Identifying High Crash Risk Roadways through Jerk-Cluster Analysis

Seyedeh Maryam Mousavi

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses

Part of the Civil and Environmental Engineering Commons

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_theses/159

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
IDENTIFYING HIGH CRASH RISK ROADWAYS THROUGH JERK-CLUSTER ANALYSIS

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science

in

The Department of Civil and Environmental Engineering

by
Seyedeh Maryam Mousavi
B.S., Shiraz University, Iran, 2012
December 2015
ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Brian Wolshon, for his assistance with my thesis completion, his support throughout my study at LSU, and all the hope he has given to me. Also, a special thanks to Dr. Pande, Dr. Wilmot, and Dr. Parr who were always available to help me, and it was an honor to have them as the committee members.

I also would like to acknowledge the significant technical contributions of Dr. Joshua Kent of the LSU Center for GeoInformatics and Katherine Spansel, P.E., formerly of LSU’s Gulf Coast Center for Evacuation and Transportation Resiliency.

Additionally, none of these would have been possible without the love and patience of my family. My parents and my brother to whom this dissertation is dedicated have been a constant source of love, support, and strength all these years. My extended family also has aided and encouraged me throughout this endeavor.

Finally, I gratefully acknowledge the financial support of the National Science Foundation (NSF) in making possible the work presented here through Grant #0927123 and Grant #0927358. Additional research support was provided by the Gulf Coast Center for Evacuation and Transportation Resiliency, a United States Department of Transportation sponsored University Transportation Center and part of the Southwest Transportation University Transportation Center (SWUTC).
# TABLE OF CONTENT

ACKNOWLEDGMENTS ........................................................................................................ ii

LIST OF TABLES .................................................................................................................. v

LIST OF FIGURES ............................................................................................................... vi

ABBREVIATIONS AND ACRONYMS .................................................................................. vii

ABSTRACT ........................................................................................................................... viii

CHAPTER 1. INTRODUCTION ............................................................................................... 1

CHAPTER 2. LITERATURE REVIEW ....................................................................................... 5

2.1. GPS and Naturalistic Driving Data Collection ................................................................. 5
2.2. Scale of the Traffic Safety Study ...................................................................................... 6
   2.2.1. Macro Scale .............................................................................................................. 7
   2.2.2. Micro Scale ............................................................................................................ 7
2.3. Crash Relevant Variables ............................................................................................... 8
   2.3.1. Jerk ........................................................................................................................ 8
   2.3.2. Average Daily Traffic (ADT) ................................................................................ 12
   2.3.3. Presence of Horizontal Curvature ......................................................................... 12

CHAPTER 3. METHODOLOGY ............................................................................................... 14

3.1. Data Collection ............................................................................................................. 14
   3.1.1. Data Collectors (GPS Data Loggers) .................................................................. 15
   3.1.2. Participants .......................................................................................................... 16
   3.1.3. Average Daily Traffic and Presence of Horizontal Curvature .............................. 17
   3.1.4. Roadway Selection ............................................................................................... 18
3.2. Data Processing .......................................................................................................... 19
   3.2.1. Filtering the GPS Data ......................................................................................... 19
LIST OF TABLES

Table 1. NB Crash Estimation Model with Jerk Ratio for LA 42 ........................................32
Table 2. NB Crash Estimation Model with Jerk Ratio for LA 1248 ........................................32
Table 3. Full Log Likelihood Values for Models with Jerk Ratio ...........................................34
Table 4. NB Crash Estimation Model with Jerk Ratio and Presence of Curvature for LA42 ......35
Table 5. NB Crash Estimation Model with Jerk Ratio and Presence of Curvature for LA 1248 ..35
Table 6. Full Log Likelihood Values for Models with Jerk Ratio and Presence of Horizontal Curvature ...........................................................................................................36
Table 7. Comparison between Jek Ratio Significance Level in LA and CA ...............................37
LIST OF FIGURES

Figure 1. Implemented GPS Data Loggers ................................................................. 15

Figure 2. Participants’ Demographic Information ..................................................... 17

Figure 3. Roadways of Analysis, Baton Rouge, LA .................................................... 18

Figure 4. GPS Noise ............................................................................................... 20

Figure 5. GPS Wandering ...................................................................................... 21

Figure 6. GPS Gap ................................................................................................. 21

Figure 7. Sensitivity Analysis Results for LA 42 ..................................................... 27

Figure 8. Sensitivity Analysis Results for LA 1248 .................................................. 27

Figure 9. Crash and Jerk Rate per Quarter-Mile Segments on LA 42 ..................... 29

Figure 10. Crash and Jerk Rate per Quarter-Mile Segments on LA 1248 ............. 29

Figure 11. Quarter Mile Segments of the LA 42 and LA 1248 ............................ 31
ABBREVIATIONS AND ACRONYMS

1. ADT: Average Daily Traffic
2. GIS: Geographical Information System
3. GPS: Global Positioning System
4. HDOP: Horizontal Dilution Of Precision
5. LSU: Louisiana State University
6. NEMA: National Marine Electronics Association
7. NSF: National Science Foundation
8. PDOP: Position Dilution Of Precision
9. SAS: Statistical Analysis Software
10. SPSS: Statistical Package for the Social Sciences
11. SWUTC: Southwest Transportation University Transportation Center
12. UTC: Universal Time Code
13. VDOP: Vertical Dilution Of Precision
14. WGS 84: World Geodetic System 84
ABSTRACT

The state-of-the-practice for most municipal traffic agencies seeking to identify high-risk road segments has been to use prior crash history. While historic traffic crash data is recognized to be valuable in improving roadway safety, it relies on prior observation rather than future crash likelihood. Recently, however, researchers are developing predictive crash methods based on “abnormal driving events.” These include abrupt and atypical vehicle movements thought to be indicative of crash avoidance maneuvers and/or near-crashes. Because these types of near-crash events occur far more frequently than actual crashes, it is hypothesized that they can be used as an indicator of high-risk locations and, even more valuably, to identify where crashes are likely to occur in the future. This thesis describes the results of research that used naturalistic driving data collected from global positioning system (GPS) sensors to locate high concentrations of abrupt and atypical vehicle movements in Baton Rouge, Louisiana based on vehicle rate of change of acceleration (jerk). Statistical analyses revealed that clusters of high magnitude jerk events while decelerating were significantly correlated to long-term crash rates at these same locations. These significant and consistent relationships between jerks and crashes suggest that these events can be used as surrogate measures of safety and as a way of predicting safety problems before even a single crash has occurred.
CHAPTER 1. INTRODUCTION

The state-of-the-practice for most transportation agencies in identifying high-risk road segments has been the use of long-term historic traffic crash data. Traffic engineers analyze these data to quantify crash frequency, rate, severity, and economic loss; and locations with higher levels of these metrics are typically classified as “less safe” compared to other locations. Since crashes are used as the primary measure to assess the need for and potential benefit of highway safety investments, these “less-safe” locations (also called “black spots/zones/routes” or hazardous locations”) usually receive more attention in the form of funding for safety improvements [(Hauer 1996), (Tarko and Kanodia 2004), and (Persaud, Lyon, and Nguyen 1999)].

While analysis of historic traffic crash data has proven to be a valuable tool in improving roadway safety, it is a retroactive measure. In effect, it can only be used after damage, injury, and loss of life have occurred; and in many cases these losses must be sustained for several years. Second, a crash might occur due to several causes since it is a complex set of interactions among drivers, vehicles, and roadway environments. And third, crashes may either not be recorded uniformly or even sometimes non-injury ones not be reported at all. Consequently, the lack of reliability and availability of enough crash data has hampered accurate statistical analysis [(AASHTO 2010), (Chin and Quek 1997), (Dixit et al. 2011), and (Gettman et al. 2008)].

Simply using all the efforts to find crash prone locations is not beneficial for detecting risky and dangerous roadway segments because drivers’ accident histories do not necessarily reflect the quality of their driving habits, and considering that the crashes alone have the potential to identify the black spots incorrectly. Therefore, looking for complementary variables is important in
detecting crash prone locations earlier and at a higher levels of accuracy [(Nguyen et al. 2014) and (RISER 1985)].

New avenues of research have focused on quantifying parameters that are correlated with roadway safety, including measures such as conflicts and time-to-collision. Such variables, referred to as “surrogate safety measures,” are observable, non-crash events that have a relationship with crashes (Tarko and Kanodia 2004). The advantages of surrogate safety measures are that they can be observed long before crashes have occurred and they often happen far more frequently than actual crashes; in some cases 10 and 15 times more frequently (Guo et al. 2010). Surrogate safety measures also enable more powerful statistical analyses to be applied because of a larger sample size and a smaller error rate.

“Risky” or “unsafe” drivers are drivers that self-report being involved in an unusually high number of crashes. Surrogate safety measures developed from the analysis of human factors reflect the characteristics of these drivers. Bagdadi and Varhelyi (2011) found that unsafe drivers experienced an unusually large number of high magnitude negative jerks. The jerk characteristic of a driver is the rate of change in acceleration and is measured in feet per second cubed (ft. /s³). Drivers who showed a pattern of high negative jerk values were also statistically more likely to have a higher number of traffic crashes. This method of identifying high-risk road users by analyzing their jerk patterns has promise in the field of human factors and insurance pricing. However, previous research has not yet been able to leverage this emerging safety research into a method to identify unsafe roadway segments. In this research, it is hypothesized that locating high concentrations of abnormal negative jerk values (i.e., jerk-clusters) would enable high-risk locations to be identified in advance and potentially with greater accuracy. By examining jerk-
clusters instead of individual drivers, this analysis can also be used to illustrate a general trend of safety for a roadway segment.

To address this problem, recent research has sought to develop measures to proactively identify high-risk roadway segments [(AASHTO 2010), (Dixit et al. 2011), and (Gettman et al. 2008)]. If successful, these techniques, in theory, can be valuable because they provide the opportunity to detect adverse safety conditions before any crashes, injuries, or fatalities occur.

This thesis describes the results of recent and ongoing research that uses naturalistic driving data collected with global positioning system (GPS) sensors. The research applies Geographic Information Systems (GIS) analysis and other quantitative methods to the GPS data to use jerks-clusters for future crash prediction. From this dataset, jerks were geo-located to a roadway network and graphed in combination with historic traffic crashes. In addition, Average Daily Traffic (ADT) and presence of horizontal curvature were added to the database. The ADT was included to compute the crash rate (number of crashes divided by ADT) per segment, which is the total number of crashes per average daily traffic (measured by the thousands) rather than just the total number of crashes. These datasets were analyzed to determine if a statistical relationship existed between the crash rate, jerk ratio, and presence of horizontal curvature. Such information would, in effect, give safety analysts a “crystal ball” to see into the future and predict locations that are likely to have a higher numbers of crashes in the future. With this knowledge, safety improvements could be planned and implemented to prevent damage, injury, and death well before any significant number of crashes had even occurred.
This thesis includes the following chapters, respectively:

- Chapter 1. Introduction.
- Chapter 2. Literature Review
- Chapter 3. Methodology
- Chapter 4. Conclusion
CHAPTER 2. LITERATURE REVIEW

Prior to beginning the development of finding surrogate measures of safety, a review of previous research was conducted. In this chapter, the extensive use of naturalistic and GPS driving data (how they are collected and why they are collected) will be explained. Traffic studies and naturalistic driving data are used in practice on two levels, macro and micro, which will be explained in this chapter as well. The most important part of this chapter is a review of literature that consider jerk, average daily traffic (ADT), and presence of curvature in crash prediction models, and their impacts on roadway safety. Eventually, the chapter will be ended up with a conclusion section.

2.1. GPS and Naturalistic Driving Data Collection

The advent of geographic information systems (GIS) and global positioning systems (GPS) have allowed driving data to be collected and processed with continuously increasing levels of detail and accuracy. This technology has been applied extensively in the traffic engineering field in areas ranging from fleet management and dashboard navigation to travel time estimation and traffic simulation. It has been suggested that this technology has potentially significant applications in the area of quantifying traffic safety (Pande et al. 2014) and has the potential to replace traditional methods of measuring safety [e.g., (Hauer 1996)]. In addition, a large naturalistic driving database provides the opportunity to find out how, why, and where crashes occurred (Jonasson and Rootzén 2014).
Studies that have adapted in-vehicle GIS and GPS technology to analyze driving behavior are classified as “naturalistic driving studies” because they use driver information collected during their natural driving routines (Dingus et al. 2006). Naturalistic driving data has many applications, and traffic safety has received the most attention in recent years.

The “100-Car Study” used 100 vehicles equipped with video recording devices and advanced GPS sensors to collect naturalistic driving behavior from 241 drivers. Advanced instrumentation was used to document 8,295 critical incidents (Dingus et al. 2006). The 100-car study led to significant advances towards crash-avoidance systems used in modern vehicles (McLaughlin, Hankey, and Dingus 2008).

There is another study called Australian Naturalistic Driving Study (ANDS) that is being carried out by Transportation and Road Safety (TARS). ANDS will be outfitting 400 cars to collect data within and outside of the vehicle. These instruments reflect data about the drivers’ behavior, the vehicles’ behavior (speed, lane position, and etc.), and road use behavior. This study analyzes the observed data in order to develop a novel method for improving roadway safety (Regan et al. 2013).

2.2. Scale of the Traffic Safety Study

Traffic safety studies with the potential of estimating the crashes proactively have been completed in two scales: the macro and micro. Macro-models attempt to estimate crash frequency over an area using zonal characteristics at the planning level [(Ladrón de Guevara 2004) and (Quddus 2008)], while micro-level models are used to estimate crash frequency on a corridor or a small portion of road.
2.2.1. Macro Scale

As mentioned, macro-level models explain crash occurrence in a zone (e.g., Traffic Analysis Zones used in long-range planning), rather than an individual corridor or segment (Hadayeghi, Shalaby, and Persaud 2007). Studies at the macro level are generally implemented using different modeling approaches including Negative Binomial (NB), Poisson Regression (PR), and Geographically Weighted (GW) models which estimate the effect of macro level variables on the total number of crashes or crash rate.

Prior research on macro-models, in general, found that the variables of population density, vehicle-miles traveled (VMT), and the number of households are positively correlated to vehicle crashes. Other variables include average posted speed and volume/capacity ratio; these were found to be negatively correlated to crash occurrences [(Hadayeghi, Shalaby, and Persaud 2003), (Hadayeghi, Shalaby, and Persaud 2007), (Ladrón de Guevara, Washington, and Oh 2004), and-(Miaou 1994)].

2.2.2. Micro Scale

Micro level research estimates vehicle crashes for a single roadway, corridor, or segment of a roadway. Hosseinpour et.al studied 448 segments of five federal roads in Malaysia to evaluate the effect of roadway geometry on head-on collisions. Results of a random-effect negative binomial (RENB) model showed that the horizontal curvature, terrain type, heavy-vehicle traffic, and access points all significantly contributed to the likelihood of head-on collisions, while shoulder width reduced the likelihood of such crashes (2014). Another study by Milton and Mannering also confirmed the positive impact of AADT on the total number of crashes per segment in Washington State (1996).
A more recent data related development for micro level scale analysis has been the use of driver behavior parameters derived from the GPS data. Specifically, a higher rate of driving events that involve higher jerk (rate of change of deceleration) have been found to be significantly related to crash frequency in road segments [(Bagdadi 2013), (Bagdadi and Várhelyi 2011), (Bagdadi and Varhelyi 2013), and (Pande et al. 2014)].

2.3. Crash Relevant Variables

Due to the rare occurrence of traffic crashes and the deficiency of adequate data, transportation professionals have sought to develop surrogate safety measures to better understand, analyze, and prevent vehicle crashes. The following sections discuss the number of variables in previous research that were suggested as surrogate measures of safety.

2.3.1. Jerk

Recently, several studies have sought to use driver behavior on roadway segments as an explanatory variable for crash frequency [e.g. (Bagdadi and Várhelyi 2011), (Bagdadi 2013), and (Bagdadi and Varhelyi 2013)]. Previous preliminary work by the authors has also used this link to identify concentration of driving events with higher rates of change of deceleration (i.e., “jerks”) as a potential surrogate measures for crash frequency at the segment level on both interrupted and uninterrupted flow segments [(Pande et al. 2014) and (Mousavi et al. 2015)]. However, before any of these measures could be confirmed as surrogates, their relationship with crash frequency needs to be thoroughly established. The following section explains the studies that evaluated the correlation between surrogate measures of safety and crashes.
Comparing to deceleration that has been used as an indicator of crash prone locations commonly, jerk (rate of change of acceleration) has shown a more promising results for detecting unsafe roadway segments. Previous research done by Dingus et al. (2006) and McLaughlin (2008) showed the deficiency of using acceleration/deceleration in distinguishing between safety critical events and acting upon them. The statistical analysis results also indicated that there is no statistically significant difference in deceleration between normal and sudden braking driving events (Bagdadi and Varhelyi 2013). In other words, focusing on acceleration/deceleration as a surrogate measure of safety, it only considers high deceleration events when a driver rapidly applies maximum longitudinal deceleration (jerk equals to zero) and does not record milder incidents with a lower deceleration (but higher rate of change of deceleration). The advantage of studying rate of change of deceleration on deceleration is that incidents involving milder braking reactions (low deceleration and high rate of change of deceleration) occur far more frequently than the events with a high deceleration (zero or small jerk) and enables detecting crash prone locations earlier with a higher level of accuracy.

Various techniques have been used to develop surrogate safety measure but the most prominent has involved the use of traffic [(Chin and Quek 1997), (Chin and Quddus 2003), and (RISER 1985), (Glauz 1980) (Parker and Zegeer 1989)] and time-to-crash [(Hydén 1987), (Svensson 2006), and (Svensson and Hydén 2006)]. These studies share the common goal of identifying where or when traffic crashes may occur and the factors that increase the likelihood of occurrence.

The Highway Safety Manual (HSM) is a reference tool for traffic professionals to facilitate improved decision-making based on safety performance. The HSM allows for a quantitative analysis of safety for highway planning, programing, project development, construction,
operations, and maintenance. The goal of the HSM is to assemble the most current methodologies and information for measuring, estimating, and evaluating roadway safety in terms of crash frequency and severity. Published in 2010, the HSM includes Safety Performance Functions (SPF) that permits the expected future numbers of crashes to be forecast on specific segments of roadway with a particular set of traffic, design, and operational conditions (AASHTO 2010).

Currently, safety critical braking maneuvers (known as jerks) are of high interest of roadway safety assessors. Safety critical driving maneuvers are events where abrupt braking is due to either the drivers or road users (Bagdadi and Varhelyi 2013).

Riser studied 201 driver errors—defined as behaviors that lead to a traffic conflict or accidents—resulting in crashes that occurred on road segments along standardized routes in Vienna, Australia. Research indicated all variables are positively correlated to the crashes. In fact, a driver with more errors was found to be involved in more traffic conflicts and reported more crash incidents (1985).

Bagdadi and Varhelyi (2011) used GPS data to analyze braking characteristics and jerks from 166 private vehicles. Using a critical jerk threshold of -32.4 ft./s³ (-9.9 m/s³), the research found drivers that showed a pattern of jerk values above this threshold were more likely to have history of self-reported crashes. More recent studies with naturalistic driving data have used critical jerk values to differentiate between controlled powerful braking and “critical braking” for crash avoidance. Results indicated that the threshold value of 1.5 g/s for critical and 1 g/s for potentially critical events are appropriate to detect the traffic conflicts. Bagdadi also developed a method for detecting crash prone locations, improving safety estimation by using Poisson regression based on a -1 g/s critical jerk threshold [(Bagdadi and Várhelyi 2011), (Bagdadi and Varhelyi 2013), and (Bagdadi and Varhelyi 2013)].
These previous studies have examined individual drivers and their jerk characteristics to identify unsafe driving behavior. However, an analysis of jerk has not yet been applied to multiple drivers over same road to examine the relationship between jerk-clusters and roadway safety.

The authors’ previous work has also found that the rate of higher jerk events on uninterrupted traffic flow is highly correlated to the total number of crashes on quarter mile segments of an uninterrupted traffic flow of Highway 101, San Louis Obispo, CA [(Pande et al. 2014) and (Loy 2013)]. Similar results were seen in an interrupted flow highway where a higher jerk ratio was associated with higher crash frequency (Mousavi et al. 2015). While the GPS data collection approach has been similar, the authors’ prior research, [(Mousavi et al. 2015) and (Pande et al. 2014)], differs from the work done by Bagdadi et al. (2013) in defining the threshold for what constitutes a “high” jerk value. The threshold value used in Mousavi et al.’s study (2015) was much lower than what had been suggested by Bagdadi. Bagdadi (2013) et al. identified spikes in the rate of change of acceleration that are indicative of atypical breaking as seen in crash avoidance and/or near miss events. Since the authors’ past and present goals have been to explain crash frequency in such events, a data-driven approach was adopted. The threshold for a “high” rate of change of deceleration was established so that the correlation is maximized between crash rate and the rate of events that involve a higher rate of change in deceleration. As a result, the authors’ prior work, [(Mousavi et al. 2015) and (Pande et al. 2014)], has used a threshold in the range of -2.5 ft./sec³, which is less than 1/10th of the threshold used by Bagdadi et al. The implication of a lower threshold is that the jerk events physically above this threshold are considered “somewhat critical” braking events. In addition, the quality of GPS data (compare 3Hz data in the current study with 5Hz data in Bagdadi’s research) affects the jerk threshold. In effect, the lower quality of data needs a lower jerk threshold to be able to capture all sudden braking maneuvers (Bagdadi 2013).
2.3.2. Average Daily Traffic (ADT)

Several studies have been conducted to evaluate the impact of ADT on the total number of vehicular crashes.

Kumar et al. studied road crashes on a sample of India’s National Highways by using Bayesian modeling techniques. This study was done to identify the most critical safety variables. The results showed that the Poisson Weibull model predicts crashes with improved accuracy. Results indicated that traffic volume, access roads, and the presence of a school positively affect the total number of crashes (2013).

Another study by Jung et al. showed that even though AADT’s impact is not very large, it has a positive significant relationship to the total number of crashes on rainy days (2014). Studying crashes by using a Bivariate Zero-Inflated Negative Binomial (BZINB) model showed Annual Average Daily Traffic of trucks had a positive impact on the total number of car-to-truck and truck-to-truck crashes (Dong et al. 2015). Daniel et al. developed two truck crash prediction models: the NB and Poisson. The length of the horizontal curve and crest curve grade rate were significant in both (Daniel, Tsai, and Chien 2002). Ferreira’s analysis was based on a binary model which also confirmed the high correlation between ADT and the total number of crashes per roads segments (2015).

2.3.3. Presence of Horizontal Curvature

Several studies have examined the effect of the geometric design of highways, including curvature, on the total number of crashes. Applying both a negative binomial model and by using a full Baye’s method evidenced that there is a significant positive relationship between the number
of crashes and the highway’s horizontal curvature (Schneider et al. 2009). Other accomplished studies also confirmed the effect of horizontal curvature and a degree of horizontal curvature having an impact on the crash occurrence [(Dong et al. 2015) and (Miaou 1994)].
CHAPTER 3. METHODOLOGY

This research seeks to estimate and forecast roadway safety by analyzing clusters of high magnitude negative jerks from multiple driver observations. The research methodology consisted of four analytical tasks including:

1) Data collection: Explains the implementation of GPS data loggers, GPS data collection participants, roadway selection, Average Daily Traffic data, and data about roadways curvature.

2) Data processing: Considers the GPS data errors and linear referencing of GPS and crash data. In addition, jerk analysis, division of the roadways, crash rate analysis, and finally interpolating average daily traffic data will be discussed.

3) Sensitivity analysis: Covers the process of implementing a sensitivity analysis to get an appropriate jerk threshold.

4) Crash frequency modeling: The final task was to develop crash-frequency models, establishing a quantitative relationship between historic crash data, jerk-clusters, and the geometric design roads.

3.1. Data Collection

The jerk data used for this study was obtained from a naturalistic driving dataset collected through GPS data loggers. The process of collecting the GPS data, recruiting and screening the participants, and attaining average daily traffic data points and horizontal curvature of roadways will be explained in the following sections.
3.1.1. Data Collectors (GPS Data Loggers)

GPS data were collected by GPS Data Loggers V3.15, Figure 1, able to record daily travel patterns for up to two weeks. The data loggers were charged by a 3.7 volt rechargeable battery and recorded data in a comma-separated format, adhering to the National Marine Electronics Association (NEMA) standards. The data logger recorded position information in World Geodetic System 84 (WGS 84) standard, collecting latitude, longitude, altitude, heading, speed, number of satellites utilized, position dilution of precision (PDOP), horizontal dilution of precision (HDOP), and vertical dilution of precision (VDOP), as well as universal time code (UTC) and date.

Figure 1. Implemented GPS Data Loggers
Readings from the device were recorded at a rate of three hertz (i.e., three readings per second) and was programmed with a “sleep” mode which would stop data recording if the position information did not change after 300 seconds. This was done primarily to preserve battery life.

The data loggers were housed in a waterproof crush-proof, airtight case to protect the unit and hard drive from damage. Drivers were asked to place the unit in their center console or glove box. Extensive testing prior to the data collection period showed that neither location affected the accuracy of the data obtained from the device. Participants were also asked to refrain from allowing anyone else to use their vehicle during the 2-week experiment period.

3.1.2. Participants

The naturalistic driving data used in this research was collected from 31 staff members and their household members at Louisiana State University. The goal of the participant selection was to recruit drivers with similar driving patterns. This was necessary to ensure that the participants in the study would consistently travel over the same routes to develop jerk-clusters. Participants were selected using a questionnaire that screened drivers based on personal characteristics and driving patterns. Participants were then segregated based on age, commuting route, and driving frequency. To maintain confidentiality, a unique random identifier was assigned to each participant that was later assigned to the GPS data collection unit.

Participants were between the ages of 20 and 65 and included 12 female and 19 male drivers. Figure 2 shows the age and gender distribution of all the participants.
The data collection period spanned from July 2012 to January 2013 with each driver contributing around 10 days of data. Similar data (for analysis of uninterrupted flow facility; US 101 freeway) were collected in San Luis Obispo, California (Loy 2013) but were not used in this analysis.

3.1.3. Average Daily Traffic and Presence of Horizontal Curvature

For this study, Average Daily Traffic (ADT) count data was accessed through LADOTD’s estimated annual average daily traffic count webpage (Albritton 2014), and roadway geometric features were drawn from Google maps (GoogleMap 2014).
3.1.4. Roadway Selection

The Baton Rouge road network used for this study was provided by the Louisiana Department of Transportation and Development (LDOTD 2014). The roadways evaluated in this study are depicted in Figure 3.

1) A 7.5 mile long stretch of LA 42, known locally as Burbank Drive, a four-lane divided highway with posted speeds varying from 45 to 55 mph;

2) A 5.15 mile long stretch of LA 1248, known as Bluebonnet, a four-lane divided highway with posted speeds varying from 30 to 45 mph.

Figure 3. Roadways of Analysis, Baton Rouge, LA
The selected roadways in this study were interrupted state highways, compared to those in the Pande et al. (2014) study, which were uninterrupted highways.

3.2. Data Processing

The collected data should be filtered and processed as input for the final regression model. In this section, the process of filtering and preparing the data will be discussed.

3.2.1. Filtering the GPS Data

The GPS data collected from each participant was processed by combining multiple data files into a single file, labeled with each participant’s unique identifier. This data file was then linked to the initial questionnaire filled out by the participant. The dataset was divided into individual trips, which were delineated by 30-minute intervals of inactivity. These were then assigned a trip number. The data was then visually inspected for errors. These errors may originate either from external sources (such as road surface roughness and vehicle vibrations) or internal measurement uncertainties caused by the measuring equipment itself (Bagdadi 2013). Three error types were found in the data:

1) GPS noise at intersections: GPS noise was the most common of the three errors. It occurred when the GPS position changed but speed and heading did not match. This error was typically associated with vehicles being stopped or moving at slow speeds and was characterized by a “cluster” of data points observed in the vicinity of the stopping location, Figure 4;
2) GPS wandering: GPS wandering occurred when the GPS data points did not correspond to the actual vehicle location. This phenomenon was characterized by data points appearing in areas where no road exists and appeared to be random in nature, Figure 5;

3) GPS Gaps: The third error, the presence of gaps in the data, was identified by segments of roadways that were void of data collection points. This error was likely caused by a temporary loss in GPS satellite availability resulting from impedance or a communication error, Figure 6.

These errors were consistent with those observed in the San Louis Obispo, CA (Pande et al. 2014).
Figure 5. GPS Wandering

Figure 6. GPS Gap
Bagdadi and Várhelyi (2011) believed that the higher frequency data have a higher level of noise, and more filters should be applied to the data to remove the existing fluctuations and uncertainties. Because this study’s data frequency was 3 Hz (compared to other studies), not many filters needed to be applied. So, the Savitzky-Golay filter was used and resulted in a marginal improvement.

Eventually, the extracted filtered data was added to a GIS file and converted into a shape file for further data processing.

3.2.2. GIS Linear Referencing

Linear referencing was used to link the data collection points to the road network of Baton Rouge, LA. Linear referencing links geographic locations (x, y) to a measured linear feature so that linear distances between each point can be calculated [(ESRI 2003) and (ESRI 2012)].

This linear referencing procedure was repeated for each individual participant. Data points were linearly referenced to the nearest road within a radius of 300 ft. This large radius value was selected to bypass any possible GPS noise or wandering errors. However, this required the GPS data points to be further filtered to remove any erroneous location information that may have resulted from referencing the data to an incorrect road. Furthermore, data points with incorrect heading information, as well as data points that reported communication with fewer than five satellites, were removed from consideration.

The removed data constituted about one percent of the overall data collected. This filtering process ensured that only accurate data with reliable position information was included in the analysis.
3.2.3. Jerk Analysis

A histogram of acceleration ($a$) and jerk ($j$) was analyzed for quarter-mile segment lengths to understand the distribution of these variables. This analysis indicated that the jerk variable ($j$) warranted further evaluation. This was logical since low magnitude negative jerk values are likely associated with gradual braking maneuvers, as opposed to high magnitude negative jerk values, indicating sudden braking. As seen in Pande et al. (2014), drivers traveling on freeway segments characterized by high long-term crash rates had to make more sudden braking maneuvers, resulting in an above average rate of high jerk values on those segments.

In this study, jerk was used as the primary variable to distinguish between harsh and normal driving maneuvers and to detect hazardous road segments. Each participants’ acceleration ($a$) and jerk ($j$) between consecutive data points calculated from the driving data (exact location and time of a vehicle with the rate of 3 Hz). Equations (1) and (2) are given for calculating acceleration ($a$) and jerk ($j$), respectively.

\[
a = \frac{dv}{dt} \quad \text{(1)}
\]

\[
j = \frac{da}{dt} \quad \text{(2)}
\]

Where:

\[a: \quad \text{Acceleration (ft/s}^2\text{)}\]

\[dv: \quad \text{Change in velocity between successive observations (ft. /s)}\]

\[j: \quad \text{Jerk (ft/s}^3\text{)}\]

\[da: \quad \text{Change in acceleration between successive observations (ft/s}^2\text{)}\]
3.2.4. Segmenting the Roadways

Safety studies are typically conducted at either the macro or micro level. In this study, the micro scale was of interest. The roads in this study were divided into three different segment lengths of short, medium, and long (an eighth, a quarter, and a half mile) in three different shape files to determine an appropriate segment length.

3.2.5. Crash Rate Analysis

To identify the roads segments of LA 42 and LA 1248, which were deemed “safe” or “hazardous,” 5-year vehicle crash data, from 2009 to 2013, were used. Crash history is a commonly accepted performance measure for relative safety (Hauer 1996).

The historic crash data were also linearly referenced to the same road network as the GPS data. The crash rate was then calculated for each segment in accordance with the United States Department of Transportation (USDOT) defined method, Equation (3) (FHWA 2011). This was used to normalize the total number of crashes by the amount of ADT, a measure of exposure.

\[ R = \frac{C \times 100,000,000}{V \times 365 \times N \times L} \]  

(3)

Where:

- \( R \): Road segment crash rate expressed for 100 million vehicle-miles traveled
- \( C \): Total number of crash on the roadway segment
- \( V \): Traffic volume (ADT)
- \( N \): Number of years of crash data
- \( L \): Length of the road segment
Based on Equation 3, the following variables were acquired for calculating the crash rate per segment:

1) Total number of crashes: using 5-year vehicle crash data from 2009 to 2013, there were a total of 1188 crashes on LA 42 and 1352 crashes on LA 1248.
2) Length of the segment: The crash rate value was calculated based on three different segment lengths of an eighth, a quarter, and a half mile.
3) Average Daily Traffic: Interpolated average daily traffic data (explained in section 3.2.6) was used for the crash rate calculation.

3.2.6. Average Daily Traffic

To calculate crash rate (number of crashes/ADT) and crash rate (Equation 3) per segment, an ADT value per segment should be available. Although the ADT data points provided by LADOTD (Albritton 2014) are finite and do not provide a value per segment, the Inverse Distance Weighted (IDW) interpolation tool in GIS was utilized to provide an ADT value per segment.

IDW interpolates measured samples for predicting values for un-sampled locations (ESRI 2007). In this case, the measured samples were ADT data points provided by LADOTD, and the un-sampled locations were estimated based on these values and calculated using the IDW method.

3.3. Sensitivity Analysis

Before conducting any crash data analysis or modeling, it was necessary to determine an appropriate jerk threshold to differentiate between normal and evasive driving maneuvers.
The jerk \((j)\) value is a continuous variable without an intuitive method for identifying which threshold value of jerk distinguished between gradual braking maneuvers and sudden braking maneuvers. Bagdadi and Varhelyi (2011) used a constant jerk threshold of -32.4 ft/s\(^3\) (-9.9 m/s\(^3\)). The threshold value used for this research was determined through data driven sensitivity analysis. Although using a lower jerk threshold value includes a higher percentage of the total number of data points, it enables the capture of all abrupt braking events in a controlled manner (Loy 2013).

In this study, twenty-one jerk value thresholds were evaluated in the sensitivity analysis; their thresholds started at -0.5 ft. /sec\(^3\) and decreasing by 0.5 ft. /sec\(^3\) until reaching -10.5 ft. /sec\(^3\). A jerk event was defined as any negative jerk value with a magnitude that exceeded a predefined threshold value. The number of jerk events exceeding the threshold value within a road segment was counted.

The count for each segment was normalized by the total number of data points observed in that segment to yield a jerk-ratio or percentage of the high jerk events, observed in each segment. The 21 jerk-thresholds were then correlated with the crash rates for each segment using a Pearson’s Correlation Coefficient analysis, by SPSS. Also, because it was unknown how the segment length would affect the analysis, three different segment lengths were selected: eighth-mile (0.125), quarter-mile (0.25), and half-mile (0.5).

All the above mentioned analyses were done for LA 42 and LA 1248 separately to investigate the consistency in relationship between crashes and jerk ratios in different roads.

The results of the sensitivity analyses are shown in Figure 7 and Figure 8.
Figure 7. Sensitivity analysis results for LA 42

Figure 8. Sensitivity Analysis Results for LA 1248
Figures 7 indicates a jerk threshold value of -2.5 ft. /s³ resulted in the highest Pearson’s Correlation Coefficient for all segment lengths of LA 42, which had the highest correlation coefficient for the quarter mile segments.

Similarly in Figure 8, a jerk threshold of -2.5 and the quarter mile segment had the highest correlation coefficient for LA 1248.

Compared to other studies, the jerk threshold value of -2.5 is much smaller and is likely too low to be indicative of an evasive driving maneuver. But a larger concentration of such braking events may be indicative of drivers having to often react somewhat urgently while encountering “less than ideal” combination of traffic and geometric conditions on a segment. On the other hand, the benefit of a reduced threshold value increases the number of jerk events, enabling high-risk segments to be identified earlier and with greater statistical certitude. Additionally, a lower frequency of GPS data (in this case 3 Hertz) gives a lower resolution, resulting in a smaller value of jerk threshold (Bagdadi and Várhelyi 2011).

The results of each of the different segment lengths suggested that the smaller, more homogenous eighth-mile segments may have been smaller than the precision of the crash location data, resulting in crashes being incorrectly assigned to the adjacent segments by linear referencing. And the half mile segments appeared to be too large and insensitive to generate the high Pearson Coefficients seen in the quarter-mile segments.

Figures 9 and 10 present heat maps to graphically illustrate the strength of the correlation between the jerk ratio and historic crash rate.
Figure 9. Crash and Jerk Rate per Quarter-Mile Segments on LA 42

Figure 10. Crash and Jerk Rate per Quarter-Mile Segments on LA 1248
3.4. Crash Frequency Modeling and Log Likelihood Ratio Test

The over-dispersion of data is one reason crash analysis is difficult. Additionally, crash data is long-tailed—meaning there are few parts of a road that have a large number of crashes—which makes the variance greater than the mean. However, there are many different statistical tools for analyzing and modeling crash data, such as negative binomial, zero-inflated, Poisson Weibull, Poisson log-normal, etc. Negative binomial remains the most commonly used model in crash analysis [(Shankar, Mannering, and Barfield 1995), (Abdel-Aty and Radwan 2000), and (Zou, Wu, and Lord 2015)].

Two different crash rate regression models were developed for both LA 42 and LA 1248, resulting in a total of four models. The first two models used the jerk ratio as the only independent variable for estimating the crash rate, while the other two models included the presence of horizontal curvature in addition to the jerk ratio.

After implementing NB regression models, a comparison test was completed for each group of the models (first group included the jerk ratio as the only independent variable and the second group included the presence of curvature as well as the jerk ratio) to see if the two models of LA42 and LA1248 were similar as vector. This comparison was done with log likelihood ratio test.

The following sections discuss the NB and log likelihood ratio test for each group of models separately.
3.4.1. Crash Frequency Modeling- Independent Variable: Jerk Ratio

In this study, crash rate values (number of crashes/ ADT) were estimated for quarter-mile segments of each road separately using a negative binomial regression model in SAS. Figure 11 indicates there were 31 and 21 quarter-mile segments of LA 42 and LA 1248 respectively. It is notable that both directions of the traffic flow were combined since the crash data does not specify the direction.

Figure 11. Quarter Mile Segments of the LA 42 and LA 1248
Two crash rate regression models were developed for 31 segments of LA 42 and 21 segments of LA 1248. In these models, the jerk ratio was used as the only independent explanatory variable for estimating the crash rate. Tables 1 and 2 represent the results of NB regression models. The results with significant p-value at the %95 level of significance are highlighted.

Table 1. NB Crash Estimation Model with Jerk Ratio for LA 42

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>S.E.</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>4.1219</td>
<td>0.4360</td>
<td>3.2675 4.9764</td>
<td>89.39</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Jerk Ratio</td>
<td>1</td>
<td>0.4110</td>
<td>0.0988</td>
<td>0.2173 0.6048</td>
<td>17.29</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Dispersion</td>
<td>1</td>
<td>0.8486</td>
<td>0.1935</td>
<td>0.5428 1.3266</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>80113.2639</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td>427.0319</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. NB Crash Estimation Model with Jerk Ratio for LA 1248

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>S.E.</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>4.7246</td>
<td>0.4988</td>
<td>3.7469 5.7023</td>
<td>89.70</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Jerk Ratio</td>
<td>1</td>
<td>0.1966</td>
<td>0.0703</td>
<td>0.0588 0.3343</td>
<td>7.82</td>
<td>0.0052</td>
</tr>
<tr>
<td>Dispersion</td>
<td>1</td>
<td>0.6722</td>
<td>0.1900</td>
<td>0.3863 1.1698</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>55345.7847</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td>299.3578</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The p and estimate values in the previous two models reveal that the jerk ratio is highly significant and has a positive correlation with the crash rate. This suggests that, whenever there is a significant crash ratio in a quarter mile segment, there should be a high jerk ratio as well, which
means the jerk ratio is a good indicator of the crash ratio and for identifying hazardous road segments.

3.4.2. Log Likelihood Ratio Test_ Independent Variable: Jerk Ratio

The two models representing LA 42 and LA 1248 separately were compared with a model that was a combination of LA42 and LA1248 to assess their fit, by a log likelihood ratio. A log likelihood ratio test was performed to compare nested models, expressing whether the parameters in the two models are the same or not. The institute for digital research and education (idre) stating that the log likelihood value is always a negative value and indicative of a model fit. The smaller the log likelihood values, closer to zero, the better the model fits (2015).

Followings show the null and alternative hypothesis and formula (4) with which the log likelihood ratio test was carried out:

\[ H_0: \beta_1 = \beta_2 \]

\[ H_1: \beta_1 \neq \beta_2 \]

\[ -2 (\text{Full Log Likelihood Pooled} - \text{LogLikelihood 1} - \text{Log Likelihood 2}) \sim \chi^2, DF = 1 \]

(4)

Table 3 shows the full log likelihood values for LA 42, LA 1248, and the combined model of the both roads:
Table 3. Full Log Likelihood Values for Models with Jerk Ratio

<table>
<thead>
<tr>
<th>Road</th>
<th>Full Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA42</td>
<td>-48.46</td>
</tr>
<tr>
<td>LA1248</td>
<td>-35.85</td>
</tr>
<tr>
<td>LA42 and LA1248</td>
<td>-88.50</td>
</tr>
</tbody>
</table>

The formula 4 then was applied to the given full log likelihood values as below:

\[-2 (-88.50 + 48.46 + 35.85) \sim \chi^2, DF = 1\]

\[8.38 \sim \chi^2, DF = 1\]

\[P - value = 0.0037\]

So, based on the p-value that is statistically significant, it can be concluded that the models of LA42 and LA1248 are different.

3.4.3. Crash Frequency Modeling- Independent Variables: Jerk Ratio and Presence of Horizontal Curvature

The other two models developed include the geometric design of the roads (in this case the presence of horizontal curvature) in addition to the jerk ratio. Tables 4 and 5 are the output, and the significant results are highlighted. As it is seen from the tables, jerk ratio remains significant and positively affects the crash rate on the both roads, while the p-value for presence of curvature is only significant for LA 42 at %95 level of significance. Therefore, there is not enough evidence to establish that the presence of curvature assuredly affects the crash rate.

The estimated value of the presence of curvature is negative in Table 4, which means that the more curves present in a segment, the fewer crashes occurred. That might have been caused by
the drivers’ behaviors, such as reducing their speeds and driving more cautiously when they operated their vehicles around curves.

Table 4. NB Crash Estimation Model with Jerk Ratio and Presence of Curvature for LA42

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>S.E.</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>4.0822</td>
<td>0.4087</td>
<td>3.2813</td>
<td>4.8832</td>
<td>99.79</td>
</tr>
<tr>
<td>Jerk Ratio</td>
<td>1</td>
<td>0.4615</td>
<td>0.0976</td>
<td>0.2702</td>
<td>0.6527</td>
<td>22.37</td>
</tr>
<tr>
<td>Curve</td>
<td>1</td>
<td>-0.6653</td>
<td>0.3185</td>
<td>-1.2896</td>
<td>-0.0411</td>
<td>4.36</td>
</tr>
<tr>
<td>Dispersion</td>
<td>1</td>
<td>0.7700</td>
<td>0.1771</td>
<td>0.4906</td>
<td>1.2084</td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>80115.1236</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td>425.3124</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. NB Crash Estimation Model with Jerk Ratio and Presence of Curvature for LA 1248

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>S.E.</th>
<th>95% Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>4.6194</td>
<td>0.5498</td>
<td>3.5419</td>
<td>5.6969</td>
<td>70.60</td>
</tr>
<tr>
<td>Jerk Ratio</td>
<td>1</td>
<td>0.2004</td>
<td>0.0708</td>
<td>0.0617</td>
<td>0.3391</td>
<td>8.01</td>
</tr>
<tr>
<td>Curve</td>
<td>1</td>
<td>0.1010</td>
<td>0.2325</td>
<td>-0.3546</td>
<td>0.5567</td>
<td>0.19</td>
</tr>
<tr>
<td>Dispersion</td>
<td>1</td>
<td>0.6672</td>
<td>0.1887</td>
<td>0.3833</td>
<td>1.1614</td>
<td></td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>55345.8796</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td></td>
<td>301.1680</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.4.4. Log Likelihood Ratio Test: Independent Variables: Jerk Ratio and Presence of Horizontal Curvature

As in Section 3.4.2., a log likelihood ratio test was also done on the models which included the presence of curvature as well as the jerk ratio. Table 6 shows the full log likelihood values for LA42, LA1248, and the both roads together. Afterward, the log likelihood ratio test was applied.

Table 6. Full Log Likelihood Values for Models with Jerk Ratio and Presence of Horizontal Curvature

<table>
<thead>
<tr>
<th>Road</th>
<th>Full Log Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA42</td>
<td>-46.60</td>
</tr>
<tr>
<td>LA1248</td>
<td>-35.75</td>
</tr>
<tr>
<td>LA42 and LA1248</td>
<td>-88.09</td>
</tr>
</tbody>
</table>

\[
-2 (-88.09 + 46.60 + 35.75) \sim \chi^2, DF = 2
\]

\[
11.4712 \sim \chi^2, DF = 2
\]

\[
P \text{ - value } = 0.003229
\]

The p-value is still significant which indicates a significant difference between the two models.

But, comparing the full log likelihood values of LA42 and LA1248 in Table 3 with their corresponding values in Table 6 indicates that the models including the presence of curvature generally performed better than the models that only include jerk ratio. Briefly, it demonstrates that the inclusion of more related variables to the crash prediction models could improve models and offer better results.
3.4.5. Comparing Interrupted and Uninterrupted Traffic Flow Crash Frequency Models

Returning to the main goal of the thesis, the crash prediction models were developed for two interrupted traffic flow condition in order to evaluate the variables impact and consistency among the models. In conclusion, the jerk ratio analyses suggest that the crash rate experienced on LA 42 and LA 1248 in Baton Rouge, LA are significantly correlated to the high magnitude negative jerk-clusters rather than variables traditionally used in crash-frequency estimation. Therefore, jerk ratio could be a good measurement of predicting crash prone locations on interrupted traffic flow.

Results of a comparison between the current and Pande et al.’s (2014) study, based on uninterrupted traffic flow, is shown in Table 7. This indicates a high correlation between the jerk ratio and crash on all the roads in addition to the consistency between uninterrupted and interrupted traffic flow. This finding is promising as it demonstrates the potential of jerk ratio as being a surrogate measure of safety.

<table>
<thead>
<tr>
<th>State/ Road</th>
<th>Flow Type</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA/ 42</td>
<td>Interrupted</td>
<td>0.4060</td>
<td>0.0857</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>LA/ 1248</td>
<td>Interrupted</td>
<td>0.1744</td>
<td>0.0672</td>
<td>0.0094</td>
</tr>
<tr>
<td>CA/ 101</td>
<td>Uninterrupted</td>
<td>0.1297</td>
<td>0.0449</td>
<td>0.0038</td>
</tr>
</tbody>
</table>
CHAPTER 4. CONCLUSIONS

The objective of this research was to investigate the consistency of relationships among the historical crash data ratio, surrogate measures of safety (or jerk data ratio collected through GPS data loggers), and the presence of curvature on two state-level interrupted flow arterials in Baton Rouge, LA. The results then were compared to the results of the Pande et al. (2014) study completed using the same variables but on an uninterrupted traffic flow section of Highway 101 in San Louis Obispo, California.

GPS data was collected from 31 staff members and their household at Louisiana State University. The data was then filtered and linearly referenced to their nearest road in ESRI ArcMap. Based on the GPS frequency of data, LA42 and LA1248 as two interrupted Baton Rouge state highways were selected for further study. The crash rate and jerk ratio were then calculated to be used in a sensitivity analysis for finding an appropriate jerk threshold and segment length.

Twenty one jerk thresholds (started from -0.5 decreasing by 0.5 until reaching -10.5), three different segment lengths (eighth, quarter, and half), and crash ratio values per segment were considered in the Pearson correlation coefficient tests. The results indicated that the jerk threshold of -2.5 and quarter mile segments accomplishes the highest correlation coefficient with the crash rate.

The data used as input for crash frequency models explored the correlation between the percentage of -2.5 jerk ratios and the presence of horizontal curvature with the crash rate based on quarter mile segments.
The results of crash frequency models demonstrated a strong correlation between the location of jerk clusters and vehicle crash rates on two interrupted traffic flow sections. The models relating the crash rate and jerk ratio produced partial correlation coefficients that were different to zero, with p-values of <0.0001 and 0.0094 for LA42 and LA1248, respectively. A comparison between the results of LA42 and LA1248 showed a consistency in the jerk ratio level of significance among the two roads. Moreover, the results were also consistent with prior research by Pande et al. (2014) on uninterrupted traffic flow in San Luis Obispo, California, which showed a similar strong correlation that could be used. While more research is required to generalize the relationship to all road types and locations, these findings suggest that a clustering of high magnitude negative jerks may precede vehicle crashes and be an indicator of future crash locations.

Compared to traditional measures used to estimate crash frequency (ADT and the presence of geometric curves), jerk-clusters may allow analysis to identify crash prone locations in advance and with greater certainty compared to other methods. In other words, jerk-clusters can minimize the crashes, along with the losses and suffering that accompanies these crashes.

A secondary goal of this research was to examine the effects of segment length on the relationship between the jerk ratios and crash rate. Segments that were too short (less than or equal to 1/8 mile) or too large (greater than or equal to 1/2 mile) reduced the ability to correlate jerks and crashes. In practice, on large segments crashes and jerks might not distributed homogeneously that makes non-uniform segments, and short segments may not include sufficient crashes- as they are rare events- to yield the significant results. For these reasons, both short and long segments cause inaccuracy and error in the statistical and crash prediction models. Altogether, the results suggest there is an ideal segment length for analysis that yields the highest correlation.
The examination of adding another explanatory variable to the crash model demonstrated an improvement in the full log likelihood values for each road separately and also the overall model. To confirm that, two log likelihood ratio tests were done for each crash frequency model (one with jerk ratio as the only independent variable and the other with the presence of curvature in addition to the jerk ratio) separately. Briefly, the results indicated adding more crash-related variables to the crash prediction models could improve the results.

In general, the results of this research could have a profound impact on the way highway safety is quantified and capital investment on roadway projects is allocated in the not-too-distant future.

There are also few limitation that should be addressed in future study. The major limitation was a low frequency and subsequently low quality of the GPS data. In fact, evasive braking maneuvers are time-sensitive and they mostly occur in a small fraction of a second, availability of high frequency of data is necessary to captured all the maneuvers. Another limitation was the average daily traffic data points that did not exist per each segment, which led to developing an interpolation function in GIS to obtain an ADT value per segment. Although the interpolation could attribute an ADT value to each segment, it may not have assigned them their actual values. So, having actual average daily traffic per segment could offer more precise results.
FUTURE WORK

Information technology advances are opening up new opportunities for traffic agencies to be able to identify locations with relatively high jerk events. With the number of GPS enabled devices, such as “smartphones” and “tablets” increasing, future research may have the ability to access this information to analyze jerk events through “crowd sourcing.” This would result in identifying unsafe roadways significantly earlier because traffic engineers would not have to wait for sufficient crash data to accumulate. Identification of unsafe roadways prior to any crashes occurring would be a lifesaving measure.

Furthermore, the following barriers should be cleared (prior to other issues) to obtain more accurate results in the future studies:

1) The frequency of the data collection (data point per second) should be set as high as possible because, as mentioned, the evasive braking maneuvers are time sensitive and occur in a small fraction of a second.
2) An appropriate segment length should be further explored through a spatial analysis tool capable of identifying the optimal length for study, as it may lead to more accurate crash prediction models.
3) Detailed data about each curve, including sharpness and radius, should be included in roadway safety models.
REFERENCES


Dixit, V. V. P., Anurag; Abdel-Aty, Mohamed; Das, Abhishek; Radwan, Essam. (2011). Quality of traffic flow on urban arterial streets and its relationship with safety. *Accident Analysis & Prevention*, 1610-1616.


Guo, F. K., Sheila G.; McGill, Michael T.; Dingus, Thomas A. (2010). Evaluating the Relationship Between Near-Crashes and Crashes: Can Near-Crashes Serve as a Surrogate Safety Metric for Crashes? Retrieved from Washington DC: http://www.nhtsa.gov D O T / N H T S A / N V S / H u m a n % 2 0 F a c t o r s / S a f e t y % 2 0 P r o b l e m % 2 0 I d e n t i f i c a t i o n / D O T % 2 0 H S % 2 0 8 1 1 % 2 0 3 8 2 . p d f .


Regan, M. A. W., A; Grzebieta, R; Charlton, J; Charlton, J; Watson, B; Haworth, N; Rakotonirainy, A; Woolley, J; Anderson, R; Senserrick, T; Young, K. (2013). *The Australian 400-car Naturalistic Driving Study: Innovation in road safety research and policy* Paper presented at the Australasian Road Safety Research, Policing & Education Conference.


Wu, K.-F. A.-V., Jonathan; Jovanis, Paul P. (2014). Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level. Accident Analysis And Prevention, 210-218.


APPENDIX A. INSTITUTIONAL RESEARCH BOARD (IRB)

ACTION ON PROTOCOL APPROVAL REQUEST

TO: Brian Woishon
    Civil & Environmental Engineering

FROM: Robert C. Mathews
    Chair, Institutional Review Board

DATE: March 13, 2012

RE: IRB# 3263

TITLE: New Methods for Measuring, Evaluating and Predicting the Impact of Road Infrastructure Systems on Driver Behavior


Review type: Full ___ Expedited X ___ Review date: 3/14/2012

Risk Factor: Minimal X ___ Uncertain _____ Greater Than Minimal _____

Approved X ___ Disapproved ______

Approval Date: 3/14/2012 Approval Expiration Date: 3/13/2013

Re-review frequency: (annual unless otherwise stated)

Number of subjects approved: 20

Protocol Matches Scope of Work in Grant proposal: (if applicable) ___

By: Robert C. Mathews, Chairman ______

PRINCIPAL INVESTIGATOR: PLEASE READ THE FOLLOWING – Continuing approval is CONDITIONAL on:

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU's Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submittal of a termination report) prior to the approval expiration date, upon request by the IRB office. (Irrespective of when the project actually begins); notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.
5. Continuing attention to the physical and psychological well-being and informed consent of the individual participants, including notification of new information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
8. SPECIAL NOTE:

*All investigators and support staff have access to copies of the Belmont Report, LSU's Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at http://www.lsu.edu/irb

48
## APPENDIX B. GPS OUTPUT SAMPLE

<table>
<thead>
<tr>
<th>LATITUDE</th>
<th>LONGITUDE</th>
<th>ALTITUDE</th>
<th>HEADING</th>
<th>SPEED</th>
<th>FIX</th>
<th>YEAR</th>
<th>MONTH</th>
<th>DAY</th>
<th>HOUR</th>
<th>MIN</th>
<th>SEC</th>
<th>MSEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>30.37929</td>
<td>-91.2003</td>
<td>168.89</td>
<td>95.72</td>
<td>18.4</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>47</td>
<td>700</td>
</tr>
<tr>
<td>30.37929</td>
<td>-91.2002</td>
<td>168.8</td>
<td>93.02</td>
<td>17.38</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>48</td>
<td>100</td>
</tr>
<tr>
<td>30.37929</td>
<td>-91.2003</td>
<td>169.1</td>
<td>97.36</td>
<td>19.21</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>47</td>
<td>300</td>
</tr>
<tr>
<td>30.3793</td>
<td>-91.2003</td>
<td>169.1</td>
<td>98.33</td>
<td>19.83</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>46</td>
<td>900</td>
</tr>
<tr>
<td>30.3793</td>
<td>-91.2002</td>
<td>168.69</td>
<td>86.73</td>
<td>15.27</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>48</td>
<td>500</td>
</tr>
<tr>
<td>30.3793</td>
<td>-91.2004</td>
<td>169.1</td>
<td>98.87</td>
<td>20.23</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>46</td>
<td>500</td>
</tr>
<tr>
<td>30.37931</td>
<td>-91.2004</td>
<td>169.19</td>
<td>98.94</td>
<td>20.8</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>46</td>
<td>100</td>
</tr>
<tr>
<td>30.37931</td>
<td>-91.2002</td>
<td>168.6</td>
<td>79.91</td>
<td>13.92</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>48</td>
<td>900</td>
</tr>
<tr>
<td>30.37932</td>
<td>-91.2004</td>
<td>169.3</td>
<td>98.88</td>
<td>21.24</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>45</td>
<td>700</td>
</tr>
<tr>
<td>30.37932</td>
<td>-91.2002</td>
<td>168.6</td>
<td>74.16</td>
<td>13.69</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>49</td>
<td>300</td>
</tr>
<tr>
<td>30.37932</td>
<td>-91.2005</td>
<td>169.39</td>
<td>98.83</td>
<td>21.64</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>45</td>
<td>300</td>
</tr>
<tr>
<td>30.37933</td>
<td>-91.2002</td>
<td>168.5</td>
<td>68.41</td>
<td>13.54</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>49</td>
<td>700</td>
</tr>
<tr>
<td>30.37933</td>
<td>-91.2005</td>
<td>169.39</td>
<td>98.51</td>
<td>22.02</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>44</td>
<td>900</td>
</tr>
<tr>
<td>30.37934</td>
<td>-91.2005</td>
<td>169.5</td>
<td>98.05</td>
<td>22.11</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>44</td>
<td>500</td>
</tr>
<tr>
<td>30.37934</td>
<td>-91.2002</td>
<td>168.39</td>
<td>62.84</td>
<td>13.46</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>30.37935</td>
<td>-91.2006</td>
<td>169.6</td>
<td>97.87</td>
<td>22.01</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>44</td>
<td>100</td>
</tr>
<tr>
<td>30.37935</td>
<td>-91.2002</td>
<td>168.3</td>
<td>58.65</td>
<td>12.94</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>50</td>
<td>500</td>
</tr>
<tr>
<td>30.37937</td>
<td>-91.2007</td>
<td>169.69</td>
<td>96.44</td>
<td>20.29</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>43</td>
<td>200</td>
</tr>
<tr>
<td>30.37937</td>
<td>-91.2001</td>
<td>168.19</td>
<td>50.54</td>
<td>13.38</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>50</td>
<td>900</td>
</tr>
<tr>
<td>30.37938</td>
<td>-91.2008</td>
<td>169.8</td>
<td>95.66</td>
<td>19.66</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>42</td>
<td>800</td>
</tr>
<tr>
<td>30.37939</td>
<td>-91.2008</td>
<td>169.89</td>
<td>95.44</td>
<td>19.27</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>42</td>
<td>400</td>
</tr>
<tr>
<td>30.37939</td>
<td>-91.2009</td>
<td>169.89</td>
<td>95.66</td>
<td>18.88</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>30.37939</td>
<td>-91.2009</td>
<td>170</td>
<td>95.94</td>
<td>18.22</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>41</td>
<td>600</td>
</tr>
<tr>
<td>30.3794</td>
<td>-91.2009</td>
<td>170</td>
<td>96.15</td>
<td>17.53</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>41</td>
<td>200</td>
</tr>
<tr>
<td>30.3794</td>
<td>-91.201</td>
<td>170.1</td>
<td>96.29</td>
<td>16.93</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>40</td>
<td>800</td>
</tr>
<tr>
<td>30.37941</td>
<td>-91.201</td>
<td>170.3</td>
<td>96.48</td>
<td>15.75</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>39</td>
<td>900</td>
</tr>
<tr>
<td>30.37941</td>
<td>-91.201</td>
<td>170.19</td>
<td>96.33</td>
<td>16.36</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>40</td>
<td>400</td>
</tr>
<tr>
<td>30.3794</td>
<td>-91.2001</td>
<td>168.1</td>
<td>45.65</td>
<td>14.33</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>51</td>
<td>300</td>
</tr>
<tr>
<td>30.37941</td>
<td>-91.2011</td>
<td>170.3</td>
<td>96.06</td>
<td>15.37</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>39</td>
<td>500</td>
</tr>
<tr>
<td>30.37941</td>
<td>-91.2011</td>
<td>170.39</td>
<td>95.11</td>
<td>15.06</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>39</td>
<td>100</td>
</tr>
<tr>
<td>30.37941</td>
<td>-91.2012</td>
<td>170.69</td>
<td>94.73</td>
<td>16.14</td>
<td>1</td>
<td>2012</td>
<td>8</td>
<td>21</td>
<td>11</td>
<td>11</td>
<td>37</td>
<td>100</td>
</tr>
</tbody>
</table>
APPENDIX C. SAS CODE SAMPLE FOR NEGATIVE BINOMIAL REGRESSION

```sas
dm 'log;clear;output;clear';
ods html close; ods html style=minimal;
ods graphics on;
ods rtf;
options nodate nocenter pageno=1 ls=90 ps=56;
ODS listing;

data Neg_binomial;
   INFILE 'All_variables.csv' dlm=',' dsd missover firstobs=2;
      input Jerk_Ratio   ADT_IDW   Curve   crashes_all;
run;

proc genmod data = Neg_binomial;
   model crashes_all =Jerk_Ratio   ADT_IDW   Curve
      / type3 dist=negbin;
run;

ods rtf close;
```

VITA

Seyedeh Maryam Mousavi, originally from Iran, received her bachelor’s degree in Urban Planning at Shiraz University, Iran, in 2012. After her graduation, she decided to study abroad in the field of transportation engineering. She is a Master candidate in the Department of Civil and Environmental Engineering at Louisiana State University in 2015 and plans to pursue her PhD upon her graduation.