such as trip duration and distance. More detailed descriptions of variables involved in the dataset can be found on the SPMD Sample Data Handbook [29].

Since a 10 Hz reporting rate results in a sizable dataset, this study randomly selects three trips with travel time being longer than 10 minutes for analysis. Thus, this study get 192 trips from a total of 14,315 trips which representing 83,384,195 driving records. As this study explores the fluctuation in lateral shifting relative to the travel lane centerline, the aborted lane change and successful lane change records are removed from the data set [54]. After data cleaning and error check, this study finally gets 1,550,107 driving records for analysis.

Roadway information from OpenStreetMap

Since the driving behavior might vary from freeway to local roadway due to different driving situations, e.g. vehicle speed, this study also links the roadway information extracted from network shape file maintained by OpenStreetMap to these trajectories data. This study extracts the roadway information by visualizing the network shape file of Ann Arbor city and these vehicles' trajectories in ArcGIS, as shown in Figure 3.1 (a). This study links each trajectory point to the closest roadway to get its roadway information, as shown in Figure 3.1 (b). From the most to least important, the OpenStreetMap classifies the roadway into: motorway, primary, secondary, tertiary, unclassified, residential, and service road. As motorway is equivalent to the freeway and the primary road are often used to link larger towns, indicating high speed limits, therefore, this study re-codes the roadway into two categories: 1) freeway with related high speed – reported as motorway and primary road, and 2) local roadway with related low speed – others, e.g., secondary road. The freeway average speed is 62 mph (show a peak at 75 mph) while average speed of local roadway is close to 28 mph (show a peak at 40 mph), which indicates the classification defined in this study is reasonable.



(a) (b)

Figure 3. 1 Link vehicle trajectories to roadway

New user-defined variables

Since the lane keeping behavior might vary under different driving situations, e.g. congestion vs. non-congestion, this study explores the effects of different driving situations on lane keeping behaviors. As the GPS data does not report the exact driving situations, this study creates the user-defined driving situations to indicates different driving situations. Based on the information of roadway types, subject vehicle speed and space gaps to front vehicle in the travel path, nine driving situations are created:

1) Type 1: freeway, congested (speed <= 40 mph), & short space gaps (space <= 10 m);
2) Type 2: freeway, congested (speed <= 40 mph), & long space gaps (space > 10 m);
3) Type 3: freeway, non-congested (speed > 40 mph), & short space gaps (space <= 10 m);
4) Type 4: freeway, non-congested (speed > 40 mph), & long space gaps (space > 10 m);
5) Type 5: local, congested (speed <= 20 mph), & short space gaps (space <= 10 m);
6) Type 6: local, congested (speed <= 20 mph), & long space gaps (space > 10 m);
7) Type 7: local, non-congested (speed > 20 mph), & short space gaps (space <= 10 m);
8) Type 8: local, non-congested (speed > 20 mph), & long space gaps (space > 10 m);
9) Type 9: others, e.g., no front vehicle.

This study uses the 40 mph and 20 mph as the congestion threshold for freeway and local roadway separately. As the duration of congested period are triggered when the vehicle average speed of weekday peak time drops below 45 mph, therefore, this study uses 40 mph (close to mph) to define the congestion threshold for freeway. Given the common speed limit of local roadway is between 35 and 40 mph, this study defines the congestion threshold for non-freeway as 20 mph which is also in the range of school zone speed limit, indicating it is a lower speed area. Figure 3.2 shows the conceptual framework which indicates the response variable and key independent variables involved in this study.

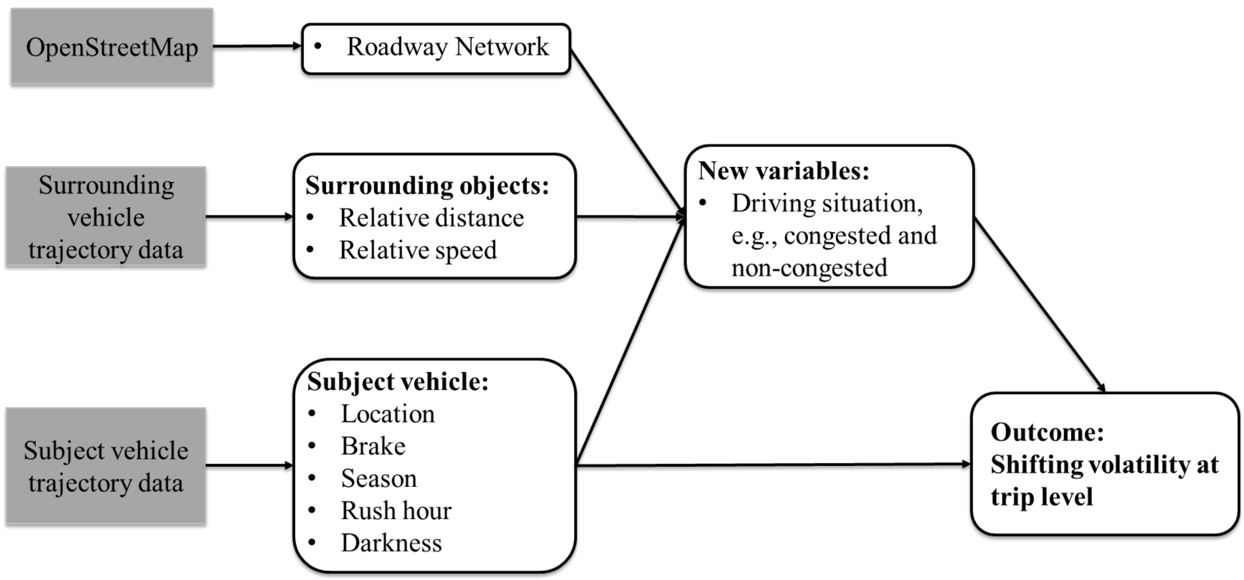


Figure 3. 2 Conceptual framework

3.3.2 Lateral shifting volatility

The critical part of this study is to develop a measurement to understand the variability in instantaneous lateral driving decisions from the aspect of lateral displacement. To explore the volatile driving decision, this study proposes the shifting volatility measured by coefficient of variations (COV), defined as the ratio of standard deviation to mean, to quantify fluctuations in the instantaneous lateral displacement [43]. The sufficient geo-referenced trajectories data collected from connected vehicles enable the analysis. Since the right and left shifting volatility might be different, two types of shifting volatility are measured in this study. The formulas for COV calculation are shown below:

Right side: $COV_{right} = \frac{Std.Dev._{right}}{Mean_{right}}$ (1)

Left side: $COV_{left} = \frac{Std.Dev._{left}}{Mean_{left}}$ (2)

3.4 RESULTS

3.4.1 Descriptive statistics

Table 3.1 shows the descriptive statistics for key variables used for modeling at the aggregated trip level. After delete missing data and error check, 167 trips are used for analysis. On average, the right shifting volatility is 0.867, while the left shifting volatility is a little higher which is 0.883. On average, nearly 51% of time that the subject vehicle is following a front vehicle in the travel path. The average number of vehicles on the right or left side is 1.5. The average speed is 0.15 m/s lower than the front vehicle. As mentioned above, this study separates the driving environment into nine categories to get an in-depth understanding of driving situation. In freeway, 9.2% of time that the subject vehicle can maintain relative satisfied speed and keep proper space gaps with front vehicle. Nearly 1.4% of time that subject vehicle follows a front vehicle with short space gaps (<= 10 m), while 1.2% of time the speed is lower than 40 mph, indicating the speed of subject vehicle is restricted which is recognized as a congested driving environment in freeway. Note, there is 0.2% of time that the subject vehicle can keep relative high speed but the subject vehicle still keeps close to the front vehicle, which indicate a relative dangerous situation. 2.4% of time that the subject vehicle is under congested environment (speed lower than 40 mph) while keeps far away from the front vehicle, indicate a conservative driving behavior of subject vehicle. In local roadway, 25.6% of time that the subject vehicle can maintain the relative satisfied speed with proper space gaps with front vehicle. However, there is

Table 3. 1 Descriptive statistics

	Variables	N	Mean	Std.Dev.	Min	Max
Shifting volatility	Cov_right	167	0.867	0.186	0.231	1.330
	Cov_left	167	0.883	0.212	0.171	1.839
Vehicle maneuvering	ABS state	167	0.626	0.480	0.000	1.000
	Brake (engaged) (%)	167	18.5%	13.0%	0.0%	58.7%
	Headlight (engaged) (%)	167	19.4%	35.1%	0.0%	100.0%
	Stable control (engaged) (%)	167	19.5%	39.5%	0.0%	100.0%
	Vehicle wiper (engaged) (%)	167	3.6%	0.187	0.0%	100.0%
Contextual factors	An exit on the left side (engaged) (% *1000)	167	8.5%	0.227	0.0%	127.0%
	An exit on the right side (engaged) (% *1000)	167	35.1%	44.3%	0.0%	236.0%
	Season (1-spring, 0-autumn)	167	0.471	0.436	0	1
	Darkness	167	0.077	0.251	0	1
	Rush hour	167	0.564	0.455	0.000	1.000
Surrounding objects	% of time with front vehicle in path (%)	167	51.7%	0.204	3.0%	100.0%
	Average of surrounding object on right side	167	1.558	0.232	1.077	2.649
	Average of surrounding object on left side	167	1.501	0.259	1.107	2.622
	Average relative speed to front object(m)	167	-0.154	0.594	-2.108	2.557
Subject vehicle driving environment	% of time in: freeway, speed <= 40 mph & distance to front vehicle < =10 m	167	1.2%	0.037	0.0%	22.5%
	% of time in: freeway, speed <= 40 mph & distance to front vehicle > 10 m	167	2.4%	7.0%	0.0%	56.8%
	% of time in: freeway, speed > 40 mph & distance to front vehicle <=10 m	167	0.2%	1.3%	0.0%	16.4%
	% of time in: freeway, speed > 40 mph & distance to front vehicle > 10 m	167	9.2%	13.6%	0.0%	61.9%
	% of time in: local, speed <= 20 mph & distance to front vehicle <= 10 m	167	8.4%	10.0%	0.0%	74.0%
	% of time in: local, speed <= 20 mph & distance to front vehicle > 10 m	167	3.2%	3.4%	0.0%	20.2%
	% of time in: local, speed > 20 mph & distance to front vehicle <= 10 m	167	0.9%	3.1%	0.0%	28.8%
	% of time in: local, speed > 20 mph & distance to front vehicle > 10 m	167	25.6%	18.1%	0.0%	85.8%
	Others, e.g., no front vehicle	167	48.9%	0.213	2.3%	100.0%

still 0.9% of time that the subject vehicle can maintain proper speed but follow the front vehicle with low space gaps.

3.4.2 Model results

Table 3.2 shows the modeling results for testing the correlation of right shifting volatility and left shifting volatility with related contributing factors. The goodness-of-fit is reasonable for right side shifting volatility model, while not significant for left side shifting volatility model. Therefore, the interpretation is mainly based on the results of right side shifting volatility model. As expected, the various driving situations have shown significant correlations with right side shifting volatility (at 5% level) and the signs of estimated parameters are expected. Note, the analysis is applied at the aggregated trip level; thus, variables significant at the aggregated trip level might not be necessary significant at the disaggregated level.

The modeling results shows that most of subject vehicle driving situations are statistically significantly associated with lower shifting volatility, compared with the based condition of the subject vehicle being traveling with proper speed but keeping low space gaps in freeway. Traveling with low speed and keeping high space gaps in local roadway has the lowest association with the shifting volatility. The modeling results also show that the shifting volatility is statistically significantly higher during autumn and non-peak hour period. No significant correlations are found regards to vehicle maneuvering and surrounding objects.

The magnitudes and signs of the estimated coefficient in subject vehicle driving environment are of interests. The presence of front vehicle and the subject vehicle speed are key contributing factors to lateral shifting volatility in the resulting model. Compared to the base condition of subject vehicle traveling at freeway with relative high speed (>40 mph) and short space gaps with front vehicle (<=10 m), the subject vehicle is less likely to be volatile in lateral shifting, especially when the subject vehicle traveling at local roadway with low speed but still be far away from the front vehicle (type 6). Under type 6 driving situation, although the subject vehicle is traveling under the congested environment at local roadway, the driver still keeps large space with the front vehicle, which indicates the subject vehicle can decelerate and come to a full stop with enough space; as a result, they might be less likely to be involved in a risk situation as they are less volatile. A logical explanation for this is that the characteristics that increase driver comfort levels, including long distance to front vehicle, cause the driver can maintain a relative low heightened awareness and focus. Rush hour is associated with lower shifting volatility. Under rush hour period, the subject vehicle is traveling with low speed and surrounded with more surrounding vehicles, which is similar to the type 1, type 2, type 5 or type 6 situations; thus, the subject vehicle might be less volatile.

These findings have potential implications regarding associations of subject vehicle driving environment with lateral shifting volatility as previous studies indicates that high volatility is associated with a higher chance of crash. The onboard device can record the historical lane keeping behavior of the subject vehicle, then the corresponding shifting volatility for each subject driver can be computed and be compared with other drivers. Thus, the driver with high shifting volatility record will receive warnings or control assistance to help them make informed lane departure decisions to avoid high risk situations, such as lane departure crashes.

Table 3. 2 Linear regression modeling results

Variables		Shifting volatility (right)		Shifting volatility (left)	
		β	P value	β	P value
Vehicle maneuvering	Brake (engaged) (%)	-0.072	0.542	0.011	0.940
	Vehicle wiper (engaged) (%)	0.009	0.908	0.023	0.807
Contextual factors	An exit on the left side (engaged) (% *1000)	-0.003	0.965	0.058	0.453
	An exit on the right side (engaged) (% *1000)	0.032	0.322	0.020	0.610
	Season (1-spring, 0-autumn)	-0.098	0.005**	0.004	0.931
	Darkness	-0.052	0.421	0.130	0.099*
	Rush hour	-0.077	0.033**	0.047	0.279
Surrounding objects	Average of surrounding object on right side	0.010	0.882	0.015	0.858
	Average of surrounding object on left side	-0.056	0.364	0.047	0.530
	Average relative speed to front object(m)	-0.018	0.473	0.040	0.193
Subject vehicle driving environment (base: Type 3: % of time in: freeway, speed > 40 mph & distance to front vehicle <=10 m)	Type 1: % of time in: freeway, speed <= 40 mph & distance to front vehicle < =10 m	-1.770	0.144	1.356	0.353
	Type 2: % of time in: freeway, speed <= 40 mph & distance to front vehicle > 10 m	-2.407	0.033**	0.702	0.604
	Type 4: % of time in: freeway, speed > 40 mph & distance to front vehicle > 10 m	-2.416	0.031**	1.216	0.366
	Type 5: % of time in: local, speed <= 20 mph & distance to front vehicle <= 10 m	-2.393	0.030**	1.400	0.291
	Type 6: % of time in: local, speed <= 20 mph & distance to front vehicle > 10 m	-3.027	0.013**	1.733	0.233
	Type 7: % of time in: local, speed > 20 mph & distance to front vehicle <= 10 m	-1.780	0.137	1.409	0.329
	Type 8: % of time in: local, speed > 20 mph & distance to front vehicle > 10 m	-2.207	0.045**	1.173	0.375
	Type 9: others, e.g., no front vehicle	-2.401	0.029**	1.267	0.338
Constant		3.382	0.002**	-0.509	0.699
Statistic summary	Sample size	167		167	
	Prob. > F	0.016**		0.630	
	Adjusted R^2	0.090		0.000	

3.5 LIMITATIONS

Several variables in the data were missing or otherwise unusable. Cruise control data did not appear reliable. Several periods of sensor failures were observed within trips. Some effort was made by the author to identify scenarios that were more likely to result in missing data, but with the limited variables available during this failure periods, this proved difficult. These missing data periods appeared to be random, but if they were related to specific circumstances within trips, potential for the introduction of considerable error would exist. Additionally, the amount of environmental data was limited, resulting in difficulty eliminating environmental effects from biasing the results.

3.6 CONCLUSIONS

This study proposes a measure called shifting volatility to quantify the variability in instantaneous lateral displacement. Correlations between lateral shifting volatility and related factors are analyzed, specifically between lateral volatility and driver comfort. Using sufficient trajectory data called BSMs collected from vehicles participating in Safety Pilot Model Deployment (SPMD) in Michigan, this study measures shifting volatility by quantifying the fluctuations in instantaneous lateral displacement through the coefficient of variation (COV), defined as the ratio of standard deviation to mean.

The resulting model identifies relationships that could inform roadway agencies of characteristics that could help reduce the number of roadway departure crashes, as well as give them a better understanding about when a driver is most likely to cause a lane departure crash. Based on the model, roadway type, vehicle speed and distance to front target vehicle correlate with lateral shifting volatility. The results reinforce the importance of driving situations in areas prone to roadway departure crashes. Additional lane departure warning system deployments may glean some more useful information. These results indicate that lateral volatility, which could potentially lead to a lane departure, is at its greatest risk when the subject vehicle is driving at relative high speeds and keeps low space gaps with the vehicle in front of it.

CHAPTER 4 GOSSIP PATTERNS IN INSTANTANEOUS DRIVING DECISIONS DURING CAR FOLLOWING EVENTS

This chapter is a revised version to be submitted by Meng Zhang, and Asad Khattak

ABSTRACT

This study proposes a new concept called “Gossip effect” to capture the peer influence of surrounding vehicle on the instantaneous driving decisions of subject vehicle. This study analyzes the two-step driving decision procedure is: 1) micro-level driving decision defined by acceleration and deceleration, and 2) aggregated event-level driving decision captured by subject vehicle making a lane change or not during a car following event. The unique aspect of this study is that it establishes a new framework to understand the naturalistic instantaneous driving decision of subject vehicle under car following scenario, which considers the psychological factors, using high resolution geo-referenced trajectory data. The data used for analysis are Basic Safety Messages (BSMs) sent by vehicle, at a 10 Hz frequency, and received by roadside equipment participating in the Safety Pilot Model Deployment (SPMD) program in Ann Arbor, Michigan. These BSMs describe a vehicle’s operation and performance such as vehicle speed, acceleration, relative distance and speed to front vehicle, which enables the analysis of driving decision at the micro-level. A sub-trajectory data representing 1,940,678 BSMs records from 192 trips by 64 vehicles is used for analysis. This study further explores the correlations of driving decisions with driving situations. The results show that the subject vehicle averagely is more likely to accelerate as front vehicle to achieve relative high speed. However, they are less likely to accelerate as front vehicle under complex and congested driving situations.

4.1 INTRODUCTION

To understand and model group behaviors and peer influence, the study explores the roles of psychological and sociological factors. Each driver, as an integral part of a network of vehicles, is assumed to obey simple rules: a) attempt to maintain internal consistency, by executing the optimum policy consistent with his/her utility measures, and b) simultaneously strive to attain social consensus. An indicative example of peer influence can be expressed by the acceleration probability of a subject vehicle when surrounding vehicles are speeding up. The subject vehicle might follow the decision of surrounding vehicle but still keep the internal cognitive equilibrium in order. On the other hand, given the scenario that surrounding vehicles are decelerating, the subject vehicle might decelerate as s/he may suppose that there is some trouble ahead, such as a crash or police control. However, the reason that the subject vehicle makes the deceleration decisions is only because s/he wants to demonstrate that s/he is not a “worse” driver than the others. Studies have tried to explore the psychological point of view for car following models [55].

Given the front vehicle in the travel path has more influence on subject vehicle, this study aims to explore the peer influence of front vehicle on the driving decisions of subject vehicle. A new “Gossip” concept which capture thus peer influence is proposed. The original gossip concept refers to people can spread information by talking to other people. This sort of information propagation can be applied to instantaneous driving decisions, that is the driving decisions of subject vehicle can be influenced by front vehicles. In addition, a two-step driving decisions procedure is analyzed: 1) micro-level driving decision defined by vehicle acceleration and deceleration, and 2) aggregated event-level driving decision captured by subject vehicle making a lane change or not during a car following event. While the driving decisions are correlated with surrounding driving situations, this study also extracts different driving situation based on relative distance and speed to each surrounding vehicle information embedded in massive trajectory data to explore their correlation with driving decisions.

4.2 METHOD

4.2.1 Data source and conceptual framework

Basic Safety Messages (BSMs)

The data used for analysis are Basic Safety Messages (BSMs) archived in Driving Dataset collected through the Safety Pilot Model Deployment (SPMD) program in Ann Arbor, Michigan. The field test includes 75 miles instrumented roadway and 26 roadside unites are installed, which are able to communicate with vehicles equipped with data acquisition systems (DAS). These data is available to public via the Research Data Exchange (RDE, available from: http://www.its-rde.net/) maintained by the U.S. Department of Transportation. These BSMs are sent by vehicles, at a 10 Hz frequency, and collected by the roadside equipment participating in the SPMD program.

Two sub-dataset archived in Driving Dataset are used for analysis: 1) HV_Primary, which describes the operation and performance of subject vehicle (reported at 10 Hz frequency), including geographic position (e.g., latitude and longitude), vehicle motion (e.g., speed and acceleration), onboard device records of lane tracking information (e.g., vehicle meets and crosses the lane boundary, and distance between vehicle centerline to lane boundary), and driving context (e.g., time stamp), and 2) HV_radar, which describes the information of surrounding vehicles (reported at 10 Hz frequency), including relative distance and speed to each surrounding vehicle at each time stamp. More detailed descriptions of variables involved in the dataset can be found on the SPMD Sample Data Handbook [29].

Given high-resolution of GPS data, the whole data set contains 83,384,195 records generate from 14,315 trips by 64 vehicles, which is very large. Due to the computational limitations, this study randomly select three trips (trip duration is longer than 10 minutes) from each driver for analysis. Therefore, this study gets 1,940,678 BSMs records from 192 trips by 64 vehicles. Since this study focuses on peer influence of front vehicle on the subject vehicle, this study only extracts scenario where a subject vehicle is following a front vehicle. In addition, this study aggregates the raw data every 1 second to address the common noise problems of GPS data. Finally, this study gets 13,458 records representing 224 hours of car following scenario for analysis.

New defined driving situations

While the driving behavior is highly correlated with surrounding driving situation, this study also extracts driving situations information embedded in trajectory data. As shown in Figure 4.1, different driving situations can be identified based on the location of surrounding vehicles. The driving decision of subject vehicle is assumed to be different when subject vehicle keeps far away from and close to the surrounding vehicle. In order to differentiate the congested and non-congested driving situations, this study use gaps equals to 10 meters as the congested threshold which indicates whether the subject vehicle has enough space to operate the vehicle. Based on the information of number of vehicles and relative distance to front vehicle in the travel path, on the right and left side, eight driving situations are created:

- Type 1: F=1 (distance to front vehicle <= 10m), L=1 (distance to left side vehicle: lon <= 10), R=1 (distance to right side vehicle: lon <= 10)
- Type 2: F=1 (distance to front vehicle <= 10m), L=1 (distance to left side vehicle: lon <= 10), R=0 (distance to right side vehicle: lon > 10)
- Type 3: F=1 (distance to front vehicle <= 10m), L=0 (distance to left side vehicle: lon > 10), R=1 (distance to right side vehicle: lon <= 10)
- Type 4: F=1 (distance to front vehicle <= 10m), L=0 (distance to left side vehicle: lon > 10), R=1 (distance to right side vehicle: lon > 10)
- Type 5: F=0 (distance to front vehicle > 10m), L=1 (distance to left side vehicle: lon <= 10), R=1 (distance to right side vehicle: lon <= 10)
- Type 6: F=0 (distance to front vehicle > 10m), L=1 (distance to left side vehicle: lon <= 10), R=0 (distance to right side vehicle: lon > 10)
- Type 7: F=0 (distance to front vehicle > 10m), L=0 (distance to left side vehicle: lon > 10), R=1 (distance to right side vehicle: lon <= 10)
- Type 8: F=0 (distance to front vehicle > 10m), L=0 (distance to left side vehicle: lon > 10), R=1 (distance to right side vehicle: lon > 10)

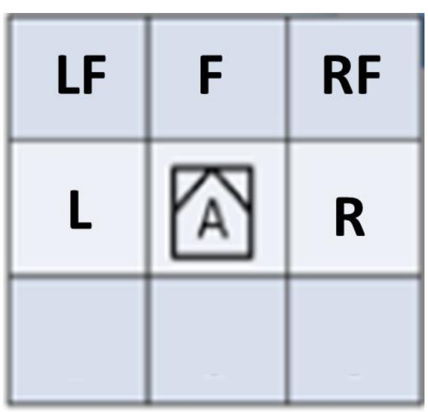


Figure 4. 1 Definition of driving situations

4.2.2 Model Structure

This study explores how the driving decision of front vehicle influences the driving decisions of subject vehicle. Four driving decisions are proposed in this study: 1) Front vehicle accelerate, and subject vehicle accelerate; 2) Front vehicle accelerate, and subject vehicle decelerate; 3) Front vehicle decelerate, and subject vehicle accelerate; and 4) Front vehicle decelerate, and subject vehicle decelerate. Considering the multinomial nature of driving decisions, this study applies multinomial logit model. In the multinomial logit model, the probability of each driving decision can be written as:

$$\Pr(Y=1)=\frac{\exp(X\beta_{(1)})}{\exp(X\beta_{(1)})+\exp(X\beta_{(2)})+\cdots+\exp(X\beta_{(n)})} \tag{1}$$

$$\Pr(Y=2)=\frac{\exp(X\beta_{(2)})}{\exp(X\beta_{(1)})+\exp(X\beta_{(2)})+\cdots+\exp(X\beta_{(n)})} \tag{2}$$

$$\ldots\ldots$$

$$\Pr(Y=i)=\frac{\exp(X\beta_{(i)})}{\exp(X\beta_{(1)})+\exp(X\beta_{(2)})+\cdots+\exp(X\beta_{(n)})} \tag{3}$$

Where,

Y is the driving decision of subject vehicle;

$\beta_{(i)}$ is a set of estimated coefficients for the i^{th} driving decision, i=1,2,3,4.

X is a vector of explanatory variables, such as driving environment;

4.3 RESULTS

4.3.1 Distribution of subject vehicle motion

Figure 4.2 shows the distributions of subject vehicle acceleration, speed and distance to front vehicle. While the front vehicle has more influence on the driving decisions, which is captured by acceleration and deceleration, of subject vehicle in the longitudinal direction, the longitudinal acceleration is considered in this study. The red points indicate the acceleration of subject vehicle is over the 95th percentile value, which is volatile [56]. The figure shows that the longitudinal acceleration is volatile when distance to front vehicle is short, which indicates the subject vehicle is more likely to be aggressive, as shown in Figure 4.2 a (red points). Figure 4.2 (b) represents the changes in subject vehicle acceleration based on the speed difference with front vehicle. It shows that the subject vehicle is more likely to accelerate when front vehicle's speed is higher, while the subject vehicle is less likely to accelerate when front vehicle's speed is much higher. The results indicate that the influence of front vehicle decrease along with the increase in the speed difference (Vf-Vs) between front vehicle and subject vehicle. Overall, the driving decision is highly influenced by front vehicle, which is analyzed in this study.

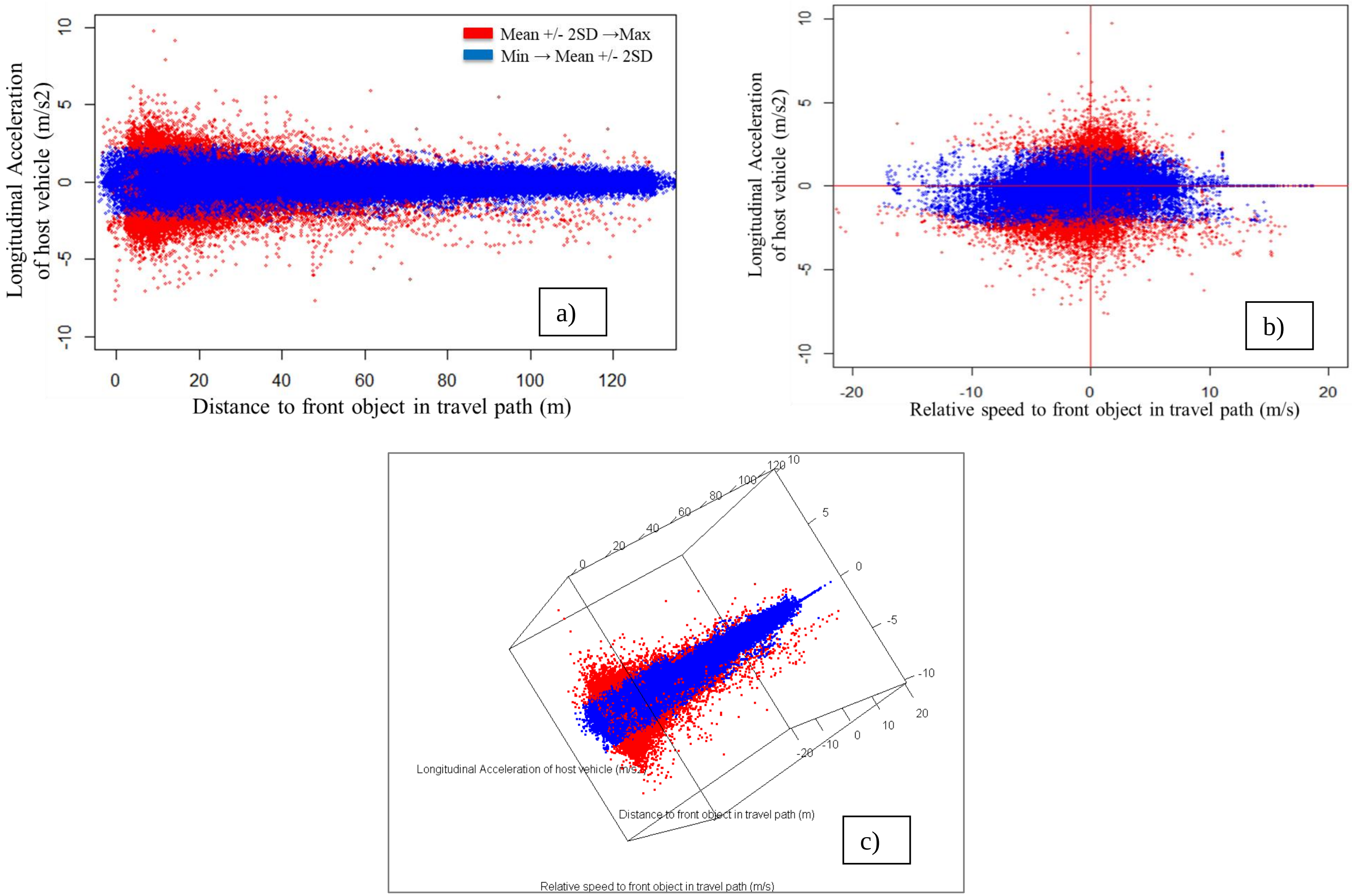


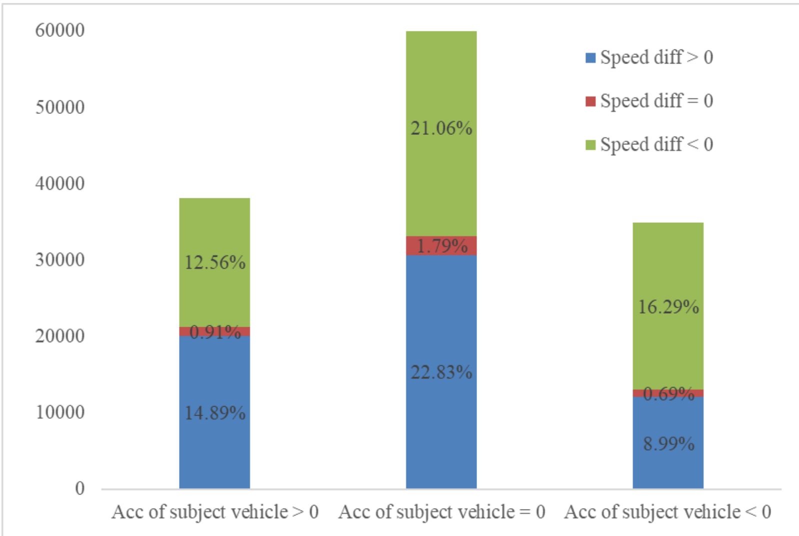
Figure 4. 2 Distributions of speed, acceleration and distance to front vehicle

4.3.2 Gossip patterns in instantaneous driving decisions

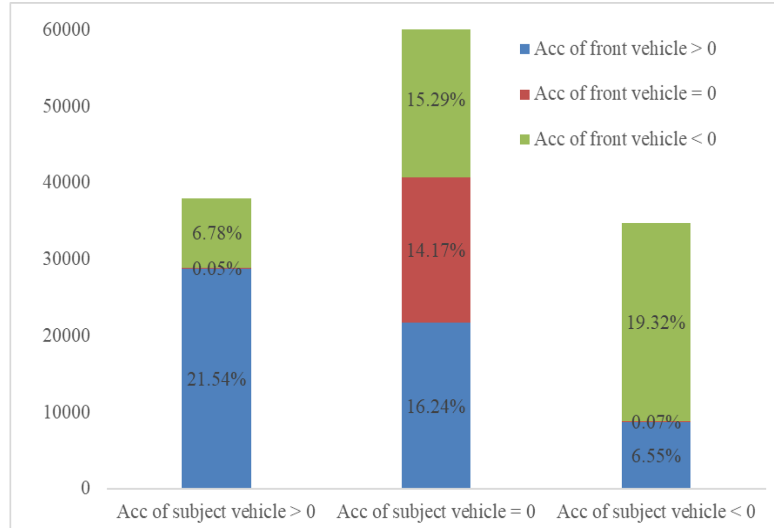
Figure 4.3 shows the driving decisions distributions of subject vehicle based on speed difference with front vehicle (a & c) and front vehicle acceleration (b & d). As this study focuses on the peer influence of front vehicle driving decision on subject vehicle driving decisions, more attention is paid to Figure 4.3 (b & d). On average, the subject vehicle is more likely to follow the driving decisions (b & d) of front vehicle but not the driving status (a & c) of front vehicle. It shows that the subject vehicle is more likely to accelerate as the front vehicle (21.54%), especially when the speed of front vehicle is higher (75.94%), which indicates that a higher speed front driver who is accelerating has more influence on the driving decisions of the subject vehicle. It is expected as the subject vehicle might want to accelerate to achieve a high speed as front vehicle. To differentiate car following model and the gossip concept, this study compares the General Motors (GM) car following model with gossip concept, as shown in Table 4.1. The common GM car following model explores the driving decisions of subject vehcile based on perception (speed differnce with front vehicle). This study investigates the driving decisions based on the decisions of front vehicle.

Table 4. 1 Comparision between General Motors (GM) car following model with gossip concept

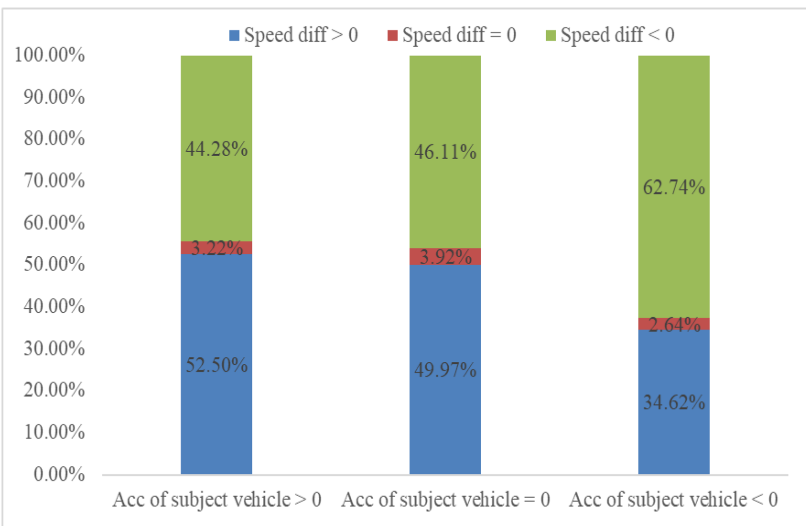
	GM Car following model	Proposed gossip concept
Similarities	Following front vehicle	Following front vehicle
	Influence on decision of subject vehicle	Influence on decision of subject vehicle
Differences	Perception (e.g., speed difference) to decision, $\ddot{X}_{n+1}(t+\Delta t) = \alpha\left[\dot{X}_n(t)\text{-}\dot{X}_{n+1}(t)\right]$	Social influence on acceleration decisions – decision to decision, $\bar{r}_i(t+\Delta t) = \frac{1}{\lvert N_i(t)\rvert}\sum_{j\in N_i(t)} \bar{r}_j(t+\Delta t)$
	Theoretical-physics driven	Theoretical-peer influence + data-driven
	Lane change not integrated in decision (separate model)	Lane change integrated in decision
	Subject vehicle should keep safe gaps with front vehicle	Driving decisions under naturalistic driving environment



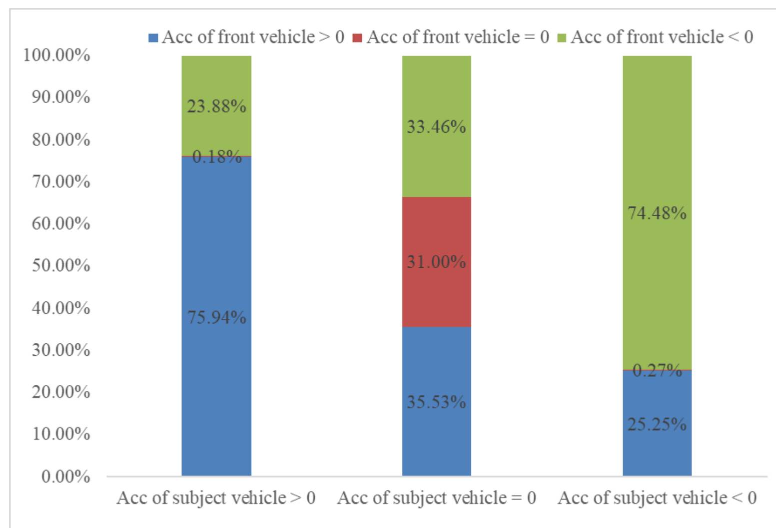
a) Total cases



b) Cases that front vehicle speed is higher



c) Total cases



d) Cases that front vehicle speed is higher

Figure 4. 3 Subject vehicle driving decision based on relative speed to front vehicle (a & b) and front vehicle acceleration (c & d)

4.3.3 Descriptive statistics

Table 4.2 shows descriptive statistics of key variables used for analysis. This study defines four types of driving decisions: 1) Front vehicle accelerate, and subject vehicle accelerate; 2) Front vehicle accelerate, and subject vehicle decelerate; 3) Front vehicle decelerate, and subject vehicle accelerate; and 4) Front vehicle decelerate, and subject vehicle decelerate. On average, nearly 56% of time that the subject vehicle follows the driving decision of front vehicle. Of these, 32.8% of time subject accelerates as front vehicle while 23.5% of time they decelerate as front vehicle, which indicates subject vehicle is more likely to follow the acceleration decision of front vehicle. Nearly 44% of time that subject vehicle does not follow the driving decision of front vehicle. Of these, 24.5% of time subject vehicle accelerates but front vehicle decelerates, which is higher than the time (19.2%) subject vehicle decelerates while front vehicle accelerates. On average, the subject vehicle keeps proper distance with front vehicle. Most of the time, the subject vehicle stays relative far away (distance to front vehicle is longer than 10 m) from the front vehicle (91.4%), while only 8.6% of time following the front vehicle closely. Of these following close to front vehicle driving situations, 6.9% of time there is no right or left side vehicles. Table 4.2 also shows the driving decisions of subject vehicle under two scenarios: front vehicle speed is higher and front vehicle speed is lower. On average, the subject vehicle is more likely to accelerate as front vehicle when front vehicle speed is higher (43.5%) compared with front vehicle speed is lower (22.8%).

Table 4. 2 Descriptive statistics of key variables

Variable		Total data (N=13,458)		Partial data: Front vehicle speed is higher (N= 6,478)		Partial data: Front vehicle speed is lower (N=6,971)			
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Min	Max
Follow front vehicle driving decision		0.563	0.496	0.586	0.493	0.542	0.498	0	1
Subject vehicle driving decision	Front vehicle acc. & subject vehicle acc.	0.328	0.469	0.435	0.496	0.228	0.420	0	1
	Front vehicle acc. & subject vehicle dec.	0.192	0.394	0.188	0.391	0.196	0.397	0	1
	Front vehicle dec. & subject vehicle acc.	0.245	0.43	0.226	0.418	0.262	0.440	0	1
	Front vehicle dec. & subject vehicle dec.	0.235	0.424	0.151	0.358	0.314	0.464	0	1
Driving situation	Type 1: F=1 (<= 10m), L=1 (lon <= 10), R=1 (lon <= 10)	0.001	0.039	0.001	0.037	0.002	0.040	0	1
	Type 2: F=1 (<= 10m), L=1 (lon <= 10), R=0 (lon > 10)	0.006	0.076	0.005	0.069	0.007	0.083	0	1
	Type 3: F=1 (<= 10m), L=0 (lon > 10), R=1 (lon <= 10)	0.01	0.099	0.009	0.093	0.011	0.104	0	1
	Type 4: F=1 (<= 10m), L=0 (lon > 10), R=1 (lon > 10)	0.069	0.253	0.052	0.222	0.084	0.278	0	1
	Type 5: F=0 (> 10m), L=1 (lon <= 10), R=1 (lon <= 10)	0.004	0.06	0.002	0.050	0.005	0.068	0	1
	Type 6: F=0 (> 10m), L=1 (lon <= 10), R=0 (lon > 10)	0.04	0.196	0.036	0.186	0.043	0.204	0	1
	Type 7: F=0 (> 10m), L=0 (lon > 10), R=1 (lon <= 10)	0.066	0.249	0.054	0.225	0.078	0.269	0	1
	Type 8: F=0 (> 10m), L=0 (lon > 10), R=1 (lon > 10)	0.804	0.397	0.841	0.366	0.770	0.421	0	1

Note: F=1 (<= 10m) referes to subject vehicles keep 10 meters away from front vehicle,
L=1 (lat <= 5, lon <= 10) refers to there is a left surrounding vehicle whitin a 10 meter range,
R=1 (lat <= 5, lon <= 10) refers to there is a right surrounding vehicle whitin a 10 meter range.

As mentioned early, this study explores two-step decisions from micro-level and aggregated event-level. This study further explores the driving decisions at the aggregated car following events level, that is a vehicle can make a lane change or continue to follow front vehicles. The lane change behaviors can be identified based on method proposed in Chapter 2. Table 4.3 shows the descriptive statistics for subject vehicle makes a lane change and continues to follow front vehicle. The average speed of subject vehicle who makes a lane change (15.48 mph) is higher than these who continue to follow front vehicle (14.10 mph), which indicates the subject vehicle might want to achieve high speed by making a lane change.

Table 4. 3 Descriptive statistics for comparison between lane change and non-lane change event

	Variables	N	Mean	Std.Dev.	Min	Max
Subject vehicle makes a lane change	Front vehicle speed (mph)	59682	15.482	10.59	0	26.47
	Subject vehicle speed (mph)	59682	15.751	10.66	0	36.81
	Speed difference (mph)	59682	-0.269	2.805	-22.583	15.86
Subject vehicle continues to follow front vehicle	Front vehicle speed (mph)	134588	14.102	10.12	0	39.9
	Subject vehicle speed (mph)	134588	14.386	10.14	0	41.99
	Speed difference (mph)	134588	-0.284	2.916	-26.789	15.86

4.3.4 Modeling results

Table 4.4 shows the multinomial logit modeling results using total data and separated data (front vehicle speed is higher and front vehicle speed is lower). Although the goodness-of-fit is on the low side, most variables have shown significant correlation with the response variables. The correlations of driving decision differ under two scenarios: front vehicle speed is higher and front vehicle speed is lower.

In the total data model, the signs and magnitudes of constant value indicate that the subject vehicle is more likely to accelerate as front vehicle or the subject vehicle accelerates but front vehicle decelerates, compared to the base of subject vehicle decelerating as front vehicle. It seems the subject vehicles are more likely to accelerate, especially when front vehicle is accelerating, which indicates that the acceleration decisions of front vehicle have a larger influence on the driving decision of subject vehicle. The total data model also indicates that the subject vehicle is less likely to decelerate when front vehicle is accelerating, compared with subject vehicle decelerating as front vehicle. The driving performance or operation of subject vehicle is restricted when front vehicle is decelerating, therefore, the subject vehicle might need to decelerate to avoid a crash with front vehicle.

The results of driving situations from total data model shows that comparing to base of type 8 driving situation which indicates that subject vehicle keeps relative far away from the front vehicle and without surrounding vehicle, the subject vehicle is less likely to accelerate as front vehicle, especially when the distance to front vehicle is lower (<=10 m) and surrounded vehicles on left and right sides, which indicates a complex and congested driving situation. One possible reasons might be that the subject vehicle might be distracted when driving situation is more complex, therefore, the subject vehicle is less likely to accelerate as front vehicle.

The results of separated model show interesting results. The signs and magnitudes of constant values indicates that comparing to the base of subject vehicle decelerating as front vehicle, the subject vehicle is more likely to accelerate and less likely to deceleration when front vehicle speed is higher, while opposite when front vehicle speed is lower. The results are consistent with the expection line. In real driving environment, drivers might want to achieve relative high speed, therefore, they are more likely to accelerate as high speed front vehicle. Similarity, the subject vehicle is less likely to accelerate as front vehicle under complex and congested driving situations not matter front vehicle speed is high or not.

Table 4. 4 Multinomial Logit modeling results for driving decisions

Variables (response variable = driving decisions)		Total data		Separated data			
				Front vehicle speed is higher		Front vehicle speed is lower	
		β	P-value	β	P-value	β	P-value
Front vehicle acc. & subject vehicle acc.							
Driving situation Base: Type 8: F=0 (> 10m), L=0 (lon > 10), R=1 (lon > 10)	Type 1: F=1 (<= 10m), L=1 (lon <= 10), R=1 (lon <= 10)	-1.800	0.005***	-2.513	0.025**	-1.119	0.158
	Type 2: F=1 (<= 10m), L=1 (lon <= 10), R=0 (lon > 10)	-1.080	0.000***	-1.295	0.002***	-0.830	0.043**
	Type 3: F=1 (<= 10m), L=0 (lon > 10), R=1 (lon <= 10)	-1.163	0.000***	-1.378	0.000***	-0.857	0.01***
	Type 4: F=1 (<= 10m), L=0 (lon > 10), R=1 (lon > 10)	-0.449	0.000***	-0.599	0.000***	-0.118	0.306
	Type 5: F=0 (> 10m), L=1 (lon <= 10), R=1 (lon <= 10)	-0.925	0.012**	-0.028	0.966	-1.466	0.019**
	Type 6: F=0 (> 10m), L=1 (lon <= 10), R=0 (lon > 10)	-0.199	0.089*	0.161	0.430	-0.388	0.023**
	Type 7: F=0 (> 10m), L=0 (lon > 10), R=1 (lon <= 10)	-0.184	0.052*	-0.119	0.469	0.003	0.980
	Constant	0.414	0.000***	1.127	0.000***	-0.269	0.000***
Front vehicle acc. & subject vehicle dec.							
Driving situation Base: Type 8: F=0 (> 10m), L=0 (lon > 10), R=1 (lon > 10)	Type 1: F=1 (<= 10m), L=1 (lon <= 10), R=1 (lon <= 10)	-0.733	0.169	-0.270	0.704	-1.659	0.118
	Type 2: F=1 (<= 10m), L=1 (lon <= 10), R=0 (lon > 10)	-1.070	0.002***	-1.225	0.02**	-0.966	0.035**
	Type 3: F=1 (<= 10m), L=0 (lon > 10), R=1 (lon <= 10)	-0.497	0.032**	-0.387	0.264	-0.626	0.054*
	Type 4: F=1 (<= 10m), L=0 (lon > 10), R=1 (lon > 10)	-0.513	0.000***	-0.442	0.009*	-0.515	0.000***
	Type 5: F=0 (> 10m), L=1 (lon <= 10), R=1 (lon <= 10)	-0.455	0.227	0.018	0.981	-0.467	0.300
	Type 6: F=0 (> 10m), L=1 (lon <= 10), R=0 (lon > 10)	-0.220	0.106	-0.239	0.345	-0.100	0.542
	Type 7: F=0 (> 10m), L=0 (lon > 10), R=1 (lon <= 10)	0.051	0.619	0.059	0.749	0.139	0.270
	Constant	-0.143	0.000***	0.269	0.000***	-0.420	0.000***
Front vehicle dec. & subject vehicle acc.							
Driving situation Base: Type 8: F=0 (> 10m), L=0 (lon > 10), R=1 (lon > 10)	Type 1: F=1 (<= 10m), L=1 (lon <= 10), R=1 (lon <= 10)	-15.234	0.978	-15.853	0.988	-14.429	0.978
	Type 2: F=1 (<= 10m), L=1 (lon <= 10), R=0 (lon > 10)	-1.249	0.000***	-2.379	0.002***	-0.747	0.048**
	Type 3: F=1 (<= 10m), L=0 (lon > 10), R=1 (lon <= 10)	-1.040	0.000***	-1.318	0.002***	-0.843	0.008***
	Type 4: F=1 (<= 10m), L=0 (lon > 10), R=1 (lon > 10)	-0.479	0.000***	-1.043	0.000***	-0.164	0.142
	Type 5: F=0 (> 10m), L=1 (lon <= 10), R=1 (lon <= 10)	-1.510	0.003***	-14.393	0.981	-1.095	0.032**
	Type 6: F=0 (> 10m), L=1 (lon <= 10), R=0 (lon > 10)	-0.175	0.161	-0.003	0.990	-0.178	0.246
	Type 7: F=0 (> 10m), L=0 (lon > 10), R=1 (lon <= 10)	-0.175	0.083*	-0.234	0.209	-0.060	0.622
	Constant	0.123	0.000***	0.507	0.000***	-0.128	0.000***
Base: Front vehicle dec. & subject vehicle dec.							
Statistical summary	Sample size	13458		6478		6971	
	Likelihood at 0	-18402.032		-8417.422		-9557.123	
	Likelihood at β	-18336.766		-8369.340		-9525.897	
	Prob. > χ^2	130.530		96.160		62.450	
	Pseudo R^2	0.004		0.006		0.003	

4.5 CONCLUSIONS

This study contributing by establishing a new framework to understand the instantaneous driving decisions of subject vehicle in car following scenario. A “Gossip” concept which captures the peer influence of surrounding vehicles on instantaneous driving decisions of subject vehicle is proposed. Instead of exploring the driving decision of subject vehicle from perception aspect (speed difference with front vehicle), this study analyzes the instantaneous driving decisions under naturalistic driving environment from decision to decision aspect. In addition, a two-step driving decision procedure is analyzed: 1) micro-level driving decisions, which defined by acceleration and deceleration, and 2) aggregated event-level driving decisions, which captured by subject vehicle making a lane change or continuing to follow front vehicle during a car following event. The sufficient geo-reference trajectory data collected from connected vehicle enables the analysis.

To explore correlations of driving decision, this study also creates new variables which define different driving situations based on relative distance and speed to front vehicles. The modeling results shows that, on average, the subject vehicles are more likely to accelerate as front vehicle to achieve relative high speed. However, they are less likely to accelerate as front vehicle when the driving situation is more complex and congested, compared with related non-congested driving situation.

CHAPTER 5 THE ROLE OF DRIVING VOLATILITY ON THE OCCURRENCE OF A LANE CHANGE CRASH OR NEAR CRASH

This chapter is a revised version to be submitted by Meng Zhang, and Asad Khattak

ABSTRACT

This study investigates relationships between lane change or merge related crashes or near crashes and driving volatility, which quantifies variability in instantaneous driving decisions, by analyzing 1,026 lane change or merging related events along with corresponding naturalistic driving trajectory data (30 seconds duration) collected from the Strategic Highway Research Program-Naturalistic Driving Study. The study measures driving volatility by analyzing fluctuations in longitudinal and lateral accelerations (reported at 10 HZ) archived in the trajectory data. A measure called the coefficient of variation, defined as the ratio of standard deviation to mean, is used to quantify the volatility of driving behavior in this study. The crash outcome contains three categories: baseline, i.e., not a crash (58%), near crash (19%) and crash (23%). To account for the multinomial nature of crash outcomes and capture the unobserved heterogeneity in the data due to unobserved factors, a rigorous multilevel mixed-effect multinomial logit regression model is estimated in this study. The modeling results show that high lateral driving volatility is associated with higher chances of lane change or merge related crashes or near crashes. Furthermore, the chances of a crash or near crash are higher when a driver makes a lane change or merging maneuver under free flow conditions when a leading vehicle is present. These results have the potential to be used in lane change or merge warning systems that help drivers make more informed lane change or merging decisions in a connected vehicle driving environment.

5.1 INTRODUCTION

The lane change or merging event is a common phenomenon in traffic flow and it can endanger the stable traffic flow and result in safety outcome. In 2015, the lane change or merging crashes accounted for 4.6% (451, 000) of all single- and two- vehicle crashes. Of these, 1.6% are fatal while 2.9% are injured [57]. Although such crashes do not account for a sizable portion of all roadway crashes, the decrease in such crashes can still have substantial benefits regarding social cost.

Previous studies have shown evidence that a lane change or merging related crash is correlated with various factors, such as driving and vehicle factors [1-8]. Variability in instantaneous driving decisions could be the leading contributor of unsafe events. Since a lane change or merging related event is an operation that a driver may show high variation in instantaneous driving decisions, i.e., abrupt acceleration or hard braking, it is very important to get an in-depth understanding of effects of these instantaneous decisions on the occurrence of a lane change or merging related crash, which is under-explored in previous studies.

The objective of this study is to explore the correlation between the propensity of a lane change or merging related crash or near crash and driving volatility which quantifies the variability in instantaneous driving decisions, as well as the traffic parameters (e.g., traffic flow density). The critical part is the measurement of variability in the instantaneous driving decision. Liu and Khattak proposed a concept called "driving volatility" to quantify extreme driving behavior by analyzing the distributions of acceleration and speed [10]. With sufficient trajectory data collected from Strategic Highway Research Program (SHRP 2) Naturalistic Driving Study (NDS), this study uses fluctuations in longitudinal and lateral acceleration (reported at 10 HZ) to measure driving volatility, that is a measure called coefficient of variation (COV), defined as the ratio of standard deviation to mean, is used to quantify the volatility of driving behavior [43]. To sum, two key questions to be answered in this study are:

1) How will the driving volatility be measured through the analysis of longitudinal and lateral acceleration?
2) What are the correlates of a lane change or merging related crash propensity with driving volatility?

5.2 LITERATURE REVIEW

Lane change decisions and execution

Many previous studies put efforts on the development of mathematical models to model or simulate lane change behavior, or study the relationship between lane change and traffic flow. For example, macroscopic models are developed to study various traffic flow characteristics in the lane change, including the exchange rate of flows between lanes [58, 59], and frequency of lane change maneuvers [60]. With the development of microscopic traffic simulation tools, lane-changing behavior attracted more attention at the micro level. Lane change is usually classified as either mandatory (MLC) or discretionary (DLC). But they are modeled based on the three steps: 1) necessity checking of a lane change; 2) choosing target lane; and 3) gap acceptance decision. Rule-based models [28] and discrete choice-based (DCB) models [61] were the most two popular models. In addition, some studies focused on the impacts of lane change on traffic state or delay. The adverse impacts of lane change on traffic flow are recognized in previous studies [62, 63]. Wang et al. explored the mechanism underlying the delays by using vehicle trajectory data extracted from the video. Results show imbalance impacts of the lane change; that is vehicles complete their lane change maneuver and return to steady state quicker when following an entering vehicle than when following an existing vehicle [63].

Although many lane change models are developed, the majority of existing models mainly focus on decision making part of a lane change. Another critical process, the lane change execution, which happens after drivers have decided to change lane and find an acceptable gap, is analyzed by few studies. The duration of lane change execution is explored. Toledo and Zohar estimated lane change duration for passenger cars and trucks respectively by applying an algorithm [64]. Moridpour et al. studied driver behavior in lane change execution and proposed a model for lane change execution behavior, but only the longitudinal movement of the vehicle is considered in this study [65]. Since a lane change related event is a relative lateral movement, this study will involve the instantaneous lateral driving decision in the analysis.

Lane change related crashes

Studies also analyzed the lane change related crashes. Chovan et al. found that a lane change related crash occurs commonly when a subject vehicle makes a lane change and hits another vehicle on the adjacent lane driving with similar speed [1, 2]; sideswipe crashes account for the highest percentage in these lanes change related crashes. Some studies compared the propensity of a lane change related crash occurring at the center lane with right or left side lane [2]. The influence of real-time traffic flow and geometric factors were analyzed. They reported that traffic flow related variables are statistically associated with a lane change related crash, while speed or occupancy related variables are not significant. Chen et al. focuses on the effects of the lane-specific real-time traffic factors and found that the propensity of a lane change related crash is associated with average flow into the target lane at the first downstream station and flow ratio at the second downstream [3].

Some studies conducted depth analysis regarding the correlation between lane change, short-term traffic flow, and the lane change related crash. Using loop detector data, Park and Ritchie observed high variation in speed during a lane change and proposed that the propensity of a lane change crash may increase along with the increase of variations in vehicle speed [66]. But the results were not validated by using real crash data.

Previous studies have analyzed lane change decisions and lane change related crashes separately. These studies indicate the occurrence of a lane change related crash is associated with various factors, such as traffic flow parameters. In addition, the important role of driving decision in lane change behavior is recognized in previous studies. A lane change event is a relative micro driver level maneuver, however, to the best of our knowledge, the in-depth understanding of instantaneous lateral driving decision during a lane change maneuver is still under-discussed. In order to fill the gap, this study analyzes the correlates of lane change or merging related crash (or near crash) propensity with driving volatility which quantifies variability in instantaneous driving decisions; which is also under-explored in previous studies. Given the sufficient naturalistic trajectory data and lane change or merging related event summary data maintained by SHRP 2 NDS, the analysis is possible. A unique aspect of this study is the in-depth understanding of variability in instantaneous longitudinal and lateral driving decisions prior to the occurrence of a lane change or merging related crash or near crash by estimating a rigorous statistical modeling using merged data collected from the naturalistic driving environment.

5.3 METHOD

5.3.1 Data source

The data used for analysis is the Second Strategic Highway Research Program (SHRP 2) Naturalistic Driving Study (NDS) data set. Approximately 3,400 participants driver participated and over 4,300 years of naturalistic driving data between 2010 and 2013 collected from six sites around the United States, such as Seattle, Washington; Tampa, Florida; and Buffalo, New York. The data is collected from over 3,300 participant vehicles equipped with a data acquisition system (DAS). The data elements include four video view (driver's face, driver's hand, forward roadway, rear roadway), vehicle network information (e.g., speed, brake, accelerator position), and information from additional sensors (e.g., forward radar, accelerometers). The data used in this study are on-board sensor trajectory data (30 seconds duration) and event summary data set provided by Oak Ridge National Labs (ORNL). A total of 9,593 trips (events) made by 1,580 drivers representing 2,190,316 driving records are provided. Nearly 90 variables (17 in trajectory data set while 76 in event summary data set) are involved in the two data sets and the corresponding key example variables are listed:

1) On board sensor trajectory data: participantID, longitudinal and lateral acceleration (reported at 10 HZ), and vehicle speed (reported at 1 HZ); and
2) Event summary data: participantID, nature of crash outcome (crash, near crash and baseline, e.g., not a crash), pre-incident maneuver (e.g., lane change), location (e.g., intersection), situational factors (e.g., free flow) and roadway geometric (e.g., grade down). More detail information is available in the description of SHRP 2 NDS data sets [67].

Since this study focuses on these lateral movement related events, this study extracts these events based on the rule such as the pre-incident maneuver is reported as changing lane or merging. After data cleaning and error check, a total of 1,026 lane change or merging related events representing 255,720 driving records are selected for analysis. The data is error-checked and validated using descriptive statistics.

Figure 5.1 shows the final data structure and conceptual framework. These trajectories driving records are aggregated to the trip level and then are linked to the event summary file based on the same variable ("participantID") within two data sets. The trajectory data is used to calculate the driving volatility of each trip based on fluctuations in longitudinal and lateral acceleration. More detailed calculation rule is shown in the following context. Note that the nature of crash outcome contains three categories as reported by the description of SHRP 2 NDS data sets [67]:

- Baseline event: refers to the "normal" driving event which is not a crash event. These baseline events are randomly selected through a sample stratified by participant and the proportion of time driven. Note the driving time only includes driving speeds above 5 mph in order to avoid the time influence of long stopping and to concentrate on the risk periods [67];
- Near crash event: refers to a non-crash event but a rapid evasive maneuver is needed by the subject vehicle, or another vehicle, pedestrian, cyclist, or animal, to avoid a crash. The definition of a rapid evasive maneuver is base on vehicle control inputs, such as the steering, braking, or acceleration;
- Crash event: refers to the contact between subject vehicle with a moving or fixed object at any speed which results in the measurable transfer or dissipation in kinetic energy.

These crashes also include situations that the subject vehicle strikes another vehicle, pedestrian or cyclist, animal, roadside barrier or object on or off the roadway, as well as non-premeditated departures of the roadway where at least one tire leaves the paved.

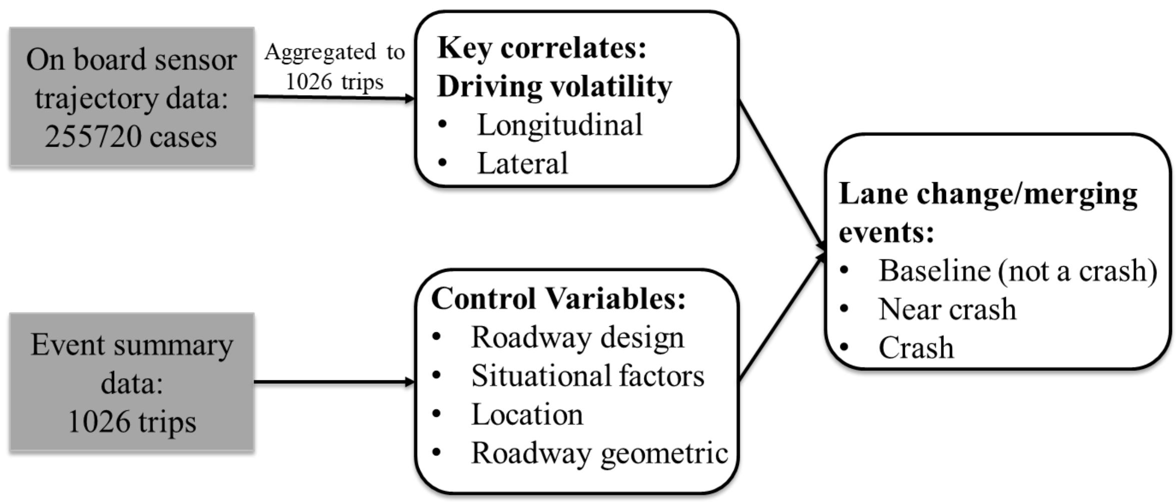


Figure 5. 1 Data structure and conceptual framework

5.3.2 Driving volatility

The understanding of variability in the instantaneous driving decision during a lane change or merging related event is a critical part of this study. Previous studies have proposed methods, such as giving a fixed cut-off value of acceleration as the threshold, to differentiate aggressive driving and calm driving [34-37]. In fact, the acceleration ability is associated with driving speed. The higher in speed, the lower in acceleration ability due to aerodynamic resistance. Noticing the variation in acceleration across different speed ranges, Liu and Khattak analyzed the relationship between speed and acceleration and proposed a speed-acceleration based method to measure driving volatility [10]. However, due to the insufficient driving records of vehicle speed (reported at 1 HZ) in SHRP 2 NDS trajectory data for each lane change or merging event, this study uses fluctuations in longitudinal and lateral acceleration to measure the driving volatility.

Thus, a measure called coefficient of variation (COV), also defined as the ratio of standard deviation to mean, is used to quantify the variability in instantaneous driving decisions [43]. COV is a standardized measure of relative dispersion in statistics [68]. Since the different patterns in longitudinal and lateral acceleration or deceleration, four types of driving volatility are measured in this study. The formulas for COV calculation are shown below:

$$\text{Longitudinal} - \text{acceleration: } COV_{acc_x} = \frac{Std.Dev._{acc_x}}{Mean_{acc_x}} \tag{1}$$

$$\text{Longitudinal} - \text{deceleration: } COV_{acc_x} = \frac{Std.Dev._{dec_x}}{Abs(Mean_{dec_x})} \tag{2}$$

$$\text{Lateral} - \text{acceleration (right side): } COV_{R\_acc_y} = \frac{Std.Dev._{R\_acc_y}}{Mean_{R\_acc_y}} \tag{3}$$

$$\text{Lateral} - \text{acceleration (left side): } COV_{L\_acc_y} = \frac{Std.Dev._{L\_acc_y}}{Abs(Mean_{L\_acc}\ )} \tag{4}$$

5.3.3 Model structure

After measuring driving volatility, this study estimates rigorous statistical model to investigate the correlates of crash propensity with related factors, especially the driving volatility. Three multinomial scales: 1- baseline (not a crash); 2 - near crash; and 3 - crash, are used in the crash outcome as the response variable. Considering the hierarchical data structure of lane change or merging related events (shown in Figure 5.2) that these events are nested in the drivers and accounting for unobserved heterogeneity due to unobserved factors, a multilevel mixed-effect multinomial logit model is estimated. The multilevel multinomial logit model is a mixed Generalized Linear Model with linear predictors $\eta_{ij}^{(m)}$ [69]:

$$\eta_{ij}^{(m)} = \alpha^{(m)} + \boldsymbol{\beta}^{(m)\prime}\boldsymbol{x}_{ij} + \xi_j^{(m)} + \delta_{ij}^{(m)} \qquad \text{Equation (1)}$$

And multinomial logit link:

$$P\left(Y_{ij} = m \middle| x_{ij}, \boldsymbol{\xi}_j, \boldsymbol{\delta}_{ij}\right) = \frac{\exp\left\{\eta_{ij}^{(m)}\right\}}{1+\sum_{l=2}^{M}\exp\left\{\eta_{ij}^{(l)}\right\}} \qquad \text{Equation (2)}$$

Where,

m = 1, 2, . . . ,M denotes the response category (crash outcome);

Y_{ij} = the crash outcome of *j^th^* event generated by *i^th^*, taking value from {1, , . . . ,M};

$\eta_{ij}^{(m)}$ = linear predictor;

$\boldsymbol{x}_{ij}$ = a set of explanatory variables, such as driving volatility;

$\boldsymbol{\beta}^{(m)}$ = a coefficient set of explanatory variables, m = 2, 3, . . . ,M;

$\alpha^{(m)}$ = Constant term, m = 2, 3, . . . ,M;

j = 1, 2, . . . , J denotes the cluster (driver);

i = 1, 2, . . . , n_j denotes the subject (lane change or merging event) of j^{th} cluster.

$\boldsymbol{\xi}_j$ and $\boldsymbol{\delta}_{ij}$ are sets of random errors capturing the unobserved heterogeneity at cluster (driver) and subject (lane change or merging event) level, respectively; $\xi_j' = \left(\xi_j^{(2)}, \dots, \xi_j^{(M)}\right)' \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\xi})$;

$\delta_{ij}^{(m)} = (\delta_{ij}^{(2)}, \dots, \delta_{ij}^{(M)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\delta})$;

The likelihood of model (1)-(2) are calculated by utilizing the conditional independence from the assumptions:

$$L(\theta) = \prod_{j=1}^{J} \int \prod_{i=1}^{n_j} \{ \int P(Y_{ij} = m | x_{ij}, \xi_j, \delta_{ij}) f(\delta_{ij}) d\delta_{ij} \} f(\xi_j) d\xi_j \qquad \text{Equation (3)}$$

Where $\theta' = (\alpha^{(2)}, \ldots, \alpha^{(m)}, \boldsymbol{\beta}^{(2)}, \ldots, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Sigma}_{\xi}, \boldsymbol{\Sigma}_{\delta})$. The coefficients are estimated using maximum likelihood method. A likelihood ratio test is applied to compare the multilevel mixed-effect multinomial logit model with traditional multinomial logit model.

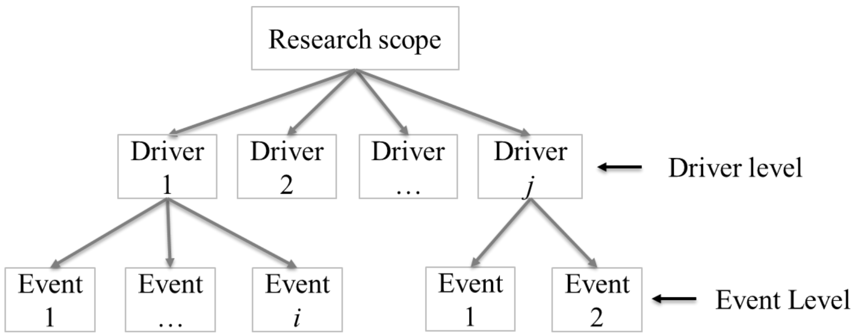


Figure 5. 2 Data structure

5.4 RESULTS

5.4.1 Driving speed, longitudinal and lateral acceleration

Figure 5.3 presents the distribution between driving speed, longitudinal acceleration, and lateral acceleration using limited available records, reported at 1 HZ. The results are consistent with the previous study [10]. The longitudinal and lateral acceleration ability decrease along with the increase of vehicle speed (shown in Figure 5.3 a and b). There is a relative rhombus relationship between longitudinal and lateral acceleration, as shown in Figure 5.3 (c). Figure 5.3 (a) shows interesting results regarding magnitudes in longitudinal acceleration and deceleration. There are many variations in longitudinal deceleration, while the longitudinal acceleration is much more stable. Generally, when a subject vehicle is approaching the front vehicle whose speed is lower, the subject driver might need to make a hard braking in order to avoid the collision with the front vehicle. Note that there is no lateral deceleration. Therefore, the positive and negative value in lateral acceleration represent acceleration to the right and left side, respectively.

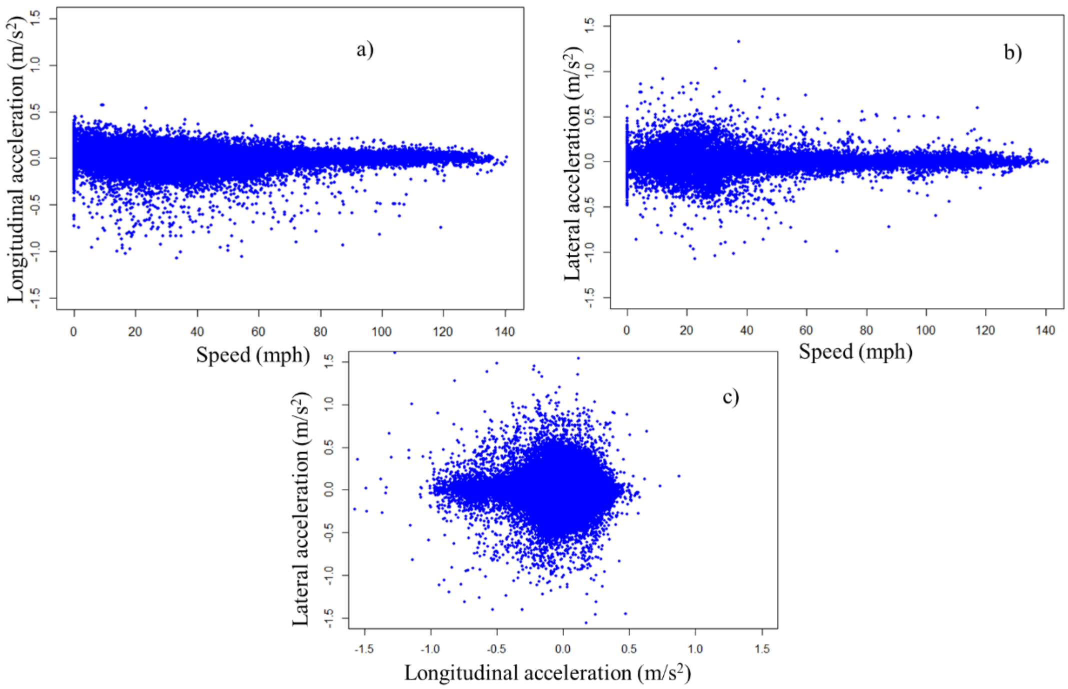


Figure 5. 3 Relationship between speed, lateral acceleration and longitudinal acceleration

5.4.2 Descriptive statistics

Table 5.1 shows descriptive statistics of key variables in lane change or merging related baseline event (not a crash), near crashes and crashes. A total of 1,026 lane change or merging related events are selected for analysis. Of these, 22.7% events result in the crashes, 19.3% are near crashes and 58% are baseline events. On average, there is no much difference in volatility between longitudinal deceleration and lateral acceleration, with a value close to 1. The volatility of longitudinal acceleration is lower with a value 0.83.

Nearly 46.7% of drivers are making the lane change or merging under free flow without leading traffic condition, only 5.7% of them will make a lane change or merging under stable flow with maneuverability or speed restriction; where they might not make a lane change or merging maneuver as they are besieged by surrounding vehicles. Subject drivers are more likely to make a lane change or merging when the grade is level (89.3%).

Table 5. 1 Descriptive statistics of lane change or merging events using NDS data (N=1026)

Variables		N	Mean	Std. Dev.	Min	Max
Crash propensity	Not a crash (baseline)	1026	0.580	0.494	0	1
	Near crash	1026	0.193	0.395	0	1
	Crash	1026	0.227	0.419	0	1
Driving volatility	COV_{acc_x}	1026	0.830	0.319	0.293	3.464
	COV_{dec_x}	1026	0.940	0.402	0.075	3.351
	$COV_{R\_acc_y}$	1026	1.081	0.416	0.197	4.314
	$COV_{L\_acc_y}$	1026	0.945	0.417	0	4.268
Roadway design	Divided (median strip or barrier)	1026	0.358	0.480	0	1
	No lanes	1026	0.111	0.314	0	1
	Not divided (center 2-way left turn lane)	1026	0.060	0.238	0	1
	Not divided (simple 2-way traffic way)	1026	0.426	0.495	0	1
	One-way traffic	1026	0.045	0.207	0	1
Traffic density	Free flow, no lead traffic	1026	0.467	0.499	0	1
	Free flow, leading traffic	1026	0.263	0.441	0	1
	Flow with some restrictions	1026	0.177	0.382	0	1
	Stable flow, maneuverability or speed restricted	1026	0.057	0.231	0	1
	Others, e.g., unstable flow	1026	0.036	0.187	0	1
Location	Intersection or junction	1026	0.557	0.497	0	1
Alignment	Straight roadway	1026	0.861	0.347	0	1
Grade	Level	1026	0.893	0.310	0	1
	Dip or grade down	1026	0.036	0.187	0	1
	Grade up or hillcrest	1026	0.071	0.257	0	1

5.4.3 Modeling results

Crash propensity

Table 5.2 shows the modeling results of the multilevel mixed-effect multinomial logit model, including fixed effects and random effects, for the crash or near crash propensity of the driver involved in a lane change or merging event. The modeling results quantifying the effects of driving volatility as well as traffic flow parameters on driver crash propensity. The reported likelihood ratio test of multilevel model vs. regular model indicates significant variability between drivers to favor a multilevel mixed-effect multinomial logit model at 95% confidence level. As expected, most explanatory variables have shown significant correlations with crash propensity at 95% confidence level and the signs of coefficients are expected. Although the explanatory variables are significant at the event level, the correlates may vary across different drivers.

5.4.4 Discussion of key variables

Driving volatility

Compared with base level (baseline event, such as the normal lane change or merging event), volatile driving behavior captured by high driving volatility (e.g., hard braking) is associated with higher chances of a crash or near crash. More attention should be paid to the volatility of longitudinal deceleration, as it has shown much high magnitude with a positive sign in coefficient. High volatility might due to the high-speed subject vehicle is approaching the low-speed front vehicle in a relatively short distance, thus the subject vehicle has to make hard braking in the longitudinal direction, or make abrupt lateral acceleration to avoid the collision with the front vehicle or to achieve the satisfied speed through changing lanes. Further study is needed when detailed data regarding the driving environment is available. If proper warning (e.g., relative distance and speed warning) could be provided to subject or front drivers to help them to adjust their driving behaviors under a connected vehicle driving environment, a crash or near crash might be avoided; this can be beneficial for connected vehicles at Level 1 or 2 automation, as the driver assistance system could help the execution of acceleration or deceleration using information about the driving environment, such as relative distance and speed to front vehicle in this case.

The subject vehicle with high volatile behavior in lateral acceleration to the left side is more likely to be involved in a crash or near crash, compared to right side. Generally, the speed on the left side lane is higher than the speed on the right side lane, the subject vehicle might need to make a more abrupt acceleration in short time to make a successful lane change or merging. The coefficient of the driving volatility of longitudinal acceleration shows abnormal signs. High volatility in longitudinal acceleration is marginally significantly associated with the lower chance of a near crash, while it is not significantly correlated with a crash event. The odds of a near crash for the driving volatility of lateral acceleration to the left side are -73% ([exp(β)-1]*100%), compared with the base condition (normal lane change or merging).

Driving situational factors

The effects of driving situational factors are also explored. Compared with base condition (free flow with no leading vehicle), although subject vehicle makes less lane change or merging maneuver under free flow with no leading vehicle or under stable flow with speed restriction, the chance of a crash or near crash is higher. The results are consistent with the expected line. Speed restriction indicates the subject vehicles are besieged by surrounding vehicles with relatively short distance, thus they are more likely to be involved in a crash or near crash when making a lane change or merging maneuver due to high chance of exposure to other vehicles. Note that stable flow with maneuverability or speed restricted is associated with higher chance of a near crash than a crash. Restricted speed indicating low speed, the subject vehicle can make a full stop easily when making a lane change or merging under that situation, as a result, a crash can be avoided.

Roadway geometric and design factors

Some crashes or near crashes can be caused by roadway geometric and design. The chance of a crash or near crash is higher when the grade is down compared to when the grade is upgrade. The subject vehicle will obtain a large additional acceleration, as a result, the speed of the subject vehicle increases and it is hard to make an instant full stop when making a lane change or merging. Therefore, the chance of a crash or near crash is higher.

The subject vehicle driving in the divided roadway or in not divided way (center 2-way left turn lane) are less likely to be involved in a crash or near crash, compared with driving in the roadway without lanes. The traffic condition might be more complex in no lane roadway, such as vehicles might not follow the roadway rules, therefore, the chance of a crash or near crash is high when driving on roadway without lanes. Unexpected, the straight roadway is associated with the high chance of a crash or near crash, compared with base (e.g., curve).

Location factors

Besides above mentioned explanatory variables, this study also untangles the effects of location attributes. Intersection or junction are associated with the high chances of a crash or near crash.

Table 5. 2 Multilevel mixed-effect multinomial logit modeling results for lane change or merging related crash propensity (N=1,026)

Variables (base: Not a crash)		Near crash			Crash		
		β	P-value		β	P-value	
Driving volatility	COV_{acc_x}	-1.322	0.062	*	0.871	0.154	
	COV_{dec_x}	10.742	0.000	***	9.338	0.000	***
	$COV_{R\_acc_y}$	2.323	0.008	***	3.824	0.000	***
	$COV_{L\_acc_y}$	7.154	0.000	***	7.811	0.000	***
Roadway design Base: No lanes	Divided (median strip or barrier)	-1.956	0.024	**	-2.822	0.001	***
	Not divided (center 2-way left turn lane)	-2.413	0.053	*	-3.652	0.003	***
	Not divided (simple 2-way traffic way)	-1.171	0.108		-1.729	0.010	***
	One-way traffic	-0.949	0.417		-1.475	0.192	
Traffic density Base: Free flow, no lead traffic	Free flow, leading traffic	1.414	0.013	**	0.569	0.299	
	Flow with some restrictions	2.845	0.000	***	1.648	0.035	**
	Stable flow, maneuverability or speed restricted	4.784	0.000	***	2.523	0.036	**
	Others, e.g., unstable flow	4.582	0.001	***	1.896	0.207	
Location	Intersection or junction	2.473	0.000	***	2.647	0.000	***
Alignment	Straight roadway	1.496	0.052	*	1.843	0.014	**
Grade	Level	-4.342	0.002	***	-4.756	0.000	***
Base: Dip or grade down	Grade up or hillcrest	-4.828	0.003	***	-4.837	0.002	***
	Constant	- 17.938	0.000	***	- 19.233	0.000	***
Random effect parameter (Driver)	Variance	10.558					
	Residual	4.776					
Summary statistics	Sample size	1026					
	Likelihood at 0	-995.329					
	Likelihood at β	-488.955					
	Prob. > χ^2	0.000***					
	Likelihood ratio test: Multilevel vs. mlogit	0.000***					

Notes: STATA software (gesm program) was used;
*** - means statistical significant associations were found (at 1% level); ** - means statistical significant associations were found (at 5% level); * - means statistical significant associations were found (at 10% level).

Given the complexity of the driving environment and high exposure in the intersection, the subject vehicle might be more likely to have a collision with another vehicle when making a lane change or merging.

5.5 LIMITATIONS

This study has explored various factors, such as driving volatility, situational factors and roadway geometric, that can lead to the occurrence of a lane change or merging related crash. However, some other factors, especially distance to surrounding vehicles and number of surrounding vehicles, might be highly correlated with lane change crash propensity are not analyzed due to the limited data. Therefore, the explanatory power of the modeling part will be restricted to these selected independent variables.

Currently, the driving volatility is quantified only based on acceleration, while the vehicle speed is not involved given low report frequency (reported at 1 HZ). In fact, the acceleration ability will vary along with different speed range [10]. Speed-based driving volatility should be considered when data is available. Although, the GPS data is guaranteed given the advanced data collection techniques, there still exist measurement errors. Since the distributions of key variables, such as longitudinal and lateral acceleration, are in the reasonable ranges based on results of descriptive statistics, the influence of measurement errors could be eliminated.

Another issue will be the accuracy in some critical variables, such as nature of the crash outcome. For example, The researcher reports a near-crash based on a rapid evasive maneuver by subject vehicle. However, this identification is subjective as they highly rely on the judgment of the researchers.

5.6 CONCLUSIONS

Previous studies have investigated the causes of dangerous lane change or merging events because they are a key threat to smooth traffic flow and safety. However, the correlation between the occurrence of a lane change or merging related crash or near crash and driving volatility, which quantifies the variability in instantaneous driving decisions, is under-explored. With sufficient trajectory data provided by SHRP 2 NDS, this study investigates the relationship between the occurrence of a lane change or merging related crash or near crash with driving volatility as well as traffic flow parameters. This study is timely and unique as it links variability in the instantaneous driving decisions with crash outcomes in a naturalistic driving environment. This study's further contributions include using a unique and rich database and rigorous statistical model to quantify the correlations between lateral and longitudinal driving volatility with the risk of lane change or merge related crashes, which should be useful to researchers and practitioners.

Using a unique data set from naturalistic driving trajectory data and event summary data, maintained by SHRP 2 NDS, this study quantifies the variability in instantaneous driving decisions for 1,026 naturalistic trajectories. The study uses the Coefficient of Variation (COV) to measure driving volatility. By considering the hierarchical data structure and accounting for unobserved heterogeneity due to unobserved factors, a multilevel mixed-effect multinomial logit model is estimated in order to explore the correlations between lane change or merge related crash propensity with driving volatility as well as traffic parameters. What follows is a summarization of key findings.

- Volatile driving behavior (captured by high lateral driving volatility) is more likely to result in the occurrence of a lane change or merging related crash or near crash.

- The chances of a crash or near crash are higher when a driver who makes a lane change or merge related event under free flow with a leading vehicle, under stable flow with speed restrictions, when the grade is lower, at intersections, or on a relatively straight roadway.
- A subject vehicle driving on a divided roadway or a roadway with a center 2-way left turn lane is less likely to be involved in a crash or near crash than a vehicle driving on a roadway without lanes.

The results have potential applications for the improvement of lane change or merging safety. The study provides insights on lateral driving volatility. Analysis found that high magnitude with a positive sign is in the coefficient of lateral driving volatility, indicating that reducing the variability in instantaneous driving decisions by the subject vehicle can improve safety. The results could be helpful for developing connected vehicles at Level 1 or 2 automation because critical information, such as relative distance from and speed of the front vehicle, can be detected and transferred by driver assistance systems which in turn helps subject vehicles make informed driving decision, such as safer merging maneuver at merging ramps [70]. In addition, alerts and warnings can be issued to surrounding vehicles (in the front or to the side) to adjust their driving behavior in order to avoid a collision with the subject vehicle in a connected vehicle environment. Note that some volatile lane change maneuvers happen because of the surrounding driving environment or geometric design, such as a short ramp. The subject vehicle has to accelerate harder to make a successful merging on a shorter ramp. We should pay more attention to ramps in which many subject drivers make volatile merging maneuvers. The roadway manager might need to redesign the ramp in order to ensure less volatile merging maneuvers. Of course, researchers should further analyze the relationship between driving volatility with the surrounding driving environment and geometric factors.

CHAPTER 6 CONCLUSIONS

This dissertation aims to explore automobile driver behaviors at a micro-level with concern to the instantaneity of lateral driving decisions by integrating and mining massive vehicle trajectory data. With advanced technology, massive vehicle driving data is available to public. Critical information embedded in the "Big Data" can be extracted and analyzed to improve transportation performance such as safety and mobility. The dissertation is timely given the high attention given to GPS data in recent years and it is necessary for the development of new methodology for extracting key information from "Big Data".

The geo-referenced vehicle trajectory data, reported at a 10 Hz frequency, describes a vehicle's position, motion and surrounding driving situations at the very detail micro-level, which makes it is possible to analyze the micro-level driving behavior, especially aggressive or extreme driving behaviors (e.g., hard accelerations or fast lane changes), from the massive GPS data. Since the lane change is fundamental to microscopic traffic flow and safety, a study was conducted to understand normal and extreme lane change behaviors, which can form the basis for generating alerts and warnings that can reduce the impacts of such behaviors. Using the high-resolution driving data, the study proposed an innovative methodology to identify normal and extreme lane change maneuvers. The lane changes are identified based on multiple criteria, including vehicle position (i.e., a sharp change in distance of a vehicle's centerline relative to lane boundary) and lane crossings recorded by onboard units (i.e., when a vehicle crosses a lane marker). Extreme lane change events are then identified as those where lateral acceleration exceeds the 95th percentile threshold at the initiation and before the end of the lane change maneuver. The results show that the test vehicles averagely generated 3.4 lane changes (0.67 extreme lane changes) with trip duration averaging 20 minutes. Based on the analysis of this data, warnings can be generated to help surrounding drivers adjust their behaviors to accommodate extreme behavior by the host vehicle driver.

Given the large portion of lane departure crash, the onboard lane keeping warning system is developed to prevent these crashes. Therefore, a study of understanding instantaneous lane keeping behaviors was conducted. A measure called lateral shifting volatility, which quantifies fluctuation in lateral displacement, is developed in the study. The study also explores the influence of driving situation on shifting volatility. The results show that the subject vehicle is more volatile when traveling at high speeds and when the vehicle keeps a low space gap with the vehicle in front of it. The shifting volatility information can be applied in onboard driving systems to help drivers make informed lane departure decisions.

While driving behavior is influenced by surrounding vehicles, a study explores the peer influence of front vehicle on instantaneous driving decision of subject vehicle is conducted. A “Gossip” concept is proposed to capture the peer influence and a two-step driving decision procedure are analyzed: 1) micro-level driving decision defined by vehicle acceleration and deceleration, and 2) aggregated event-level driving decision captured by subject vehicle making a lane change or not during a car following event. This study further explores the correlations of driving decision with various driving situations. The results show that the subject vehicle is averagely more likely to accelerate as front vehicle to achieve high speed, however, they are less likely to accelerate as front vehicle when the driving situation is more complex and congested. This study establishes a new framework to understand the driving decisions during car following events.

Since the variability in instantaneous driving decisions could be the leading contributor of unsafe events, a study was further conducted to explore the correlations between the occurrence of a lane change or merging related crash with the instantaneous driving decisions, which is under-explored in previous studies. The results show that high lateral driving volatility is associated with a higher chance of the lane change or merging related crashes or near crashes. Furthermore, the chances of a crash or near crash are higher when a driver makes a lane change or merging maneuver under free flow conditions with a leading vehicle present, compared with no leading vehicle.

LIST OF REFERENCES

1. Chovan, J.D., *Examination of lane change crashes and potential IVHS countermeasures*. 1994: National Highway Traffic Safety Administration.
2. Lee, C., et al., *Effects of Lane-Change and Car-Following-Related Traffic Flow Parameters on Crash Occurrence by Lane*. 2009.
3. Chen, Z., X. Qin, and M.R.R. Shaon, *Modeling Lane-change Related Crashes with Lane-specific Real-time Traffic and Weather Data.* Journal of Intelligent Transportation Systems, 2017(just-accepted).
4. Fitch, G., et al., *Analysis of lane-change crashes and near-crashes.* US Department of Transportation, National Highway Traffic Safety Administration, 2009.
5. Sen, B., J.D. Smith, and W.G. Najm, *Analysis of lane change crashes*. 2003.
6. Lee, C., M. Abdel-Aty, and L. Hsia, *Potential real-time indicators of sideswipe crashes on freeways.* Transportation Research Record: Journal of the Transportation Research Board, 2006(1953): p. 41-49.
7. Wang, J.-S. and R.R. Knipling, *Lane change/merge crashes: problem size assessment and statistical description*. 1994: US Department of Transportation, National Highway Traffic Safety Administration.
8. Pande, A. and M. Abdel-Aty, *Assessment of freeway traffic parameters leading to lane-change related collisions.* Accident Analysis & Prevention, 2006. **38**(5): p. 936-948.
9. WHO, *Road traffic injuries.* 2016.
10. Liu, J. and A.J. Khattak, *Delivering improved alerts, warnings, and control assistance using basic safety messages transmitted between connected vehicles.* Transportation Research Part C: Emerging Technologies, 2016. **68**: p. 83-100.

11. Luo, Y., et al., *A dynamic automated lane change maneuver based on vehicle-to-vehicle communication.* Transportation Research Part C: Emerging Technologies, 2016. **62**: p. 87-102.

12. Gong, S. and L. Du, *Optimal location of advance warning for mandatory lane change near a two-lane highway off-ramp.* Transportation Research Part B: Methodological, 2016. **84**: p. 1-30.

13. Xi, Y. and M. Crisler. *A Review of Lane Change Definitions and Identification Methods*. in *Transportation Research Board 92nd Annual Meeting*. 2013.

14. Bogard, S. and P. Fancher, *Analysis of data on speed-change and lane-change behavior in manual and ACC driving*. 1999.

15. Worrall, R. and A. Bullen, *An empirical analysis of lane changing on multilane highways.* Highway Research Record, 1970(303).

16. Miller, R.J. and G. Srinivasan. *Determination of Lane Change Maneuvers Using Naturalistic Driving Data*. in *Proceedings: International Technical Conference on the Enhanced Safety of Vehicles*. 2005. National Highway Traffic Safety Administration.

17. Ayres, G., B. Wilson, and J. Le-Blanc, *Method for identifying vehicle movements for analysis of field operational test data.* Transportation Research Record: Journal of the Transportation Research Board, 2004(1886): p. 92-100.

18. Thiemann, C., M. Treiber, and A. Kesting, *Estimating acceleration and lane-changing dynamics from next generation simulation trajectory data.* Transportation Research Record: Journal of the Transportation Research Board, 2008(2088): p. 90-101.

19. Knoop, V., et al., *Quantifying the number of lane changes in traffic: Empirical analysis.* Transportation Research Record: Journal of the Transportation Research Board, 2012(2278): p. 31-41.

20. Xuan, Y. and B. Coifman. *Lane change maneuver detection from probe vehicle DGPS data*. in *2006 IEEE Intelligent Transportation Systems Conference*. 2006. IEEE.

21. Lee, S.E., E.C. Olsen, and W.W. Wierwille, *A comprehensive examination of naturalistic lane-changes*. 2004.

22. Xuan, Y. and B. Coifman, *Identifying Lane-Change Maneuvers with Probe Vehicle Data and an Observed Asymmetry in Driver Accommodation.* Journal of Transportation Engineering, 2012. **138**(8): p. 1051-1061.

23. Van Winsum, W., D. de Waard, and K.A. Brookhuis, *Lane change manoeuvres and safety margins.* Transportation Research Part F: Traffic Psychology and Behaviour, 1999. **2**(3): p. 139-149.

24. Wang, C. and B. Coifman. *Estimating the Number of Lane Change Maneuvers on Congested Freeway Segments*. in *11th World Conference on Transport Research*. 2007.

25. Chen, R., K.D. Kusano, and H.C. Gabler. *Driver Behavior During Lane Change from the 100-Car Naturalistic Driving Study*. in *24th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. 2015.

26. Salvucci, D.D. and A. Liu, *The time course of a lane change: Driver control and eye-movement behavior.* Transportation Research Part F: Traffic Psychology and Behaviour, 2002. **5**(2): p. 123-132.

27. Dingus, T.A., et al., *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment*. 2006.

28. Hidas, P., *Modelling vehicle interactions in microscopic simulation of merging and weaving.* Transportation Research Part C: Emerging Technologies, 2005. **13**(1): p. 37-62.

29. Henclewood, D., *Safety Pilot Model Deployment – One Day Sample Data Environment Data Handbook*. 2014, Research and Technology Innovation Administration, US Department of Transportation: McLean, VA.

30. Liu, J., X. Wang, and A. Khattak, *Generating Real-Time Driving Volatility Information*, in *2014 World Congress on Intelligent Transport Systems*. 2014: Detroit, MI.

31. Staplin, L., et al., *Highway design handbook for older drivers and pedestrians*. 2001.

32. Gattis, J. and S. Low, *Intersection angle geometry and the driver's field of view.* Transportation Research Record: Journal of the Transportation Research Board, 1998(1612): p. 10-16.

33. Wang, X., et al., *What is the Level of Volatility in Instantaneous Driving Decisions?* Transportation Research Part C: Emerging Technologies, 2015.

34. Lajunen, T., J. Karola, and H. Summala, *Speed and acceleration as measures of driving style in young male drivers.* Perceptual and motor skills, 1997. **85**(1): p. 3-16.

35. Langari, R. and J.-S. Won, *Intelligent energy management agent for a parallel hybrid vehicle-part I: system architecture and design of the driving situation identification process.* Vehicular Technology, IEEE Transactions on, 2005. **54**(3): p. 925-934.

36. Kim, E. and E. Choi. *Estimates of Critical Values of Aggressive Acceleration from a Viewpoint of Fuel Consumption and Emissions*. in *2013 Transportation Research Board Annual Meeting*. 2013. Washington DC.

37. De Vlieger, I., D. De Keukeleere, and J. Kretzschmar, *Environmental effects of driving behaviour and congestion related to passenger cars.* Atmospheric Environment, 2000. **34**(27): p. 4649-4655.

38. Liu, J.A.K., *Delivering Improved Alerts, Warnings, and Control Assistance Using Basic Safety Messages Transmitted between Connected Vehicles.* Accepted for presentation at the Transportation Research Board Annual Meeting, National Academies, Washington, D.C., 2016., 2015.

39. Agresti, A., *An introduction to categorical data analysis*. Vol. 135. 1996: Wiley New York.

40. StataCorp, *Stepwise Estimation*. 2013.

41. *Roadway Departure Safety*. 2017.

42. *Status Report, Vol. 52, No. 6: Stay Within the Lines - Lane Departure Warning, Blind Spot Detection Help Driver Avoid Trouble*. 2017.

43. Kamrani, M., A.J. Khattak, and B. Wali, *Can Data Generated by Connected Vehicles Enhance Safety? Proactive Approach to Intersection Safety Management.* Transportation Research Record: Journal of the Transportation Research Board, 2017.

44. McCall, J.C., *Human Attention and Intent Analysis Using Robust Visual Cues in a Bayesian Framework*. 2006, University of California: San Diego.

45. Edwards, C., J. Cooper, and A. Ton, *Human Factors of Vehicle-Based Lane Departure Warning Systems*. 2015, Local Road Research Board Minnesota Department of Transporatation: St. Paul.

46. Lee, S., E. Olsen, and W. Wierwille, *A Comprehensive Examination of Naturalistic Lane-Changes*. 2004, Virginia Tech Transportation Institute: Blacksburg, VA.

47. Lord, D., et al., *Analysis of Roadway Departure Crashes on Two-Lane Rural Roads in Texas*. 2011, Texas Transportation Institute: Austin.

48. Nodine, E.E., A.H. Lam, and W.G. Najm, *Safety Impact of an Integrated Crash Warning System Based on Field Test Data*. 2011, National Highway Traffic Safety Administration: United States.

49. Sayer, J., et al., *Integrated Vehicle-Based Safety Systems Light-Vehicle Field Operational Test Key Findings Report*. 2011, University of Michigan Transportation Research Institute: Ann Arbor.

50. Trivedi, M.M., T. Gandhi, and J. McCall, *Looking-In and Looking-Out of a Vehicle: Computer-Vision-Based Enhanced Vehicle Safety.* IEEE Transactions on Intelligent Transportation Systems, 2007. **8**(1): p. 108-120.

51. Hallmark, S., et al., *Evaluation of Driving Behavior on Rural 2-lane Curves Using the SHRP 2 Naturalistic Driving Study Data.* Journal of Safety Research, 2015. **54**(1): p. 17-27.

52. Navarro, J., et al., *Influence of lane departure warnings onset and reliability on car drivers' behaviors.* Applied Ergonomics, 2017. **59**(1): p. 123-131.

53. Moller, H., et al., *Simulator Performance, Microsleep Episodes, and Subjective Sleepiness: Normative Data Using Convergent Methodologies to Assess Driver Drowsiness.* Journal of Psychosomatic Research, 2006. **61**(1): p. 335-342.

54. Zhang, M. and A.J. Khattak, *Identifying and Analyzing Extreme Lane Change Events Using Basic Safety Messages in a Connected Vehicle Environment*. 2018.

55. Naseri, H., A. Nahvi, and F.S.N. Karan, *A new psychological methodology for modeling real-time car following maneuvers.* Travel behaviour and society, 2015. **2**(2): p. 124-130.

56. Wang, X., et al., *What is the level of volatility in instantaneous driving decisions?* Transportation Research Part C: Emerging Technologies, 2015. **58**: p. 413-427.

57. Administration, N.H.T.S., *Traffic Safety Facts 2015.* Available from : https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812384, 2016.

58. Daganzo, C.F., *A behavioral theory of multi-lane traffic flow. Part I: Long homogeneous freeway sections.* Transportation Research Part B: Methodological, 2002. **36**(2): p. 131-158.

59. Coifman, B., *Estimating density and lane inflow on a freeway segment.* Transportation Research Part A: Policy and Practice, 2003. **37**(8): p. 689-701.

60. Kang, K.-P. and G.-L. Chang. *Observations of macroscopic non-mandatory lane-changing behaviors on the Capital Beltway*. in *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*. 2004. IEEE.

61. Toledo, T., H. Koutsopoulos, and M. Ben-Akiva, *Modeling integrated lane-changing behavior.* Transportation Research Record: Journal of the Transportation Research Board, 2003(1857): p. 30-38.

62. Zheng, Z., S. Ahn, and C.M. Monsere, *Impact of traffic oscillations on freeway crash occurrences.* Accident Analysis & Prevention, 2010. **42**(2): p. 626-636.

63. Wang, C. and B. Coifman, *The effect of lane-change maneuvers on a simplified car-following theory.* Intelligent Transportation Systems, IEEE Transactions on, 2008. **9**(3): p. 523-535.

64. Toledo, T. and D. Zohar, *Modeling duration of lane changes.* Transportation Research Record: Journal of the Transportation Research Board, 2015.

65. Moridpour, S., M. Sarvi, and G. Rose, *Modeling the lane-changing execution of multiclass vehicles under heavy traffic conditions.* Transportation Research Record: Journal of the Transportation Research Board, 2010(2161): p. 11-19.

66. Park, S. and S.G. Ritchie. *Exploring the relationship between freeway speed variance, lane changing, and vehicle heterogeneity*. in *83rd Annual Meeting of Transportation Research Board*. 2004.

67. Hankey, J.M., M.A. Perez, and J.A. McClafferty, *Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets*. 2016, Virginia Tech Transportation Institute.

68. Washington, S.P., M.G. Karlaftis, and F. Mannering, *Statistical and econometric methods for transportation data analysis*. 2010: CRC press.

69. Grilli, L. and C. Rampichini, *A multilevel multinomial logit model for the analysis of graduates' skills.* Statistical Methods & Applications, 2007. **16**(3): p. 381-393.

70. Md Salman Ahmed, M.A.H., Jackelin Rios-Torres, Asad and Khattak, *A Cooperative Freeway Merge Assistance System using Connected Vehicles.* Working paper, 2017.

VITA

Ms. Meng Zhang was born on November 5, 1989, and grew up in a small city called Ruichang in Jiangxi Province of China. Ms. Zhang started her transportation study at Dalin University of Technology, where she received her Master Degree in Transportation Planning and Management in 2014. In the same year, Ms. Zhang continued her study at The University of Tennessee, Knoxville, pursuing her Ph.D. degree in Civil Engineering major in Transportation Engineering and her Master's degree in Statistics. Ms. Zhang's research contains intelligent transportation systems, transportation safety, and connected vehicle.

During her Ph.D. study, Ms. Zhang received several scholarships and awards, including T. Darcy Sullivan TSITE Scholarship and Graduate Student Senate Travel Award. Ms. Zhang is the member of ITE student group. She served as the website master of ITE Student Chapter at The University of Tennessee. Ms. Zhang also served as a reviewer for academic journals and conferences.