Using SHRP2-NDS Data to Investigate Freeway Operations, Human Factors, and Safety

Final Report

Report: ATLAS-2018-23

by

Raul Avelar, PhD, PE
Associate Research Engineer
Texas A&M Transportation Institute

Bahar Dadashova, PhD
Associate Transportation Researcher
Texas A&M Transportation Institute

Laura Higgins
Associate Research Scientist
Texas A&M Transportation Institute

and

George Gillette
Texas A&M Transportation Institute

ATLAS CENTER
Advancing Transportation Leadership and Safety

University of Michigan	**Texas A&M University**
2901 Baxter Rd	**Texas A&M Transportation Institute**
Ann Arbor, MI 48109-2150	**College Station, TX 77843-3135**

July 2018

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGMENTS

1. Report No. ATLAS-2018-23	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Using SHRP2-NDS Data to Investigate Freeway Operations, Human Factors, and Safety. Final Report.		5. Report Date July 2018	
		6. Performing Organization Code	
7. Author(s) Raul Avelar, Bahar Dadashova, Laura Higgins, and George Gillette		8. Performing Organization Report No.	
9. Performing Organization Name and Address Texas A&M Transportation Institute The Texas A&M University System College Station, TX 77843-3135		10. Work Unit no. (TRAIS)	
		11. Contract or Grant No. DTRT13-G-UTC54	
12. Sponsoring Agency Name and Address Advancing Transportation Leadership and Safety (ATLAS) Center 2901 Baxter Rd., Room 124, Ann Arbor, MI 48109-2150 USA		13. Type of Report and Period Covered	
		14. Sponsoring Agency Code	
15. Supplementary Notes Supported by a grant from the U.S. Department of Transportation, OST-R, University Transportation Centers Program			

16. Abstract

This project identified and quantified relationships between traffic operations variables (such as operating speeds), human factors characteristics (e.g., driver demographics), and safety variables (crash or near-crash outcome) making use of the recently available SHRP2 databases.
Researchers studied driver performance and investigated how the operational characteristics may explain safety outcomes.

Using nearly 800 events from freeway trips in Washington and Florida at non-curve and under uncongested conditions, researchers fitted one set of models to free-flowing driving situations to investigate variability in speed of choice (SOC). Driver characteristics such as years of driving experience, driver age, vision conditions, and history of traffic violations correlate with the SOC. However, facility type, traffic density, and especially the posted speed limit are most influential to free-flow driving. For another analysis, researchers calibrated a dynamic mixed-effects model to car-following situations and found that the estimated reaction time for a driver increases with increasing age (from 1.1 s for drivers younger than 20 years, up to 2.2 s for drivers older than 69 years). Compared to other age groups, younger drivers were found over-sensitive to relative speed and following gap as the magnitude of their speed adjustment due to these factors was found largest. Drivers 70 years of age or older performed speed adjustments similar to drivers ages 20–39, though these adjustments lagged more for the older group (i.e., reaction time of 2.2 s, compared to 1.2 s for ages 20–39). Additionally, a consistent trend in adjusting car-following speed toward the predicted free-flow SOC was found for all age groups except for younger drivers. Future research should focus on studying car-following behavior in situations that resulted in crashes or near crashes.

The second study compiled a set of speed data at freeway ramps in Pennsylvania. The results showed that the traffic conditions and the geographic location characteristics are the main predictors of a driver's speed choice on ramps. Based on the influence of these factors, drivers were found to adjust their speed more uniformly. Driver characteristics were found to influence the SOC to a certain degree. Drivers in the group of speeds with larger variability were found to be easily distracted based on their Barkley Attention Deficit Hyperactivity Disorder scores.

17. Key Words Freeway Operations, Driver Performance, Driver Age, Reaction Time, Speed of Choice, Car-following Behavior, Ramp Speed		18. Distribution Statement Unlimited	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 122	22. Price

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND INITIALISMS

ADF	Augmented Dickey-Fuller test
ADHD	Attention Deficit Hyperactivity Disorder
AIC	Akaike Information Criterion
ARIMA	Auto Regressive Integrated Moving Average
ARMA	Auto Regressive Moving Average
CHAID	Chi-squared Automatic Interaction Detector
CPM	Crash Proximity Measure
DTW	Dynamic Time Warping
DWT	Discrete Wavelet Transformation
FS	Freeway to street
GPS	Global Positioning System
HF	Human Factors
LOS	Level of Service
NDS	Naturalistic Driving Study
NN	Neural Network
OP	Operations
PSL	Posted Speed Limit
RID	Roadway Information Database
RPS	Risk perception score
SA	Safety
SF	Street to freeway
SHRP2	Second Strategic Highway Research Project
SOC	Speed of Choice
TTC	Time to Collision

CHAPTER 1: INTRODUCTION

Traditionally, transportation research has investigated the operations (OP), human factors (HF), and safety (SA) aspects of transportation issues in relative isolation. Even though it is possible to study HF and OP together, the use of surrogate measures to translate those findings to the SA realm is often unavoidable because of scarcity of crash data, compared to the wealth of OP and HF data. Clearly, there is a need to better understand how and to what extent are variables studied in OP and HF conducive to the observed crash outcomes, typically studied by SA after the fact. Questions about risky behaviors and other OP and SA factors associated with crash propensity can be answered using naturalistic data, such as the Second Strategic Highway Research Project (SHRP2) databases. However, in the face of the ever-changing traffic OP conditions and demographics, it is important to investigate how OP, HF, and RS interact, and how those interactions are likely to affect the variables of interest for traffic engineers, behavioral scientists, and SA researchers. To articulate the interrelations among OP, HF, and SA, one needs to bridge the multiple temporal scales involved in those interrelations. On one hand, HF and OP phenomena occur in time resolutions of the minutes or seconds, but SA studies crashes observed over periods of months or years.

Project Objective

This project identified and quantified relationships among traffic OP variables (such as operating speeds) and HF characteristics (e.g., driver demographics). The intent was to shade light on these relations over SA outcomes (crash or near-crash) making use of the recently available SHRP2 databases. Figure 1 illustrates the concept of this study's objective.
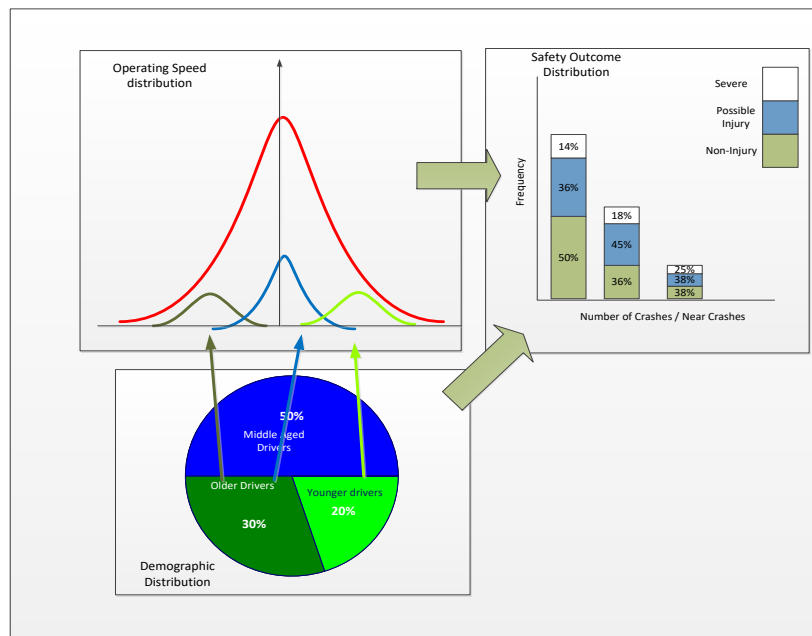
Figure 1. Interrelations between Demographic, OP, and SA Outcome Distributions.

Researchers relied on recently available SHRP2 products. The two main database components of the SHRP2 products are the Naturalistic Driving Study (NDS), which contain driving details from volunteers

on seven states, and the Roadway Information Database (RID), which is a rich data set from those seven states containing multiple years of roadway features, yearly traffic, crashes, and geolocation of such features. Combined, these databases encompass crashes and near-crashes documented in detail, as well as multiple peripheral data that can be directly linked to specific locations, driver, and vehicle characteristics. Researchers assembled the data sets of interest by combining and querying these databases.

Structure of the Report

This report consists of eight chapters. Chapter 1 introduces this research and the structure of the report. Chapter 2 provides basic background and describes the work plan. Chapter 3 summarizes the literature review performed prior to the data reduction. Chapter 4 summarizes the data characteristics. Chapter 5 provides details about the methodology implemented in the analyses. Chapter 6 describes and document the analyses performed for the study on driver speed of choice (SOC) in freeways (Study 1). Chapter 7 documents the analyses performed to study driver speed on freeway ramps (Study 2). Chapter 8 summarizes the conclusions and outlines future directions from this work. This report includes an appendix with a research paper derived from Chapter 7 that was submitted for publication (and accepted) at the *Transportation Research Record. The Journal of the Transportation Research Board*.

CHAPTER 2: NDS BACKGROUND AND WORK PLAN

This chapter summarizes the general background to this research and outlines the work plan prepared to accomplish the research goal. The next subsection summarizes the features of naturalistic data (i.e., data from real-world drivers who have agreed to get their vehicles instrumented for research purposes) since that is precisely the type of data set used in this research.

Naturalistic Driving Studies

Naturalistic driving studies pertain to a research approach that attempts to understand typical driver behaviors through inconspicuous sensor data collection [1]. Instrumented vehicles record microscopic data for later download and analysis. These microscopic data range from survey data and psychological profiles to raw sensor data from global positioning system (GPS) loggers, accelerometers, and infrared devices. Ideally, drivers would be representative of various locations to capture and represent geographical and geometric diversity. However, the cost of such wide representation is significant and often means that a compromise should be achieved between the wide representativeness and scope. Although NDS may provide a better understanding of how and why crashes happen, the data collected can be difficult to prepare for analysis and results hard to interpret. A published paper offering prospective on these types of studies recognizes three main methodological groups: a) studies on surrogates for events of interest (e.g., crashes) and the hierarchical relationship between those events and their surrogates; b) studies that pursue the interpretation of the driving context and its influence on drivers; and c) studies that assess the risk of events of interest (e.g., crashes) and the representativeness of such risk estimates (i.e., based on features of the sampling design of data) [2].

Prior NDS Studies

There are two landmark NDS efforts that demonstrate the progression of the field in terms of scope, approach, and depth of research questions: a) the 100 Car NDS, and b) SHRP2 NDS.

The 100 Car NDS was the first large-scale instrumented vehicle study conducted for the express purpose of collecting naturalistic driving data [3, 4]. The study collected approximately 2,000,000 vehicle miles, 43,000 hours of data, with 241 primary and secondary drivers spanning a 34-month period starting in August 2001. Upon completion of the data collection effort, there were 10 primary research objectives for the project:

1) Characterize crashes, near-crashes, and incidents.
2) Quantify near-crash events by operational characteristics.
3) Characterize driver inattention.
4) Qualify driver behavior and how it changes over time.
5) Understand the causal factors and dynamic conditions to rear-end conflicts.
6) Understand the causal factors and dynamic conditions for lane-change conflicts.
7) Understand the role of inattention for rear-end lead vehicle scenarios.
8) Characterize the rear end scenarios in relation to Heinrich's Triangle.
9) Evaluate performance of hardware, sensors, and data collection system.
10) Evaluate performance of data reduction plan, triggering methods, and data analysis.

In their attempts to analyze the data collected, researchers found difficulty in addressing question 2. Noise in the data and sensor issues limited their ability to relate near-crash events to operational characteristics. For example, in a lane-switching scenario, the radar that provides the distance to the lead vehicle would not collect data about the lead vehicle until the lane change maneuver is complete. This situation then results in incomplete information in the data set. Additionally, researchers concluded that the kinematic signatures associated with near-crash events are virtually identical to many common driving situations that are not indicative of crash risk.

The second major NDS conducted was the SHRP2 NDS [5]. To increase the volume of data collected, for the 100-car study, SHRP2 recruited significantly more drivers (nearly 2,360 participants either having completed their participation or on the road as of September 2012). This study was organized across six institutions: Indiana University, Pennsylvania State University, University of South Florida, Westat, CUBRC, and Battelle. The overarching interest for this research study was to study how driver behavior is affected by driver characteristics, vehicle, roadway, and environmental factors, and how changes in those behaviors are related to crash risks under various conditions.

Research Plan

To address the proposed research questions, researchers obtained NDS data from the SHRP2 NDS databases with the intent to investigate the relationships of interest to this research project. Two different studies were devised as a result.

Study 1: SOC, Car-following Behavior, and Crash Risk on Highway and Freeway Operations

Researchers prepared queries in the Insight website [6] to assess how many events of potential interest are available. The Insight website allows its users to examine the data preliminarily, before they can request specific data sets for purchase. The queries researchers prepared contained a set of approximately 900 events of interest. Among all the events, there were three distinct kinds: baseline events (i.e., random snippets of normal driving behavior), near-crash events (i.e., driving behavior immediately before, during, and immediately after a near-crash situation), and crashes (i.e., driving behavior immediately before, during, and immediately after a situation that resulted in a crash). Primarily, the queries were designed to include operational speed, radar gap, and headway with resolution of 10 ms. Each event consists of up to 30 seconds of data, but several of these events were limited to partial availability for either speed, gap, or both types of data. In addition, the queries were prepared to include facility type, SA outcome, level of service (LOS), light condition, driver gender, driver age, and driver risk-taking questionnaire answers, among other variables. A total of 105 variables were requested for potential analysis. After applying some filters, a final set of 847 events was requested from Virginia Tech for analysis.

The following are the steps that researchers proposed for a comprehensive analysis of these driving events:

1. Extract speed, gap, and headway data from all 847 events requested.
2. Develop a data set merging the data from these events and the corresponding RID segment characteristics.

3. Develop models for baseline speed that differentiate between car-following and free-flowing situations (considering gap and headway, when appropriate) for highway and freeway cross-sections.
4. Use the free-flowing model to obtain insights on driver SOC, potentially considering humans factors characteristics (gender, age, risk-taking questionnaire answers, visual acuity, etc.). The number of factors that can be determined influential in such baseline model would depend on how many baseline events are found that represent free-flowing conditions.
5. Considering the results from the SOC analysis, perform an additional analysis on how a car-following situation (i.e., subject vehicle following a vehicle in front) seems to affect speed, gap, and headway of the study participants.
6. Study the relation of near-crash and crash events with operational models developed in the prior steps.
7. Perform an investigation of how the marginal and conditional distributions of OP, HF, and SA variables relate to each other.

Study 2: SOC on Freeway Ramps

The second project concerns the speed choice on freeway ramps. Researchers identified four interchanges (16 ramps) of interest using SHRP2 RID and trip density maps from the Insight website [*6*]. After identifying the freeway ramps of interest, researchers requested the NDS trip and driver characteristics data. To inform the request of driver characteristics data, researchers conducted a comprehensive literature review on the driver characteristics associated with the speed choice and speeding behavior. As the result, researchers obtained over 2,000 trip data taken by over 60 drivers. The trip distances were approximately 2 miles, while the time of the trip varied based on the driver. As for Study 1, the trips were measured in the timeframe of seconds.

For the second study, researchers prepared the following data analysis plan:

1. Extract the GPS and network speed data from the requested trips.
2. Reduce the trip time series data by identifying those trips that were recorded poorly and had inconsistencies.
3. Apply data transformations to aggregate the trip time series data into a more manageable set. As indicated, the trips were available on a secondly basis. However, due to congestion or driver's SOC, the lengths of some of the trips could potentially extend up to 15 minutes. A data transformation may be warranted to match the relatively longer time series with the shorter ones to be jointly analyzed.
4. Match and cluster the trip time series data with similar patterns, into a reasonable number of clusters to capture different levels of risk driving types.
5. Conduct the data analysis using machine learning tools to identify the driver characteristics, and ramp geometric characteristics associated with the three speed choice states (clusters).

Potential Areas of Impact from These Studies

The following lists potential areas of impact from these studies:

- The influence of HF as it relates to OP and SA may yield findings that are useful to road design and highway capacity estimation practice. Previous research has indicated that older drivers

have higher reaction times and tend to maintain a larger gap between themselves and the vehicle ahead of them. With the aging of the general population and increased numbers of older drivers, does this increased vehicle gap play a role in road capacity? With more vehicles occupied by older drivers, does the increased gap between older driver and lead vehicles allow for less overall vehicles on the roadway? An increased share of older drivers could have implications in road design on their increased reaction times that this research may help quantifying.

- Improved speed, gap, and headway models are potentially useful to develop updated microsimulation models for the various conditions represented in the data set of 847 events. More realistic microsimulation models are likely to improve transportation professionals' ability to perform operational assessments of impacts of management strategies or even the effects of implemented policy decisions.
- Knowing the list of the factors that can affect the speed choice on the ramps can help the highway engineers to develop better designs to accommodate different types of drivers.
- Realistic, validated crash and near-crash risk functions offer the opportunity to develop estimates of crashes and near-crashes based on operational simulations.
- The potential quantitative formulation of a mathematical form for the interrelationship of the marginal and conditional distributions of OP, HF, and SA variables has the potential to advance current knowledge about how these domains affect each other. Currently, there is working knowledge, mostly qualitative, of these mutual effects, but there is little literature on qualitative estimations.

CHAPTER 3: LITERATURE REVIEW

This chapter summarizes literature relevant to this research's questions of interest. Researchers performed a thorough search of literature sources within the last 10 years on the following topics:

- Distracted driving based on SHRP2 NDS data published by 2016 or earlier.
- SHRP2 studies on younger drivers, and older drivers from the same timeframe.
- Other studies on distracted driving, younger and older drivers; studies focusing on driver reaction times.
- Studies focusing on driver car-following behavior.
- Studies focusing on distance to collision.

The following subsections organize the literature found in coherent topics to inform the subsequent data reduction, analysis, and interpretation of results.

Speed Choice and Freeway Safety

Crash report analysis conducted from 2003 to 2007 by the American Automobile Association Foundation for Traffic Safety revealed that aggressive driving was reported in 56 percent of fatal crashes [7], with the number one factor being excessive speed. Aggressive driving behavior may ultimately create a risky situation not only for drivers themselves but also for other road users. In 2012, there were more than 33,000 fatalities and 2.2 million injuries in the United States, with driver behavior being a significant factor in more than 90 percent of these crashes [8].

The impact of a driver's speed choice has been noted across multiple studies to associate with SA [9]. The SA association of speed is evaluated as both the absolute speed and speed dispersion (the measure of difference between an individual driver's speed choice and the prevailing traffic conditions). There are many factors that can impact the SA of a driver beyond their speed, such as road design and surface condition. These factors include situational, demographic, and personality characteristics and could be potentially divided into two main categories:

- Personal characteristics of driver: age, gender, sensation seeking, risk perception, etc.
- External factors: other road users, road type, passengers, congestion, speed limit, etc.
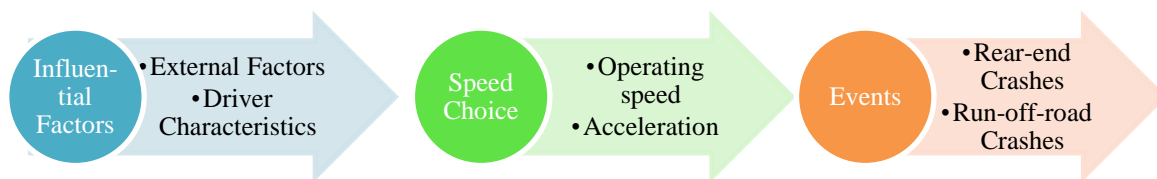
Figure 2. Speed and its SA Implications.

From the operational standpoint, the relationship between a driver's speed choice and their crash risk is well understood and is sensible [10]: as the speed increases, drivers have less time to react or stop in response to changing roadway conditions and thus are at a higher risk of collision. However, a definitive

answer has not been achieved from SA research that focusing on crashes only. Speed limit is not always found to have a marginal association with crashes in the expected direction [9, 11, 12].

Freeway crashes tend to occur more frequently at on-ramp and off-ramp locations. These crashes account for 18 percent of all interstate crashes, 17 percent of injury crashes, and 11 percent of fatal crashes at interchange locations [13]. A 2004 study by McCartt et al. found that about 50 percent of ramp-related crashes occurred while the vehicles were exiting the freeway, 36 percent occurred while the vehicles were entering the freeway, and 16 percent occurred at the midpoints of access roads [14]. They also observed that 48 percent of the crashes were run-off-road crashes, and 36 percent of them were rear-end collision. These findings suggest that speed adjustments that occur at interchange locations, such as freeway-to-ramp transitions, may be associated with an increase in crashes. Kim et al. found that 85 percent of all freeway rear-end crashes occurred within 2000 ft of the on-ramp gore [15]. This study found that there was a strong association between rear-end crash rates and deceleration rate. Overall, speed indicators and the acceleration rates have been found to be important predictors of highway SA [16, 17, 18, 19, 20].

Surrogate Measures of Safety

Surrogates are objective measures that are sought in research to relate crashes, near-crashes, and pre-event factors, ranging from the driver's personal tendency toward unsafe driving behaviors to environmental factors such as rainfall or other inclement weather conditions. Various single metric surrogate measures have been proposed in the past, including but not limited to: time to collision (TTC), deceleration rate, lateral position, and jerk [21, 22, 23, 24, 25].

Wu et al. concluded that good SA surrogates must have five general characteristics: a) the surrogate should have a short period of data collection and be more frequent than accidents, b) the surrogate should be correlated with a clinically meaningful outcome, c) the surrogate should have a statistical and causal relationship to crashes, d) the surrogate should fully capture the effect of a treatment in a way similar to how the treatment would affect crashes, and e) the surrogate should function as a marker correlated to a crash with a time scale underpinning [26].

Post-encroachment Time
Songchitruksa and Tarko propose a method that leverages a traffic event's operational characteristics with an understanding of extreme value theory to indicate the expected number of crashes on a roadway segment [27]. The process involves five steps: 1) decide crash type of interest for analysis, 2) determine crash proximity measure (CPM), 3) measure CPM at multiple other sites, 4) generate relationship between CPM and historical data of crashes, and 5) apply CPM relationship to measured data at site of interest. For this study specifically, researchers decided to apply the method to right-angle crashes at intersections. Accordingly, the CPM applied was post-encroachment time. This value refers to the amount of time that vehicles spend occupying a given space within the intersection. If the value is large, then there is safe passage for both vehicles; if the value is zero or slightly negative, then a crash has occurred. The extreme value theory approach mathematically transforms the extremes observed in a short observation period to attempt to reflect the events over the course of an extended timeframe. To observe whether this approach could be relied upon, researchers conducted a Monte Carlo simulation using traffic and crash data from 12 sites. This approach is not limited to just right-angle crashes and post-encroachment time CPM, and it can be generalized to any type of crash and any CPM

with statistical adjustments to risk estimation, if it adheres to the following general guidelines on the CPM:

- The CPM must be defined relative to the type of crash for the analysis. For example, post encroachment time may not be a relevant CPM for other crashes.
- The CPM must be observable and continuous, such that it adequately represents both crash-free OP and collisions.
- A definitive boundary in the CPM must exist between the crash and non-crash conditions.

After observing how the extreme value theory can be applied to approximate the number of crashes, the next point of interest is whether there can be a definitive relationship between these operational characteristics and SA and whether the likelihood of a crash or near-crash event can be determined from these flow characteristics.
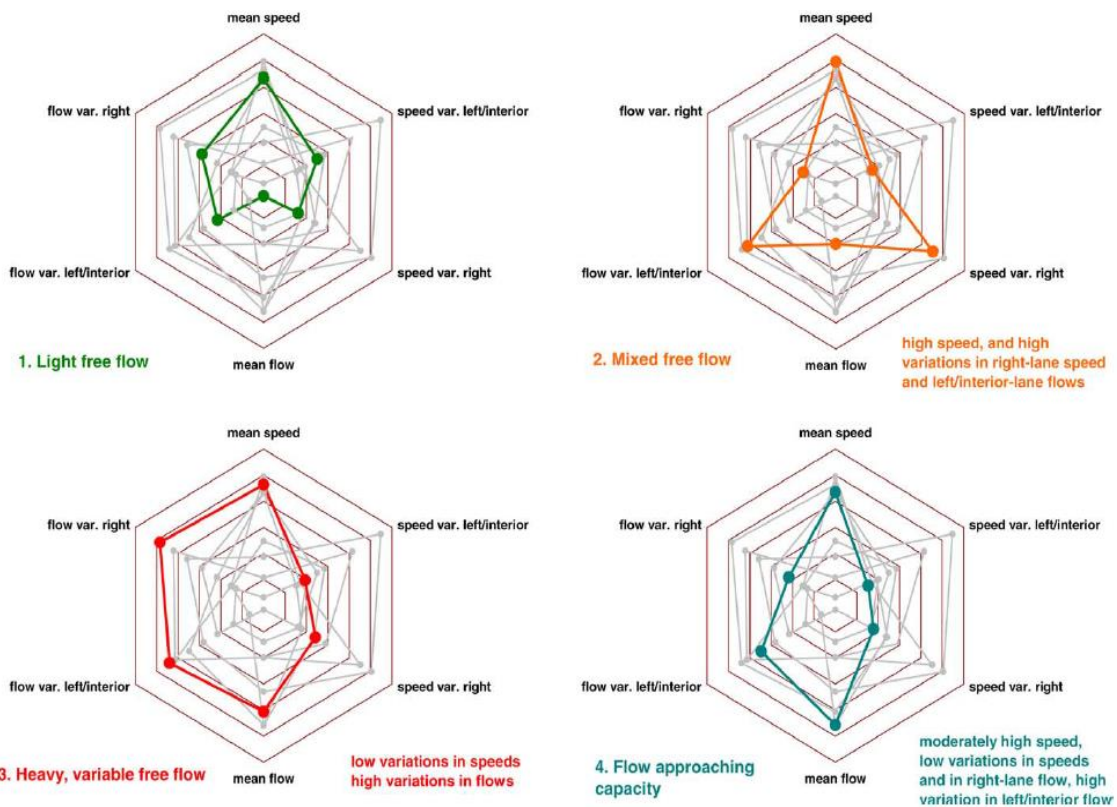
Figure 3. Eight Traffic Flow Regimes [*28*].

Traffic Regime and Crash Risk

Golob et al. studied traffic flow conditions through inductive loop detectors and analyzed them for the number of crashes [*28*]. To calibrate the analysis approach, this data set included over 1000 crash events over six major freeways in Orange County, California, in 1998.

Researchers segregated the data set by crash type, crash location, and severity. Researchers used traffic flow data collected in 30-second intervals of the count, and occupancy, and then grouped the

operational data into four blocks of general information: central tendency of speed, variation in speed, central tendency of volume, and variation in volume. Researchers then performed a cluster analysis to find groups of traffic flow conditions. These eight regimes were as follows: light free flow, mixed free flow, heavy variable free flow, flow approaching capacity, heavy flow at moderate speed, variable speed in congested flow, variable volume in congested flow, and heavily congested flow (see Figure 3).
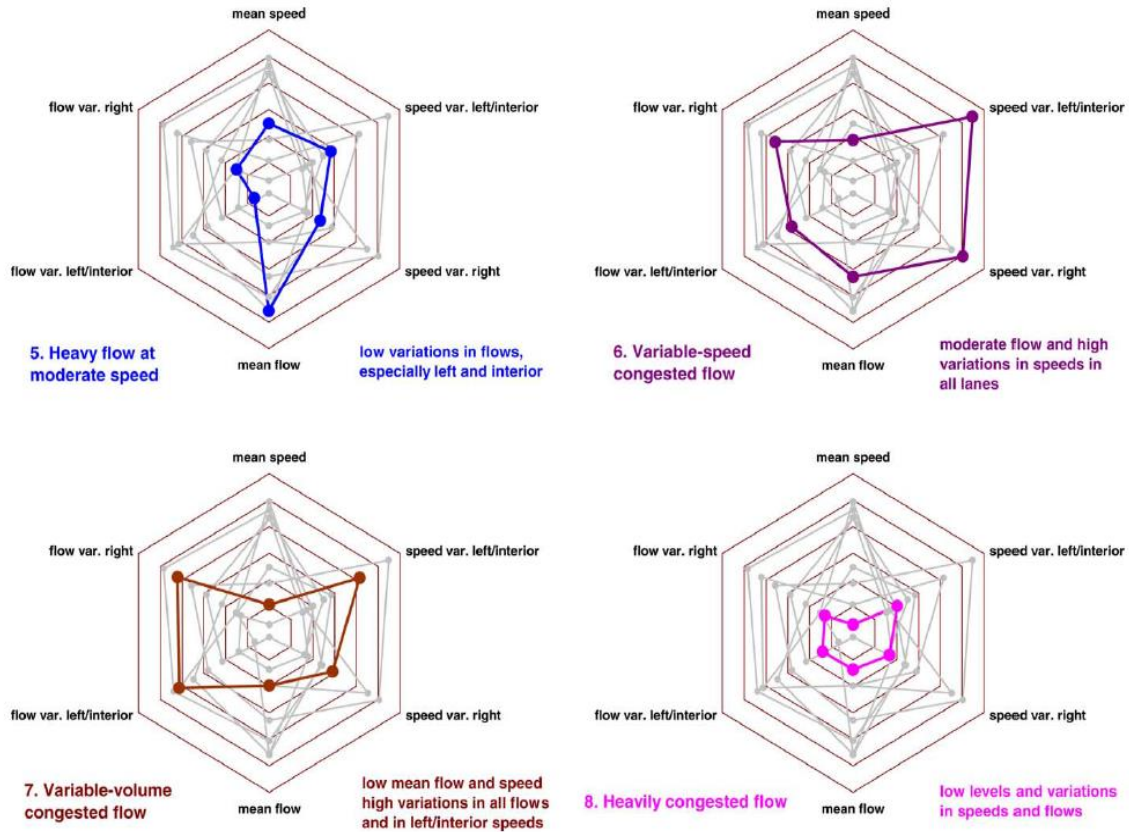
Figure 3. Pight Traffic Flow Regimes (continued) [28].

For these regimes, different types of crashes were found to be more prevalent in different regimes. The percentage of prevailing crashes by type is shown in Figure 4. For example, if the flow is characterized by a heavily congested regime (eight cluster in Figure 3) or as heavy with variable free flow (third cluster in Figure 3), the prevailing percentages of rear-end crash type are 83.3 percent and 78.8 percent, respectively. Golob et al. extended their research to develop a tool called Flow Impact on Traffic Safety [29].
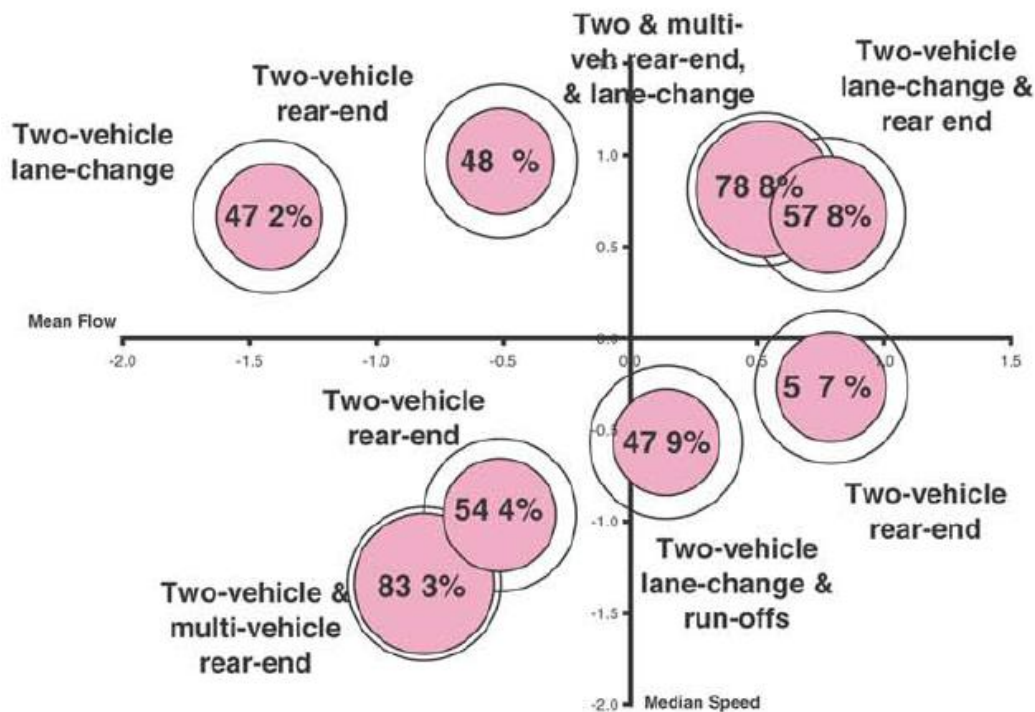
Figure 4. Prevalence of Types of Crashes with Traffic Regimes [*28*].

Brake Operation

Cheng et al. evaluates the driver brake operation in near-crash events through a forward-facing camera equipped to 50 taxis in Beijing [*30*]. Data were collected over the course of one year, amounting to 4.5 million kilometers traveled in the study. This data set includes 50 professional drivers with the following recorded characteristics: gender (48 male, 2 female), age (43 ± 5), and driving experience (13 ± 4). The vehicles were instrumented with equipment that would be automatically triggered by the following conditions: the Video Drive Recorder would record if the longitudinal acceleration reached .4G within 0.5 seconds, or if the instantaneous acceleration reaches 2G at any point. Upon analysis of the data, researchers found that the number of near-crash events to be approximately 60 times as large as actual crash events. Researchers indicated that there could be an underlying relationship that could be potentially mined to characterize the relationship between near-crash and crash events. Drivers were then categorized as conflict-prone or conflict-infrequent based upon the number of near-crash events they were involved with. In a further analysis of the same data set, researchers modeled the relationship between the time headway of an individual driver and their distribution of rear-end conflicts. The time-headway for an individual driver was shown to vary depending on which type of event they were involved.

Near Crashes

One surrogate measure that has been considered in the past is near-crashes. Guo et al. conducts three separate analysis: sequential factor analysis, frequency relationship of behaviors between crash and near-crash events, and a sensitivity analysis [*31*]. This is done to assess whether a sufficiently similar

causal mechanism exists in both near-crash and crash events, and to also evaluate the assumption of a constant crash-to-near-crash ratio. Researchers found a positive relationship between the number of crashes and near-crashes, which varies by situation. In general, the authors advocate for the combined analysis of crashes and near-crashes to improve accuracy of estimates.

Car-following Speed

Early studies of driver behaviors on the road began to define the concepts of TTC, safe stopping distances, and vehicle headways in terms of driver perceptions and reactions. Gibson and Crooks [*32*] postulated that drivers have a general sense of the minimum distance that would be required to stop their vehicle, depending on the vehicle's speed and the roadway conditions, and will be inclined to slow or steer the vehicle, gradually or suddenly, if they perceive an object encroaching within the minimum safe distance they have estimated for themselves. Gibson and Crooks also recognized that driver inattention could interfere with detection of approaching objects in the driving path and result in a vehicle headway that is shorter than the real or perceived minimum stopping distance.

In the late 1950s and early 1960s, several different mathematical models were developed to attempt to describe and predict a complete picture of driver following distances, speeds, and acceleration and deceleration behaviors as responses to the roadway environment and surrounding vehicles [*33*]. These models were generally developed using controlled experiments with instrumented vehicles on test tracks or open roads and may have been limited in part by their focus on optimum decision-making by drivers under controlled test conditions, rather than driver behaviors in more natural driving circumstances.

Driver Reaction Times

Similar to studies used for early car-following models, many driver reaction time studies have necessarily been conducted using volunteer drivers in simulated or test-track conditions. More recently, naturalistic driving studies have provided the opportunity to measure drivers' reactions to real-world roadway environments and events without the constraints of any testing scenario. Many simulator and test track studies have explored the effects of cell phone use on driver performance, including reaction times. Two meta-analyses, one of 33 studies in which drivers talked on hands-free or handheld cell phones while driving and one of 28 texting-and-driving studies, found that phone-related distractions tended to increase drivers' reaction times to events [*34, 35*]. In 2011, a test-track texting and driving study [*36*] estimated a doubling of reaction time for drivers writing or reading text on a cell phone, compared to undistracted drivers.

Age can affect the speed at which drivers process information, which in turn can degrade abilities to perceive hazards, shift attention, and manage complex driving tasks. As a result, older drivers tend to require more time to scan for visual cues when driving, tend to pay more attention to static cues than to dynamic ones, and can be slower to respond to changes in the driving environment, particularly to unexpected events [*37, 38, 39*]. An examination of reaction times of 351 drivers found that reaction times "increased progressively between the ages of 20 and 80" [*40*].

Reaction times depend partly on physical reflexes and motor skills, but also depend on how quickly a driver can sort through the possible responses to a stimulus, such as braking, steering, or both to avoid an object. When the number of possible responses is low, differences in reaction times between

younger and older adults are relatively small; as the number of possible responses to a stimulus increases, older adults are at an increasing disadvantage [*39, 41*].

Car-following Distances and TTC

TTC is dependent on the distance between any two vehicles (or a vehicle and any other object) and on their relative velocities and trajectories. Drivers influence TTC between their own vehicle and a lead vehicle by controlling their own speed and following distance. Detecting that a lead vehicle is slowing or stopping is crucial to avoiding a rear-end collision. It is not surprise that past research have use TTC as a variable to study such situations.

Several studies have shown that secondary-task distraction degrade a driver's perception and assessment of roadway hazards, even when the distraction is not visual. Studies examining lane change behaviors [*42*] and gap acceptance [*43*] found that drivers who were talking or listening to verbal messages were more likely to miss or misjudge potential hazards. The effects of distraction on steady-state following distance (as opposed to TTC immediately before or after a lane change or other maneuver) are less predictable. Several simulator and test-track studies have observed that drivers' following distances were similar or slightly longer when they were using cell phones, compared to when they drove undistracted [*34, 42, 44, 45*]. Other studies have found that drivers engaged in a secondary task are slower to adjust their following distances when conditions change, such as when a lead vehicle slows [*46*].

Elderly drivers in an instrumented vehicle study drove more slowly and with less speed variability when they were distracted versus when they drove undistracted; in addition, both elderly and middle-aged drivers in the study exhibited reduced steering when distracted [*47*]. A simulator study of drivers' lane-keeping behaviors found that drivers were better at maintaining lane position when distracted compared to when they were not, but only when the roadway environment remained predictable. When unpredictable elements were added to the experience (the study used simulated wind gusts), undistracted drivers were more successful at maintaining their lane position [*48*]. This result suggests that drivers distracted by secondary tasks become less attentive to steady-state driving tasks such as lane position and speed since these do not require as much moment-to-moment mental effort (if conditions do not change).

There is evidence that drivers begin to change their braking behavior as they get older, resulting in longer TTCs. An examination of data from the 100-Car NDS found that drivers aged 30 and older begin to brake an average of 1.7 seconds earlier than drivers under 30, in response to a decelerating lead vehicle [*49*].

Human Factors

The following section discusses which factors affect the operational characteristics and SA of drivers and how researchers are currently quantifying these relationships.

Personality Traits and Driver Experience

In terms of personality traits, Heino et al. discusses the concept that there are two broad groups of people: sensation avoiders and sensation seekers [*50*]. This is analogous to the conflict-prone and conflict-infrequent distinction in near-crash evaluation; however, the key difference is that sensation

avoiders and sensation seekers categories define the underlying behaviors, whereas conflict-prone and conflict-infrequent designations describe the frequency of those occurrences. The definition of a study participant as sensation avoider or seeker was determined by the completion of a psychological test, known as the Sensation Seeking Scale [51]. As expected, the study found that sensation avoiding individuals tended to prefer longer following distances. However, the physiological (heart rate variability) and cognitive (verbal risk ratings) measures of the two groups of sensation seeking and sensation avoiders during prescribed car-following tasks were not shown to vary. This indicates that the perceived level of risk and amount of cognitive effort is roughly equivalent for the two groups of drivers, despite different following distances; this suggests that the headway is chosen to have an optimum balance of personal risk relative to personal driving characteristics. Accordingly, it is not that the drivers are intentionally pursuing higher risk behaviors; instead, they do not see the behaviors as necessarily risky.

Heino et al. reveals through a Poisson regression analysis that the personality trait of sensation-seeking or sensation-avoidance influences the relationship between driver experience and crash involvement [50]. The generally assumed trend is that a more experienced driver is less likely to be engaged in a crash. When broken down by their sensation-seeking or avoiding personality trait, among drivers who have driven less than 80,000 km, sensation avoiders are less likely to be involved in a crash. However, among drivers who have driven more than 80,000 km, sensation seekers are less likely to be involved in a crash. A possible explanation is that sensation-seeking individuals have improved their measured braking in response to situational changes over time, due to their closer car-following habits. In a controlled driving simulator study, Winsum et al. focuses on evaluating the choice of time headway in car-following vehicles and determining whether the choice of headway is related to the braking behaviors [52]. The researchers found evidence that short followers (analogous to sensation-seeking individuals) tended to have a finer tune control of braking in critical situations. This thereby supports the theory that sensation-seekers, given enough driving experience, can better control their braking response and thus reduce their likelihood of a crash relative to the sensation-avoiding individuals.

Quimby et al. investigated how personality characteristics and personal behaviors affect speed through a combination of on-road observations and survey techniques [53]. To ensure that the speed was the chosen speed, free-flow speed was sampled unobtrusively from the side of the road with a video recording of the vehicles. The video recording allowed the researchers to find the license plates for the vehicles and send out surveys. These surveys included eight psychological self-assessments and allowed the researchers to evaluate the impact of certain personality characteristics on driver speed choice. The eight psychological assessments applied were: decision making style, mild social deviance, willingness to commit driving violations, sensation seeking propensity, intolerance, driving stress, hazard involvement, and general driving style. To compare the spot speeds of the drivers in the study to other drivers, the concept of relative speed is applied; the speed of the individual driver is evaluated relative to the speed of the other drivers across that segment also observed in that time segment. When these variables were applied independently in a regression equation to predict speed, the violation scale was shown to have the largest impact as an 8 percent effect; the model indicates that the more that drivers report themselves as engaging in violations, the more likely that their relative speed will be higher. Mild social deviance is also indicated to function as a positive predictor of speed; however, when age and other explanatory variables are added into the model, the factor is no longer statistically significant. Many of the psychological tests are strongly interrelated and measure many of the same driver behaviors from

different perspectives. Of the other psychological tests, the stress test was indicated to have a negative impact on speed. This implies that drivers who find driving to be a stressful task tend to drive more slowly. When the drivers were grouped into five relative speed categories from highest speed to lowest speed, the violation, sensation seeking, and stress psychological self-assessment tests proved to be the most robust explanatory variables to determine the placement of a driver in a given speed group.

Age and Gender

In addition to their personality traits, a driver's age and gender play a role in their overall involvement in crashes, and their choice of speed and time headway. The aggregate analysis of crashes and the role that age and gender play in those occurrences suggests that older drivers have the highest fatal crash involvement rate and that younger groups of drivers have the highest rate of involvement in all crashes [54]. However, a more recent study found that the risk of crash involvement for older drivers tend to decrease at times of the day when congestion is high, months of the year when winter conditions are prevalent, and in proximity to freeway ramps when the crash occurred over a highway connected to the freeway [55]. These changes in risk seem to suggest that older drivers may be offsetting their increased risk of crashing by self-regulating to avoid challenging environments to drive. Additionally, the results by Massie et al. point to an elevated risk of fatal crashes for men as compared to women, whereas women were shown to have a higher rate of involvement in injury crashes [54]. Montgomery et al. investigated how the brake response of the drivers varied for various scenarios; the analysis used brake applications from the 100-Car NDS and conducted a mixed model analysis to examine the differences by age and gender groupings [49]. The braking response was evaluated by the TTC at the moment of the response. Table 1 summarizes the results on age and gender of participants.

Table 1. Impacts of Age and Gender on TTC at Braking.

Gender	Age < 30	Age > 30
Male	3.2 ± 0.4 s	4.1 ± 0.4 s
Female	3.8 ± 0.5 s	5.7 ± 0.6 s
Notes: n_{male} = 52, n_{female} = 32,		
p < .001 for the differences in TTC at the time of braking for both age and gender		

Additionally, the TTC at the braking response was evaluated in four age categories: novice (18–20), young (21–30), middle (31–50), and mature (51+). The TTC variable is shown to increase with age. However, the increase is shown to not occur linearly with age. The TTC is shown to increase slightly between novice (18–20) and young (21–30). Between the young (21–30) and the middle (31–50) categories, there is a much more noticeable increase of almost a full two seconds. This is followed by a slight increase between middle (31–50) and mature (51+).

There are several potential explanations for the well-documented trend of increasing TTC with age. Charlton et al. propose the compelling argument that they may be compensating for their perceived age-related decline [56].

Quimby et al. also investigated the impact of how age and gender relate to speed using a combined data set assembled from observational and a survey [53]. Researchers applied five general age groupings to analyze the impact of age and gender on relative speed choice: 17–29 years, 30–39 years, 40–49 years, 50–59 years, and 60+ years. Researchers found that males in the age group 17–29 had the highest mean

speed of any age and gender pairing in the observed speed sample at 44.9 mph. Table 2 describes the mean speed by age group and gender, including 5064 total drivers. While females show a lower mean speed than males at every age category, the same pattern of decreasing speed in increasing age categories was found to apply to both genders.

Table 2. Impacts of Age and Gender on Mean Speeds.

Age Group	Males	Females	Both	# of Cases
17–29 years	44.9	43.7	44.2	744
30–39 years	44.3	42.7	43.4	1185
40–49 years	42.4	41.0	41.6	1194
50–59 years	42.5	41.1	41.9	871
60+ years	39.4	39.2	39.4	1070
All Ages	**42.2**	**41.8**	**42.0**	**5064**

In order to evaluate how the driver's propensity to be in a certain speed grouping is influenced by the driver's personal characteristics (age, gender, psychological traits, etc.), researchers conducted two analyses: Chi-squared Automatic Interaction Detector (CHAID) and logistic modeling. CHAID is a statistical process designed to divide a set of cases into mutually exclusive groups. Each of these groups will have a unique feature that differs from all the other groups for a specific parameter. Since only categorical data can be applied, the variables of age and psychological assessments are grouped together. Additionally, the algorithm outputs the percentage belonging to each of the five speed bands from within the groups.

The algorithm found some relations between the defined age and assessment groups. Some of the age categories are statistically similar. For example, the age groups 40–49 and 50–59 were combined. This is a similar impact as observed by Montgomery et al. [49] on TTC at the point of braking. This may indicate a natural grouping for other driver characteristics. Additionally, the impact of the psychological behavior can vary greatly depending on age. The higher scores on the violation self-assessment test were shown to correspond to higher speeds. However, when combined with age information, the relationship is less prominent. For example, 38 percent of drivers with a high violation score in the psychological assessment in the 30–39 age group belong to the fastest speed group, whereas only 3.5 percent of the same violation score group in the 60+ age group belong to the fastest speed group. This indicates that age has a larger impact on the determination of speed choice over the psychological assessment.

Young drivers, particularly those in their teens, drive less overall miles than drivers in other age groups, but are overrepresented in crashes. In 2015, drivers aged 16–19 had a fatal crash rate that is three times higher than that for drivers who are 20 and older. The primary reason for this much higher crash risk is likely inexperience [57]. An analysis of 539 crash events from SHRP2 NDS video data [58] found that drivers aged 16–19 experienced significantly more crashes overall and significantly more rear-end crashes compared to drivers aged 35–54.

At the other end of the spectrum, older drivers tend to be overrepresented in crash-related injuries and fatalities (due to increasing physical fragility), but not necessarily in overall crash rates except for the oldest age groups (typically 75 or 80 and older). A 2009 study of crash data over the years 2002–2006 found that drivers aged 60–69 had a lower likelihood of being involved in a crash compared to drivers in

other age groups. Crash involvement increased slightly for drivers aged 70–79, and an additional increase was found for drivers aged 80 and older [*59*].

Inattention

Driving is an attention-demanding multitask activity for the driver. It is not surprising then that past studies have found that a disruption of this process in the form of inattention can cause variations in car-following behavior and speed choice of a driver. In an analysis of the impact of inattention of near-crash and crash events, researchers applied the data set gathered in the 100-Car NDS [*60*].

The terms distraction and inattention are frequently used interchangeably in the research literature. Accordingly, researchers broadly defined inattention as any point in time when a driver engages in a secondary task, exhibits symptoms of moderate to severe drowsiness, or looks away from the forward roadway. Inattention can be more specifically described into four categories: secondary task distractions, driving-related inattention to the forward roadway, drowsiness, and non-specific eye-glance away from the forward roadway.

Secondary task distraction includes a wide array of behaviors, including hand-held devices, talking to a passenger, and eating, among others. Within secondary tasks, it has been acknowledged that certain tasks occupy more manual or visual attention than others. In this study, secondary tasks are sorted by their difficulty as simple (e.g., adjusting radio or drinking), moderate (e.g., talking on handheld device or personal hygiene), or complex (e.g., reading or reaching for a moving object). Additionally, although some tasks may be pertinent to the driving task, they also qualify as inattention. For example, glances away from the roadway toward the speedometer are considered driving-related inattention to the forward roadway. This category also includes glances to the rear-view mirrors or windows. While not a tangible item that draws away the driver's attention, drowsiness is considered inattention, with behaviors that include eye closures, repeated yawning, and minimal body or eye movement. When analyzing the baseline behaviors of drivers, 73 percent of all six-second segments collected in the study contained at least one form of driver inattention. Upon conducting an analysis based on the odds ratio, researchers found that each of the inattention behaviors had unique impacts on the likelihood of involvement in a crash or near-crash event. Table 3 displays the odds ratios for each of the distractions, along with an upper and lower confidence interval for the estimate.

Table 3. Odds Ratio Point Estimates and 95 Percent Confidence for Likelihood of At-Fault Crash (N = 49) or Near-Crash (N = 439) under Various Driving Inattention Conditions.

Type of Inattention	Odds Ratio	Lower CL	Upper CL
Complex Secondary Task	3.10	1.72	5.47
Moderate Secondary Task	2.10	1.62	2.72
Simple Secondary Task	1.18	0.88	1.57
Moderate to Severe Drowsiness (in isolation from other types of inattention)	6.23	4.59	8.46
Moderate to Severe Drowsiness (all occurrences)	4.24	3.27	5.50
Non-Specific Eye Glance Away from the Forward Roadway – Greater than 2 Seconds	0.85	0.20	3.65
Non-Specific Eye Glance Away from the Forward Roadway – Less than 2 Seconds	0.43	0.17	1.06

Odds ratios reflect the measure of association between a certain outcome and an exposure condition. In this analysis, the exposure conditions are the inattention categories while the outcome is the participation in an at-fault crash event. Table 3 shows how each inattention has a unique association with crash involvement. For example, non-specific eye glances away from the forward roadway were found to have no significant change in crash involvement, regardless of glance duration. Additionally, as expected, the categories of secondary tasks increased the odds of crash involvement with increasing complexity of the task performed. Simple secondary tasks were not found to have a significant association with crash involvement. However, engaging in a moderate to complex secondary task resulted in an increase in likelihood for involvement in a crash event by 2 to 3 times.

In order to quantify the impact of driver visual attention upon the identification of proper car-following headway, Summala et al. [61] evaluated how drivers perceived the lead car's braking lights when they were looking at the lower part of the windscreen, at the speedometer, or at the middle portion of the console. These three locations were determined to evaluate in-vehicle driver distraction—glances at the lower part of the windscreen were associated with lane keeping glances, whereas glances at the speedometer and console were determined to be distracted. Additionally, the drivers were classified by the amount of driving experience they had, including beginners (0–4,299 km), novices (4,300–112,999 km), and experienced drivers (113,000+ km). This classification allowed researchers to observe whether the ability of shifting foveal focus for safe braking is a learned behavior with driving experience. Researchers observed in the results an increased reaction time to the lead vehicle's braking based on the in-vehicle distraction. The average delay in brake reaction time was found to be 0.9 seconds for the lower windscreen position, 2.1 seconds for the speedometer, and 2.9 seconds for the mid-console position. This finding suggested that the farther away the driver's foveal view, the more delayed the braking reaction time would be. Additionally, it was determined that the driver's experience did not influence their ability to respond appropriately to the lead vehicle braking event. Although the use of peripheral vision in driving can be practiced, normal driving conditions do not necessarily offer a learning opportunity for this skill, which may explain why driver experience did not seem to improve the reaction time associated with in-car distracted driving.

With the improvement of in-vehicle technologies for passenger entertainment, the number of in-vehicle distractions is constantly on the rise. For any given distraction, there are multiple components—visual, physical, and cognitive. For example, to tune the radio to a new channel, the driver moves their visual focus to the radio display or adjusts their view to include it in their peripheral vision, causing a visual distraction from the roadway environment. Additionally, the driver physically reaches for the tuner knob and thereby is physically distracted from responding to a potential roadway hazard. The choice of station and decision-making is expected to contribute to the cognitive distraction. Accordingly, many have thought the solution should be to incorporate more hands-free and voice-activated devices in vehicles because it should reduce two out of three of these components of distraction. If reality abides by this scenario, then the overall influence the driver distraction might be minimized accordingly. To evaluate this concept, Harbluk et al. [62] designed a study to quantify the impacts of cognitive distraction on visual patterns and vehicle control through a series of tasks varying in cognitive complexity. The three cognitive load conditions were complex addition (e.g., 47 + 38), simple addition (e.g., 6 + 9), and the control group with no additional task. For each driving task record, data were gathered on the visual behavior, vehicle control, and driver self-assessment of workload, SA, and distraction. For visual behaviors, the general result was that drivers were narrowing their field of focus.

This was represented by the increasing percentage of glances that fell within the central 15 degrees of the forward roadway, the reduced number of glances at instruments and mirrors, and the decreasing number of saccades (quick eye movements characterizing scanning pattern glances) with increasing complexity. Vehicle control was characterized by the braking performance of the drivers. Accordingly, the continuous driving data from an 8-km segment was coded to find instances of hard braking, defined as longitudinal decelerations exceeding 0.25 g. Across the 16 participant data sets and 291 braking events, the mean number of braking events increased from approximately 5 to 7. This may indicate that drivers employ compensatory cautionary behaviors when faced with significant inattention. In review of the self-assessments, it was found that as the task complexity increased, the workload estimation increased, the feeling of SA ratings decreased, and the distraction ratings increased.

While the insight into the braking patterns of distracted drivers is beneficial, other research has focused on how distractions impact vehicle following behavior and driving control. Greenberg et al. conducted a study of 48 adults (age 35–66) and 15 teenagers in a simulated driving environment with eight driver tasks [*63*]. The study participants were required to respond to sudden movements in the surrounding traffic, while maintaining control of their vehicle. Vehicle control was measured in terms of heading error and following distance. Heading error is defined as the difference between the roadway tangent and vehicle heading measured in degrees at an instantaneous point in time. This value was shown to be approximately 0.5° under no distraction conditions and increased up to 0.7° under the handheld phone dialing distraction condition. In terms of mean headway distance, it was found that during one of the tasks (handheld voicemail), drivers had an increase in mean headway distance of almost 50 ft over the no distraction condition drivers. At the operational level, multiple studies have also shown [*64*] that drivers attempt to reduce their workload and mitigate their exposure to risk when they interact with their in-vehicle devices. Some of these compensatory behaviors include: decreasing speed, increasing car-following distance, changing the relative amount of attention given to driving and non-driving tasks, and accepting a temporary degradation in certain driving tasks.

Distraction

Driver distraction has long been recognized as a frequent contributor to vehicle crashes. Three consecutive years of United States vehicle crash data attribute 10 percent of all fatal crashes and 17–18 percent of injury-causing crashes at least in part to distracted drivers [*65*]. Analysis of driving data from the 100-Car NDS in the early 2000s found that drivers who were engaged in secondary visual and/or manual tasks had a crash and near-crash risk three times as high as those who were not. Distraction in secondary tasks was a contributing factor in 22 percent of the recorded crashes and near-crashes during the year-long study [*60*].

Driver distractions due to secondary tasks are generally classified as visual distractions (taking the eyes away from the roadway), manual distractions (taking at least one hand off the steering wheel), and cognitive distractions (diverting mental attention from the driving task). Distractions that involve more than one of these classifications are generally observed to be higher-risk; for example, writing text messages on a cell phone while driving involves cognitive, visual, and manual distraction, and has been found to associate with higher-risk of resulting in a crash. An observational study of 2229 crashes involving teen drivers over the years from 2007 to 2015 found an increase in the proportion of rear-end crashes over that period. The overall proportion of crashes involving distracted drivers did not change over the study's period, and neither did the proportion of crashes in which cell phone use was the

primary distraction; however, the ways in which crash-involved drivers were using their cell phones did change, with texting and other visual phone-related tasks increasing just over 4 percent per year. The study also found increases over the years in driver reaction times (from 2.0 to 2.7 seconds) and in the percentage of rear-end crashes in which drivers did not react prior to the crash (from 12.5 percent to 25 percent) [66].

Initial studies of naturalistic data obtained in the SHRP2 NDS study have supported many prior studies' conclusions on distraction and crash risk. A 2015 analysis found that visual-manual tasks including texting on a cell phone, locating/reaching for a cell phone, and adjusting an in-vehicle radio all significantly increased the odds ratios predicting crashes and near-crashes [67]. Talking or listening on a cell phone decreased these odds ratios for the events studied. A different analysis of the NDS data by the Insurance Institute for Highway Safety [68] examined the 6-second time periods leading up to 1465 crashes that occurred during the study. When comparing these pre-crash time periods to other periods of normal driving, researchers found that any secondary activity by drivers increased the odds of a crash, though some activities like talking on a cell phone did not cause a significant rise in crash risk. However, cell phone use, including simply talking on a phone, became a more significant crash risk factor when researchers excluded the least-dangerous crashes (such as tires striking a curb) and focused on the more severe crashes in the data set. Another research analysis of the NDS data [69] supported this conclusion, finding that the risk of non-trivial crashes increased by a factor of 2.2 when the driver was talking on a handheld phone. Similar results were found by Higgins et al. in another NDS study that tried to account for roadway environment as well [70]. This study found that the median reaction time increased by 40 percent among drivers engaged on visual-manual tasks with an increased risk of crash or near-crash 4.7 times as large as for undistracted drivers. Interestingly, drivers aged 16–19 were found to have faster reaction times yet higher crash risk (about 8.2 times as large as the risk for their older counterparts).

Distraction-related crash risk may also be affected by the age of the driver. An NDS data analysis [71] found that drivers who were younger than 30 and older than 65 years of age were more adversely affected by secondary-task distraction, compared to middle-aged drivers. An earlier driving simulator study found that when asked to perform secondary vocal and visual tasks while driving, drivers over 65 had slower reaction times than drivers between 35 and 45 years old [72].

Traffic Conditions and Environmental Variables
Looking from the inside of an individual vehicle, there are multiple potential factors that could lead an individual driver to select a longer car-following distance or to choose a lower driving speed. These include the age and gender of the driver, the amount of driving experience, the psychological tendencies of the driver, and the level of attentiveness during the driving task. Beyond the factors that can be understood within an individual driver's vehicle, there are two general external factors for consideration: prevailing traffic conditions and environmental variables. The prevailing traffic conditions are described as the operational characteristics of the traffic stream, including but not limited to: density, flow, volume, and average speed. The impact of prevailing traffic conditions is both physical and psychological. As an individual driver attempts to maneuver on a freeway, there are physical limitations to which paths are possible and available to the driver. However, the impact of prevailing conditions can also be psychological. In the presence of more vehicles, drivers may employ more cautionary car-following distances. In terms of the environmental variables, this category includes both inclement weather conditions, such as rain or snow, and the geometric configuration of the roadway. One

prominent area in which geometric configuration strongly affects the operational characteristics of the drivers is curves, as the speed profile reflects the driver's management of the curve.

Many of the studies that seek to understand time headway and operational characteristics attempt to quantify these variables under free-flowing conditions to remove any possible confounding effects. In order to gain a better understanding of how congested flow conditions affect the car-following behaviors, Brackstone et al. conducted an instrumented vehicle study on two different road facility types—urban arterial and three-lane motorway [73]. The study collected data about both active and passive subjects; active subjects referred to the six participants recruited to drive the test vehicle, whereas passive subjects referred to the 123 drivers observed car-following the test vehicle during the study. Researchers investigated four primary hypotheses:

1) Increase in flow and density may lead to driver increasing car-following headway.
2) Driver car-following behavior varies with road characteristics.
3) Driver car-following behavior is affected by the type of lead vehicle.
4) Drivers are inconsistent in their choice of headway.

Although researchers also found that driver car-following behavior does not vary with road characteristics, they attribute this result to the large variation observed between the active subject car-following behaviors. Additionally, researchers evaluated how the type of lead vehicle influenced the car-following behavior; generally, heavy vehicles were followed more closely than passenger cars. Many explanations have been offered for this seeming non-intuitive result, ranging from reduced workload by following a single heavy vehicle to reduced risk by following a heavy vehicle under the assumption of professional driving skills. Drivers were found to be inconsistent in their choice of adopted headway.

Beyond the natural occurrence of congestion during peak periods, there are other conditions that may cause unique fluctuations in the car-following behaviors. For example, inclement weather conditions have been shown to make a significant impact on the free-flow traffic stream parameters. Perrin et al. examined the ability to tune signal timings to account for the changes caused by inclement weather conditions [74]. They found that the saturation flow decreased significantly in inclement weather. One potential explanation is a combination of factors, such as a reduction in free-flow speed, a decrease in acceleration rates, and larger headways. These changes, in turn, can be explained by the drivers adjusting their driving due to reduced visibility. In heavy rain or snow storms, the road environment can often be hard to see. To account for this lack of visibility, many drivers may reduce their speed or increase their car-following distance to compensate.

Outside of the impact on car-following behavior and traffic SA, the surrounding environment plays a significant role in the choice of traffic speed. Hagland et al. investigated the impact of other road users on the speed choice of drivers through a comparison of observed, reported, and normal speeds [75]. In this study, 1029 drivers were stopped, interviewed, and given a questionnaire about their normal speed and speeding attitudes. The speeding attitude questions evaluated the perception of the interviewed participant about the proportion of drivers who significantly exceed the speed limit, other people's speed habits, and the overall issue of speeding. Approximately 50 percent of the variability in normal speed could be accounted for by the interviewed participant's perception of what percentage of other vehicles speed (24 percent), and the participant's perception of having to keep up with other drivers (26 percent).

Discussion

The relationships among the driver characteristics, operational performance, and SA are clearly complex and have many significant interrelationships. The influencing factors for a driver's speed choice and car-following behavior are quite varied, including everything from the psychological measures of the driver from self-reported surveys to their fatigue. Accordingly, it is of interest to the transportation research community to attempt and develop models that quantify this relationship. In many of these studies, the importance of a singular factor to the analysis is highlighted and emphasized; however, in the application of the lessons from the data set, many real-world variables are available and will be taken into consideration.

CHAPTER 4: DATA CHARACTERISTICS

The need for SHRP2 was identified in Transportation Research Board Special Report 260 [*76*], published in 2001, and was based on a study sponsored by Congress through the Transportation Equity Act for the 21st Century (TEA-21). SHRP2 was designed to complement existing highway research programs and have four focused areas of applied research:

- Safety—improve road SA by understanding driver behavior.
- Renewal—address the aging infrastructure through rapid design and construction methods.
- Reliability—reduce congestion through incident reduction, management, response, and mitigation.
- Capacity—planning and designing new transportation capacity by integrating mobility, economic, environmental, and community needs.
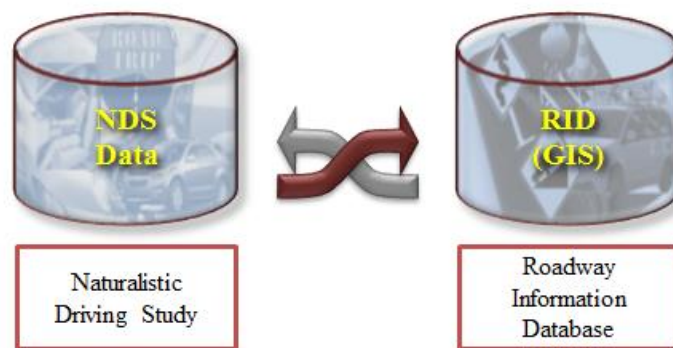
Figure 5. SHRP2 Data (FHWA).

The SHRP2 program consists of an NDS and a companion RID (Figure 5). The NDS data were collected from more than 3,500 volunteer passenger-vehicle drivers aged 16 to 98 during a three-year period, with most drivers participating for one to two years (2010–2012). The study was conducted at sites in six states: Indiana, New York, North Carolina, Washington, Pennsylvania, and Florida. The two predominantly rural sites were in Indiana and Pennsylvania and covered about 10 counties each. The other four urban or mixed sites covered one to three counties each. The total study area encompassed more than 21,000 square miles. Specifically, NDS collection sites were Bloomington, IN, Buffalo, NY, Durham, NC, Seattle, WA, State College, PA, and Tampa, FL. Collected data included vehicle speed, acceleration, and braking; vehicle controls, when available; lane position; forward radar; and video views forward, to the rear, and on the driver's face and hands. The NDS data file contains approximately 35 million vehicle miles, 5.4 million trips, 2,705 near-crashes, 1,541 crashes, and more than 1 million hours of video. Altogether, these amount to two petabytes of data. The NDS data have been divided into four main categories: vehicle data, driver demographics and survey data, trip data, and event data. In this study, researchers used driver information and trip related data. The data used in the study will be described in more detail below.

RID contains detailed roadway data collected on 12,538 centerline miles of highways in and around the study sites—approximately 200,000 highway miles of data from the highway inventories of the six study

states. NDS trip data can be linked to roadway data from the RID using a LinkID variable, which is a unique road segment identifier. The RID also provides environmental data such as time of day and weather. RID mobile data (coded S04B) were collected from the roads frequently driven by NDS participants (Table 4). Guidance was developed both for the allocation of total road data collection mileage apportioned to each of the NDS sites and for allocation within each study area. Allocation within each study area was determined using a sample of GPS traces from the NDS participants' vehicles.

Table 4. Mobile Van Data (S04B) Site Coverage.

NDS site	Miles collected
Florida	4,366 miles
Indiana	4,635 miles
New York	3,570 miles
North Carolina	4,558 miles
Pennsylvania	3,670 miles
Washington	4,277 miles

In addition to the data provided from the mobile data collection project, roadway data from existing public resources (state department of transportation, Highway Performance Monitoring System, Federal Railway Administration) and a list of supplemental data items have been acquired and included in the RID. In this project, NDS and RID data have been jointly explored to carry out the data analysis.

Table 5. Types of Factors and Examples in SHRP2 NDS.

Type of Factor	Factors	Potentially Impacts:
Driver	Age, Gender, Inattention, Distraction, Fatigue, Impairment, Personal Driving Characteristics	- **Risk level for participation in crash/near-crash events** - Time of reaction - Car-following behavior
Roadway	Edge-Marking, Rumble Strips, Lane Width, Shoulder Type, Shoulder Width, Curvature, Grade, Median, Signing	- **Operational speed** - Amount of exposure
Intersection	Control Type, Number of Approaches	- Number of conflicting movements
Vehicle	Type (car, SUV, van), Crash Prevention Technologies, Braking, Handling, and Visibility Characteristics	- Decelerations - Involvement in crash instead of near-crash (prevention technology)

Table 5 lists these factors and how they may potentially impact the data under four broad types: driver, roadway, intersection, and vehicle. The bolded potential impact areas demonstrate the connection to the current study.

Study 1: NDS Data Characteristics

The data set received from Virginia Tech Transportation Institute consisted of 847 events from volunteer drivers from Washington and Florida containing 105 variables for a wide range of analyses. Of those 754

are baseline events (i.e., normal driver, unrelated to any crash or near crash); 82 near-crash events; 10 crash events; and 1 event classified as crash-related. Up to 300 speed, gap, and headway readings at 10 ms intervals were available for each event.

As described later, researchers matched the NDS and RID data for various analyses. Depending on the specific analysis, the data sets of interest would be of different sizes because not all variables were key or of interest for all analyses, as well as some variables were available for the complete database. With that in mind, Table 6 shows typical descriptive statistics for a subset of 9,265 data that have all the variables shown available.

Table 6. Data Set Descriptive Statistics (n=9,265).

Variable	Mean	Std. Deviation	Min	Max
Follow Speed (m/s)	25.51	5.78	2.69	42.36
Lead Speed (m/s)	25.81	5.76	-2.32	67.94
TTC (s)	37.69	4683.72	-79406.22	339238.10
Number of Lanes	3.40	1.26	1	7
Through travel Lanes	3.30	1.13	1	6
Years of Driving (years)	18.42	17.99	0	69
Age Class (years)	35.35	17.91	17.5	82
PSL (mph)	57.80	7.95	25	70
SOC (m/s)	26.67	2.54	18.14	31.63

PSL = posted speed limit

More details on this and the complete data sets are given in the exploratory and formal analyses described in the next chapter.

Study 2: NDS and RID Data Characteristics

For the ramp study, the authors used NDS trip information (or *traversals*) collected from ramps at four interchanges on two freeway sections in Altoona, PA. For the analysis of speeding behavior, authors considered the state where the enforcement of the speed traffic law was not very strict to diminish the impact of this factor on drivers' speeding behavior [77]. Data were obtained from four interchanges on two freeways (Figure 33):

- Urban freeway: William Penn, Blair County.
- Rural freeway: Bud Shuster Freeway, Blair County.

The length of each ramp located on these freeway interchanges is approximately 2 miles long, which includes locations on the intersecting freeway and street, and the ramp connecting the two. To fulfill the objectives of this study, the authors identified trips where each driver had traversed the same ramp in both directions of travel. These trips are identified as:

- Merging: Street to Freeway (SF).
- Diverging: Freeway to Street (FS).
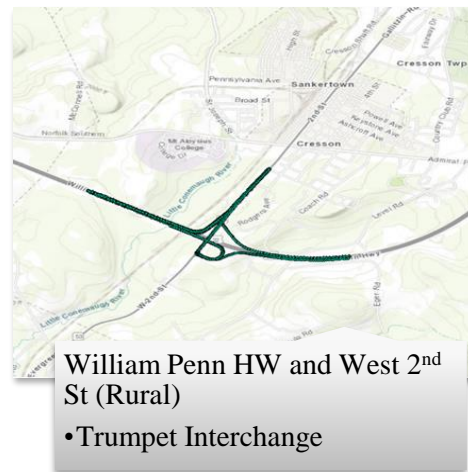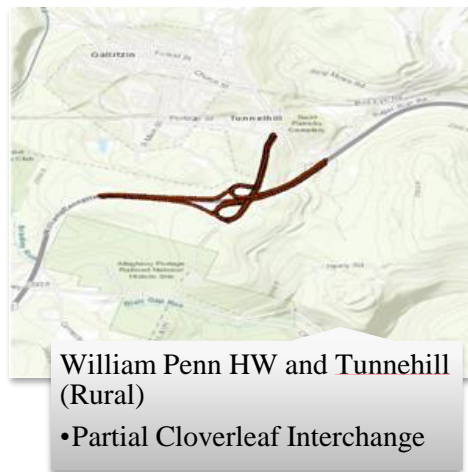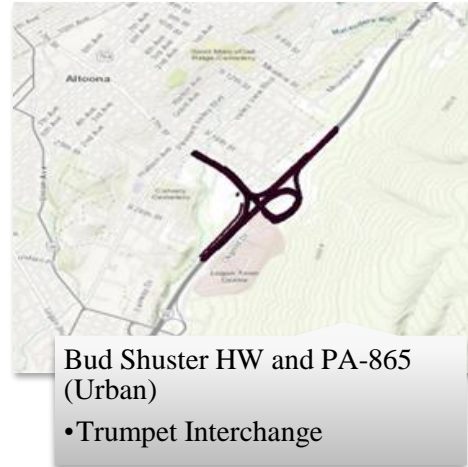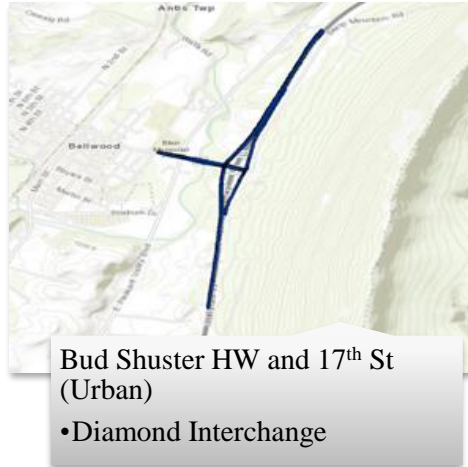
Figure 6. Ramp Trajectories.

For the analysis of speeding behavior, researchers used speeds recorded by the global position system (GPS) transponder located in each study vehicle. The NDS trip time series provides both GPS speed and the speed recorded by the vehicle's own network. However, since the speeds recorded by the GPS also have the matching latitude and longitude information for every second, the authors elected to use this data source as the primary indicator of speeding behavior.

To obtain a consistent study sample, researchers only selected trip data that spanned the entire duration of the target freeway ramps with travel times of 60–70 seconds or greater. This produced 859 trips taken by 32 participants. In many cases, drivers traveled the study route multiple times over the study period. The number of trips per driver for most of the drivers ranged from 2 to 70 trips, with the average of 30 trips per driver. Two of the study participants regularly used one of these routes; they had accumulated 341 trips (117 and 224 trips), which accounted for almost 40 percent of all trips. Such an overrepresentation in the trip numbers can present a bias in the results where the speeding will be analyzed as the function of driver characteristics among other explanatory factors. Therefore, due to the potential for undue and biased influence in the results from these two drivers if all their trips were used, researchers decided to reduce the number of trips by random sampling. For this purpose, researchers

assigned 0 and 1 to each trip randomly and selected the trips that were assigned value 1. As the result, 53 and 77 trips were left. The final study data set used in this study included 256 SF and 392 FS trips.

Next, researchers assembled a trip summary for each trip. The associated data included the year, month, weekday, time bin, and each trip's maximum speed, mean speed, and speed variance. Since there is no exact hour for the trips, instead each trip was assigned a three hours long time bins.

NDS Driver Characteristics Variables

To explore how the individual characteristics of a driver may influence their speeding behavior, researchers used the following driver information obtained through interviews and psychological testing of the SHRP2-NDS participants:

- **Driver demographics:**
  - **Age Group** is the age of the subject driver, categorized in five-year increments (16–19 years, 20–24 years, 25–29 years, etc.).
  - **Gender.**
- **Barkley's Attention Deficit Hyperactivity Disorder** (ADHD) screening: Individuals with attention deficit disorder and ADHD are prone to frequent inattention and distraction while performing tasks. In turn, inattention is a known factor associated with speeding behavior. This observation leads to the potential hypothesis that those who score high on Barkley's ADHD screening test may have higher speeding incidences than those who score low on the test. The six items included in the Barkley's ADHD screening are: 1) easily distracted; 2) difficulty organizing; 3) loses objects; 4) quick screen–difficulty waiting turn; 5) feels restless; and 6) difficulty enjoying leisure activities. Each of these six items is scored by the participant by using one of the following three answers: Never or Rarely (1), Sometimes (2), and Often (3). The Barkley's ADHD score was then calculated using the answers provided to these items [78].
- **Risk perception score (RPS)**: Risk perception is well-documented in the literature to have a strong impact on speeding behavior [79]. Those with low risk perception tend to have a high perception of their individual driver control. As part of this characteristic, these drivers tend to dismiss risks, exude a high self-confidence (especially about their driving ability), and demonstrate unrealistic optimism [80]. It is hypothesized that those with low risk perception (which equates to a low RPS score) will have higher number of speeding incidents than those who have a high RPS score. Table 7 lists the questions used for the RPS. Respondents answered to the questionnaire by assigning No Greater Risk (1), Moderately Greater Risk (4), and Greater risk (7) to each question.

Table 7. Elements of Risk Perception Questionnaire.

- Running red light
- Risks for fun
- Sudden lane changes
- Running stop sign
- Speeding for thrill
- Tailgating
- Illegal turn
- In a hurry
- Risk of passing on right

- Yellow light acceleration
- Driving after taking drug or alcohol
- Road rage
- Driving to reduce tension
- Passenger interaction
- Racing
- Speeding <20> mph over limit.
- Not wearing safety belt
- RPS

1- No Greater risk; 4- Moderately Greater Risk; 7- Greater Risk.

- **Driving Behavior:** An illegal, high-risk, or otherwise detrimental driving behavior that the subject driver was observed to be engaging in at the time of the event, such as driving while distracted. (This can also be coded as none meaning that no improper/detrimental driving behavior was observed at the time of the event.)
- **Secondary Task:** Any non-driving task that the subject driver was engaged in at the time of a given event (reaching for object, interactions with passengers, adjusting radio or other in-vehicle instruments, interacting with cell phones or other devices). For this analysis, the only secondary tasks that were included were cell-phone related.
- **Sleeping Habits**: Previous research has found personality differences between long and short sleepers. Specifically, short sleepers were found to be efficient persons that handle stress by keeping busy and by denial. Long sleepers, on the other hand, had higher instances of depression and anxiety and scored higher on most pathology tests [81]. This portion of the evaluation is an exploratory analysis on both sleep schedule (i.e., whether the participant keeps a regular sleep schedule, Yes or No) and sleeper type (i.e., light, normal, or heavy) to determine if these factors encompass various traits related to speeding propensity.
- **Depth Perception**: Depth perception is the ability to visually perceive the world in three-dimensional space and is necessary to accurately determine the distance to an object. Depth perception is a personal characteristic that directly affects an individual's visual perception, and therefore may be connected to speeding behavior. Participants were shown a picture of four rings (top, bottom, left, and right) and were asked if the bottom ring seems to be floating toward them. If the participant answers yes, they moved to a second picture and were asked which ring seems to be floating toward them. Drivers who cannot see the ring floating toward them received no score. For all others, they were scored in seconds of arc, where the smaller the seconds of arc, the better depth perception of the participant.

CHAPTER 5: METHODOLOGY

This chapter describes the methodologies applied in this project. It comprises two main subsections, one for each of the two studies in this research.

Study 1 Speed Choice and Car-following Behavior

This section summarizes the methodology used for the evaluations under Study 1.

Conceptual Model

As a first step of the analysis, researchers devised and refined a conceptual model to encapsulate the elements of interest to this project. Figure 7 depicts the summary diagram of relationships of interest.
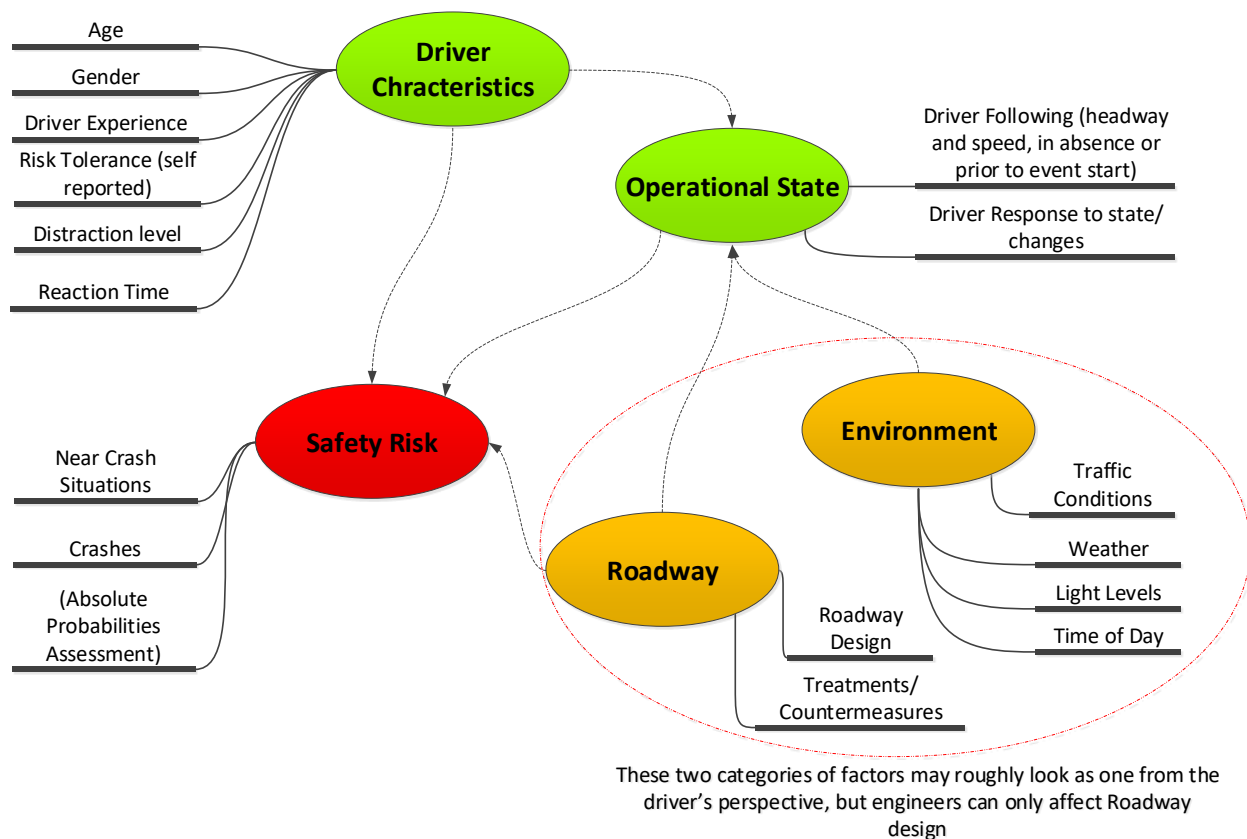
Figure 7. Relationships of Interest.

Figure 7 shows three broad categories of variables (color-coded). The Driver Characteristics and Operational State categories (in green) represent the most fundamental information about the driver and are the building blocks to their behavior. These variables are key to understand the emergence of SA performance. The directional arrow from Driver Characteristics to Operational State indicates that the former category of variables influences and helps determine the operational state, as measured by the two variables under that category: driver car-following and driver response to stimuli.

The SA Risk category (in red) represents variables that directly quantify SA performance (crashes, near crashes, and sampling characteristics that determine measures of absolute crash risk). As indicated by the directional arrows, variables in the SA Risk category are influenced by variables in the other categories.

Finally, the Roadway and Environment categories (in orange) indicate infrastructure and environmental elements that have a direct effect on both SA performance and OP.

Having Figure 7 as a framework, researchers refined its conceptual model of the process of driver car-following on freeways and highways. The refined conceptual model.
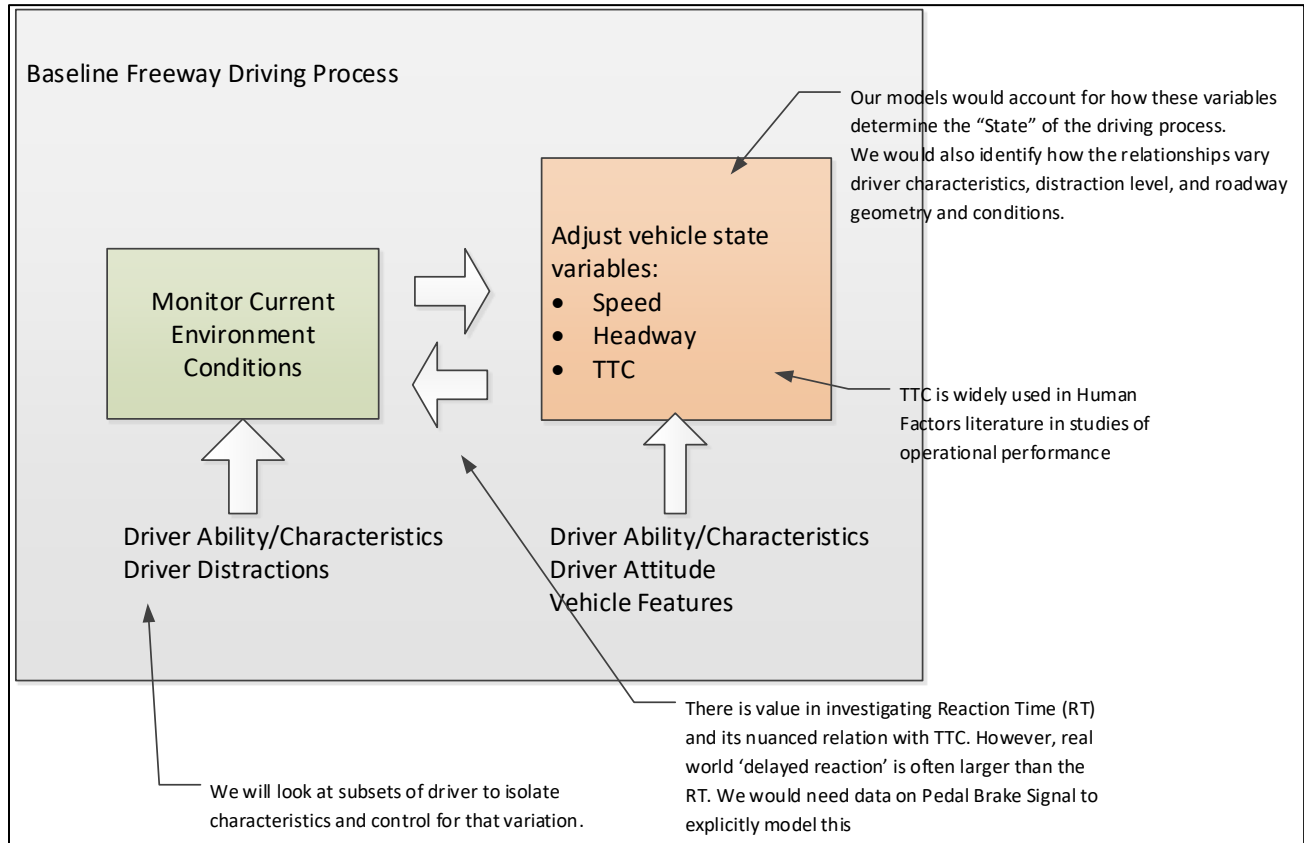
Figure 8. Baseline Freeway Driving Conceptual Model.

The conceptual model in Figure 8 represents the task of driving as two processes performed by the driver within a feedback loop: 1) monitoring current driving and environmental conditions, and 2) adjust vehicle state variables. Driver ability and characteristics affect both the first and second processes. In the early stages of the analysis phase, researchers devoted significant effort to quantify some of the relations in this figure using a baseline model calibrated only to non-critical driving events.

Closely related with Figure 8, Figure 9 shows this conceptual model in the face of a SA critical event (defined as an event that demands a response from the driver to avoid a collision).
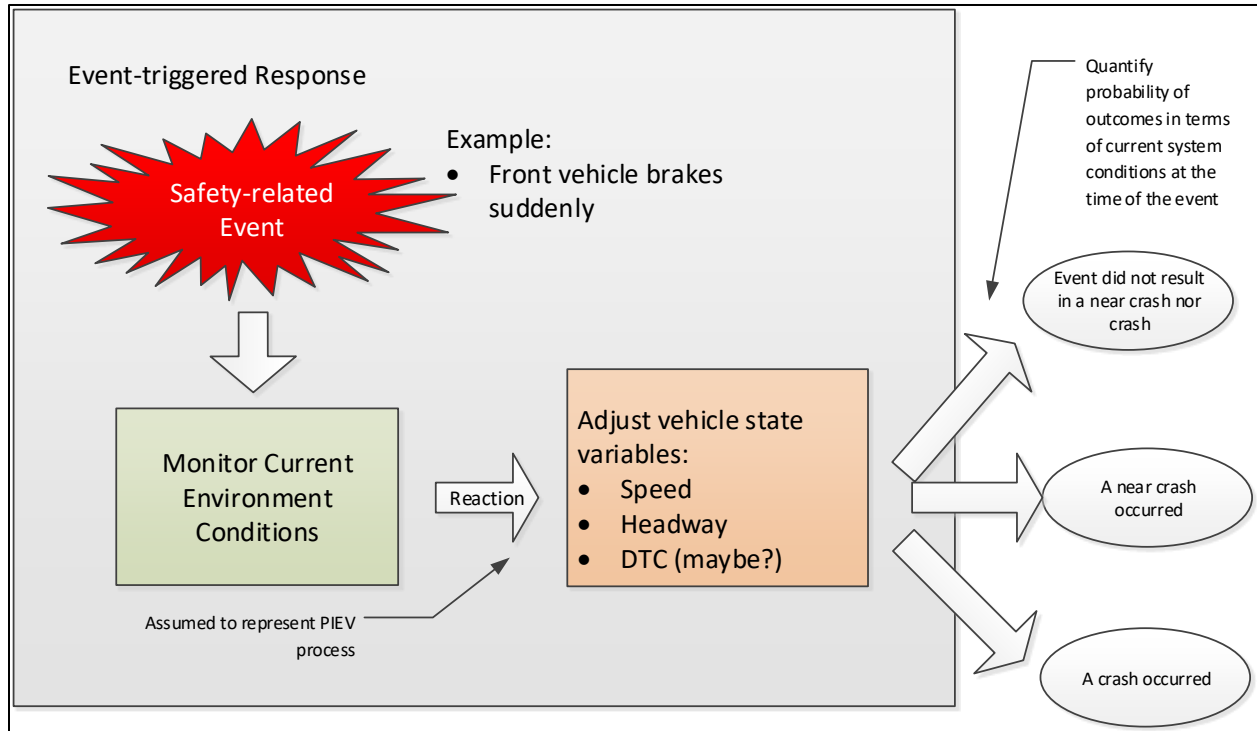
Figure 9. Conceptual Model for Safety Critical Events (Rear-End Crash).

In the case of a SA critical event, the time it takes to recognize such an event and react to it could make all the difference in determining the SA outcome, as shown in Figure 9. An additional step on the analysis phase consisted of using the baseline model to examine the crash and near crash situations in the data set.

Car-Following Models

Brackstone et al. summarized the microscopic car-following behavior models into five broad categories: Gazis-Herman-Rothery model, collision avoidance model, linear model, psychological or action-point model, and fuzzy-logic-based model [*33*]. Equations in Table 8 represent the Gazis-Herman-Rothery model, collision avoidance model, and linear model of car-following behavior, respectively, with the other two categories (psychological and fuzzy-logic based models) discussed afterward.

Table 8. Car-Following Equations [*33*].

$$a_n\,(t) = cv_n^m(t) * \frac{\Delta v\,(t - T)}{\Delta x^l\,(t - T)}$$

$$\Delta x\,(t - T) = av_{n-1}^2(t - T) + \beta_1 v_n^2\,(t) + \beta v_n\,(t) + b_0$$

$$a_n(t) = C_1 \Delta v\,(t - T) + C_2(\Delta x\,(t - T) - D_n(t)$$

$$D_n(t) = \alpha + \beta v\,(t - T) + \gamma a_n(\,-T)$$

Where:

- $\alpha, \beta, \gamma, C1, C2, c, m, n = constants$
- $a_n, v_n, D_n = acceleration, velocity, and\ desired\ following\ distance\ of\ vehicle\ n$
- $\Delta x, \Delta v = change\ in\ relative\ distance, speed\ between\ vehicle\ n\ and\ n-1$
- $t, n = instantaneous\ time, vehicle\ number$
- $T = driver\ reaction\ time$

The psychological (action-point) model is built around conceptual thresholds of perception, where drivers perceive the differences in relative speed and base their driving and car-following behavior from that point. There are five general phases in car-following with different threshold guidelines: free driving, following I, following II, closing in, and danger [33]. Free driving is uninfluenced by the leader vehicle, and following I and II conditions present no need for much driver adjustment. However, for closing in and danger conditions, deceleration becomes necessary. While this method may better represent the human process of making following decisions, it is difficult to estimate and calibrate the individual thresholds associated with this model. The fuzzy-logic-based model applies fractional degrees of membership to values; for example, if a separation between leading and following vehicles is only 0.5 seconds, fuzzy classification would still allow for the use of categorical evaluation by granting partial membership of 0.5 seconds to class 0 and 1. Similar to the action-point psychological model, the most difficult part is in the quantification of membership functions. These models are generally limited in scope, however. For example, very little concerted effort has been performed to develop a complete driver model since the 1960s. Instead, most models focus on understanding a singular aspect of driving.

Prior to applying a car-following algorithm, it is critical to understand how these types of algorthims relate to true conditions. To understand the differences between available car-following models and their errors to the true data from the instrumented vehicles, Panwai et al. collected field data and compared it to prominent models, such as AIMSUN 2, VISSIM, and PARAMICS [82]. The last two software are predominately driven by a psychophysical model that employs one of four general regimes, depending on the perception of changing distance and speed between the current vehicle and the lead vehicle.
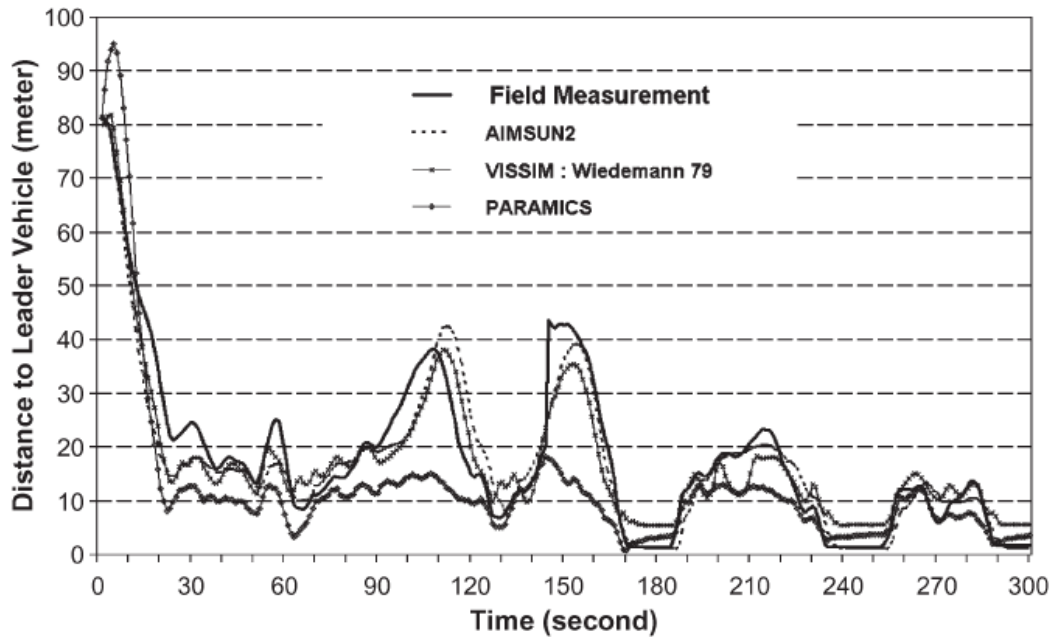
Figure 10. Comparison of Car-Following Models [*82*].

The AIMSUN2 and VISSIM models seem to follow very closely the field measurement in Figure 10. The logic behind the AIMSUN2 model is based on the Gipps model that determines the desired speed by adjusting the current speed by the driver's desired speed and the relative speed and position of the lead vehicle. These elements will be considered in the analysis section.

Cointegration and Dynamic Relationships

Two time series are assumed to be co-integrated if the residual term, when regressing one of the time series variables over another is determined to be stationary. For example, as described in Figure 11, the series y_t and z_t are random walks with drift. The residual term that is depicted on the right panel is a stationary process. Cointegration of two variables is determined by conducting unit root test such as Augmented Dickey-Fuller (ADF) test [*83*].

The term dynamic relationship refers to two time series data where the leading (independent) event has a lagged effect on the response variables. For example, the car-following driver can react to the leading vehicle's speed changes after few seconds. In that case the leading vehicle will have a lagged effect on the car-following vehicle. This relationship can be tested using the Granger causality test.

Mixed Effects Models

Within the frame of generalized linear models methods, a distinction can be made between models with fixed effects, random effects, and mixed effects. Commonly, the coefficients obtained from generalized linear models can be thought of as fixed effects. The variables corresponding to fixed effects are implied to have time-invariant effects (e.g., roadway design elements). The model coefficients are estimated and interpreted as metrics of underlying parameters from a latent data-generating process.

a) Cointegrated series   b) Residual term
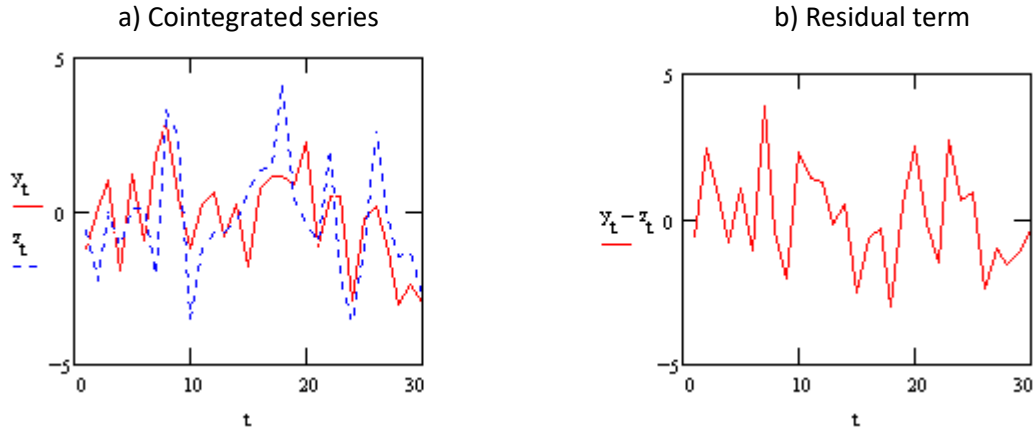
Figure 11. Cointegrated Series and Residual Term.

In contrast, random effects models estimate the effects of factors that are deemed the observed realizations of a random variable. As such, it is typically not of interest to quantify how the response variable shift with the observed realizations in the data set, but rather to account for the impact of such variation in the model. The simplest example of random effects in linear models is the use of blocking in ANOVA designs. Typically, the effect of each block is not the focus of the analysis. However, it is of interest to account for the variability from the blocking to quantify the response variability from the independent variable of interest. Mixed effects models include both fixed and random effects [*84*]. Mixed effects models approach the analysis of repeated measures cross-sectional data by including a random effect per every unit of data aggregation (i.e., the blocking units in the data, such as individual study locations with more than one datum in the analysis). Orthogonal to the random effects, the model estimates fixed effects for the treatment and any additional fixed effects covariates. As described in the previous section, the use of propensity score matching can produce a more robust data set for analysis with mixed effects and other analytical alternatives for cross-sectional data.

Study 2 Ramp Speed Choice

Time Series Reduction

Time series reduction tools are used to reduce the dimension of time series data. For example, due to the speed differentials, the length of the series (i.e., number of seconds) included in the analysis of ramp speed choice ranged between 70 to 850 seconds. Such a significant inequality among the series lengths could cause a concern. This problem can be dealt with by reducing the time series dimension (length-wise) using the algorithms such as Discrete Wavelet and Discrete Fourier transformation.

In this study, researchers used the Discrete Wavelet Transformation (DWT) to reduce the dimension of relatively longer series. Namely, the Haar wavelet is used to conduct the dimension reduction. Haar wavelet allows the time series of length T to be represented in terms of its orthonormal basis by calculating a set of averages and coefficients (usually $\sqrt{2}$ to ensure energy conservation) [*85, 86*]. After the first iteration, the time series length reduces to $T/2$. The resulting time series is referred as the first level of wavelet transformation. The DWT method can be applied recursively until a single coefficient and average is obtained.

Neural Networks

To describe the relationships between the predictors (inputs) and the speeding state (output), the multi-layer perception Neural Network (NN) architecture can be trained by a backpropagation algorithm. NN methods are known for their ability to deal with a relatively large number of predictors. The NN framework or architecture has three elements: input, hidden layer, and output (Figure 12). Input and output refer to the predictors and response variable, respectively. Hidden layers are the collection of neurons organized and connected to each other using the arrows that are referred to as the weights. Weights can also be understood as the parameter estimates although they should not be interpreted as such.
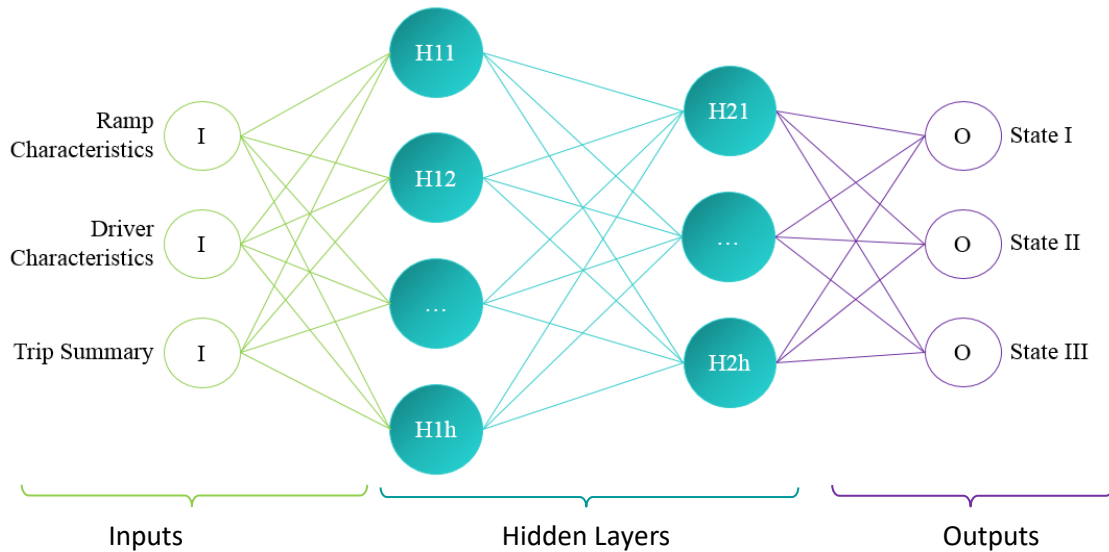
Figure 12. Neural Networks Framework.

In the NN architecture, the relative contribution of the inputs to the output depends on the magnitude and the direction of the connection weights [87]. Connection weights are computed using the weights of individual inputs in each hidden layer. Greater connection weight indicates the higher intensity of the association. Negative connection weights represent an inhibitory (reducing) effect while the positive connection weight represent an excitatory (increasing) effect of the neurons on the output. Figure 12 shows the NN architecture.

CHAPTER 6: STUDY 1 FREEWAY SPEED ANALYSIS

To gain insights from the data and to understand the problem further described later in this chapter, researchers initially examined a subset of the data with a series of graphical comparisons described in the following section.

Initial Exploratory Analysis

This section shows a preliminary exploration of this data set. Researchers selected a subset of events under very specific conditions: a) occurred at interstates or highways with no traffic signals; b) no traffic control recorded; c) with divided median, either a barrier or a buffer strip; d) two through lanes in the direction of travel without additional auxiliary lanes; e) no adverse weather; and f) traffic flow classified as Level-Of-Service A1: free flow, no lead traffic.

The selected subset consisted of 20 baseline events, each from a different driver. An additional 4 near-crash events were selected that correspond to a subset of 4 among the 20 drivers with baseline events. Finally, a near-crash event without a corresponding baseline event was also included in the preliminary subset.

The number of subsequent speed readings available for each of the selected events ranged from 176 up to 300. A total of 5,430 10-ms intervals were extracted for this subset of baseline events.
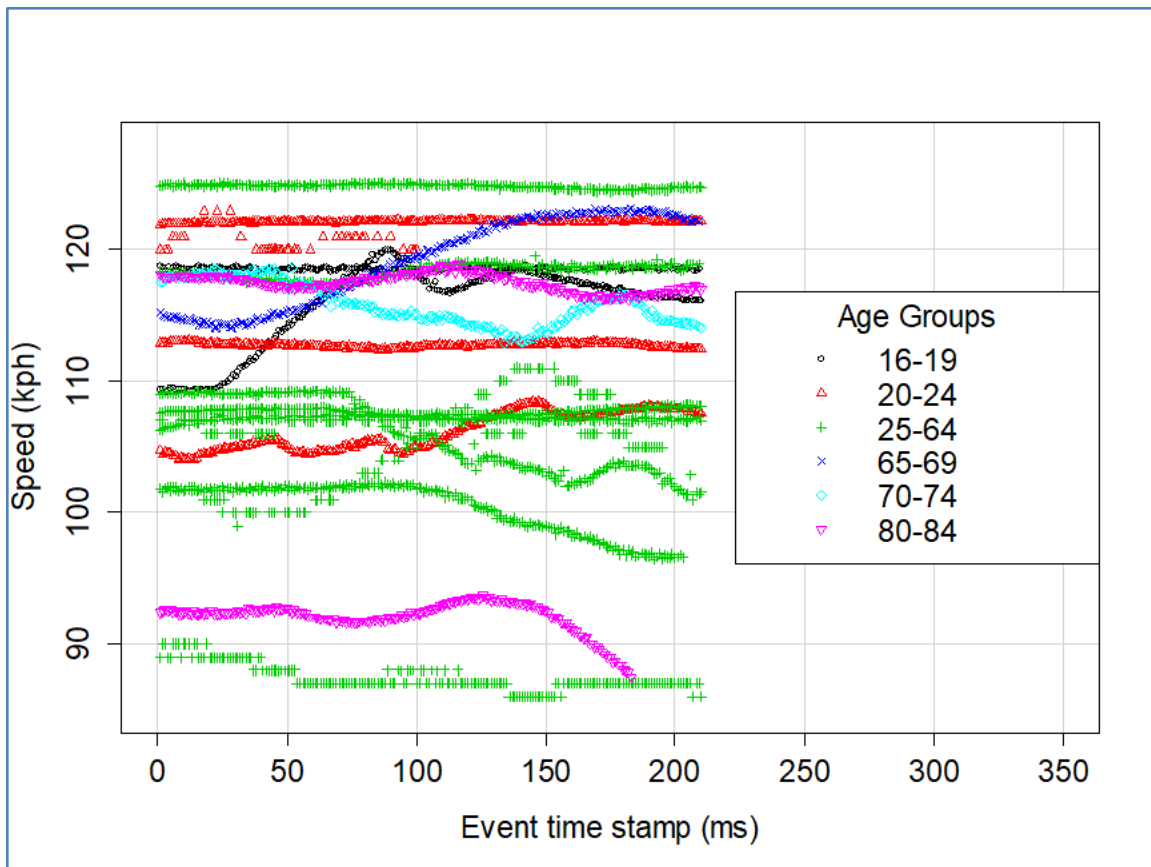
Figure 13. Baseline Events Speed Profiles (n=19).

Figure 13 shows a plot with 19 of the baseline speed profiles. One profile was removed because it showed a monotonic increasing trend (i.e., continuous acceleration). Since these events were selected such that no other vehicles were around, each represent the desired free-flowing speed for different drivers. Interestingly, the overall speed distribution shows a negatively skewed distribution (i.e., heavy lower tail). When looking at the data by age groups, the wider group (ages 24–64) seems to be roughly centered (i.e., not skewed). Little can be said about older drivers (ages 70–74 and 80–84), since only three profiles were present for these groups. Two of those three profiles lie above the grand average for the group of profiles, while one lies below the average. What is most interesting in this plot is the fact that younger drivers (ages 16–19 and 20–24) tend to have higher speed profiles that other age groups.
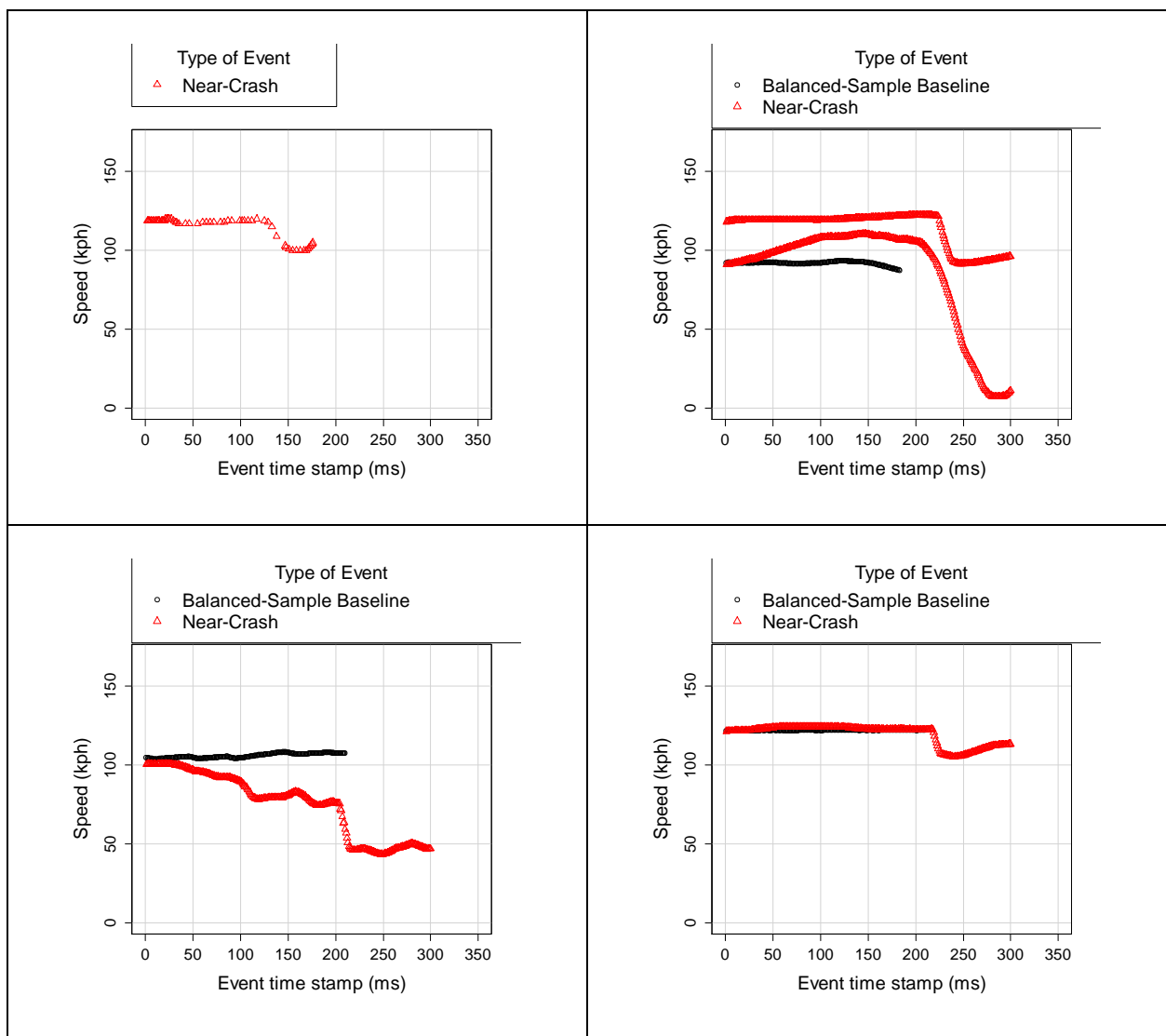
Figure 14. Four Drivers with Both Baseline and Near-Crash Events under Free-Flowing Conditions (n=4).

The plots in Figure 14 show the four drivers with near-crash events. In total, 8 events (5 near crashes and 3 baseline) are shown. In the first cell of this figure, there is a near-crash event with no corresponding baseline event. The second cell in Figure 14 (top right) shows that, compared to his or her baseline speed profile, this driver was going at a higher speed just before the event happened. There is no apparent difference between the baseline speed and the speed before the event for the remaining two cases (lower left and lower right).

It is not clear that speeding is a factor in each of the four cases shown. However, further investigation on the type of event, at other facility types and under heavier traffic conditions, would be required to make more solid inferences about the relationship of speed and likelihood of near-crash or crash events.

Pre-Processing

This section summarizes the analysis tools that were used to pre-process the data sets for analysis.

Piecewise Linear Representation

Researchers considered the application of Piecewise Linear Representation to discretize the time series into segments of known length and slope. To obtain these estimations, it is necessary to first assume that the time series of interest has some number of breakpoints whereby the segments between these points can be represented by a stable regression relationship. In the driving environment, this is a reasonable assumption. Given the two available inputs for the driver to adjust their speed (brake and gas), there is a set number of actions for a driver: accelerating, decelerating, and remaining constant.
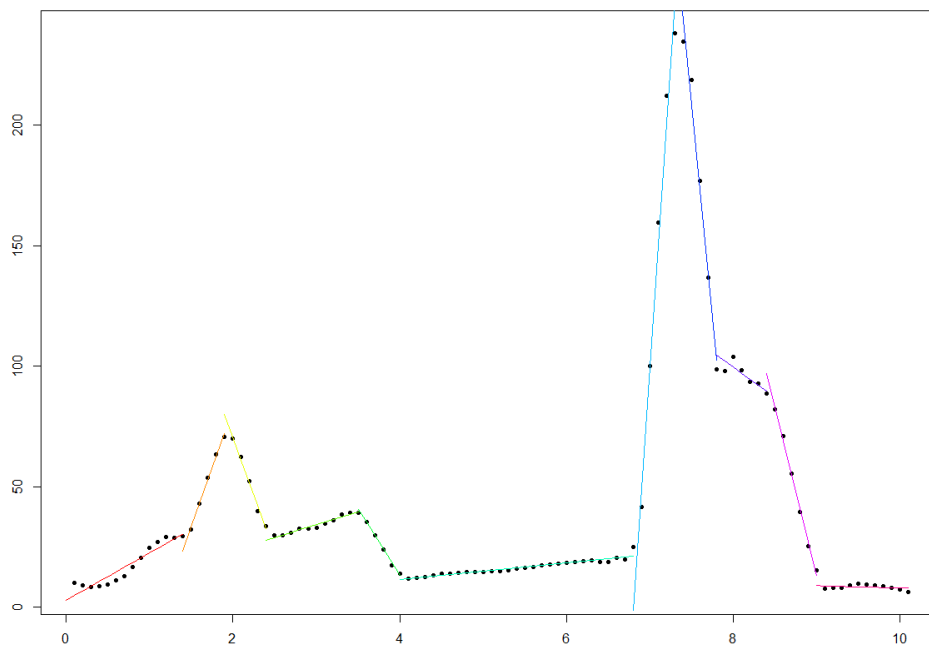
Figure 15. Reducing Time Series to Set of Linear Segments.

In each of these states, a linear relationship can approximate the impact on the speed. Accordingly, a linear model with an unknown number of breakpoints is fit through the algorithm developed by Bai et al and implemented in R by Zeileis et al. [*88*]. A sample time series (TTC) is fit with these lines in Figure 15.

Each of the line segments above have a known length and slope. From this information, a letter can be assigned to the distinct combinations of length and slope. As pre-supposed about the driver behavior, there are three types of slope that can be observed (negative, relatively constant, and positive). Accordingly, a state name can be assigned to each of these distinct slopes. For the purposes of simplicity, the letters A, B, and C were assigned to each of these slope types, respectively. This tool will be considered in breaking down events into subevents of interest for analysis.

Time Series Smoothing

While there is a multitude of available methods to interpolate missing data and clean errors in time series, such adjustments must be done carefully and considering the fundamental properties of time series. In this exploratory analysis, two techniques were explored: LOcally weighted non-parametric regrESSion (LOESS) and Kalman filter. Figure 16 shows four TTC time series from the data set as a scatterplot and their respective interpolated counterparts as the solid line.

(a) TTC Data

(b) Speed Data
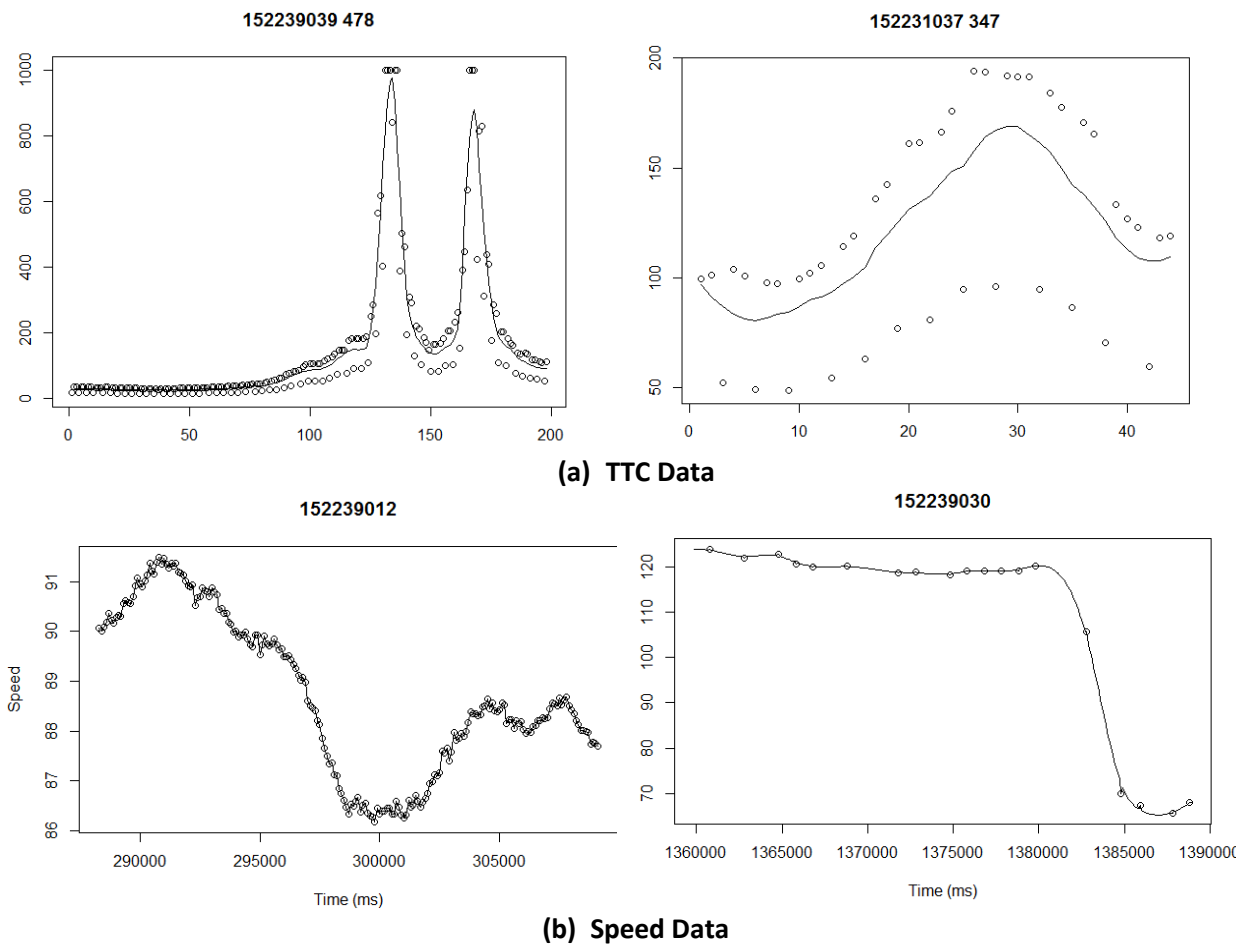
Figure 16. Original and Cleaned Continuous Profiles.

The Kalman smoothing algorithm functions in a series of steps. An initial estimate is developed for the state parameters in a forward pass, similarly to the traditional Kalman Filter. Following this step, a backwards pass estimate is conducted, and the error is minimized through Expectation Maximization. In

this manner, the various hidden causes for changes in speed can be accounted for and cleaned without prior knowledge of their influence on the speed curve. The Kalman smoother is known as a robust estimator in time series data and can account for sensor noise.

With respect to LOESS, there are a few parameters that must be determined prior to its application to this data set, such as the function order to fit and the number of points to consider. Due to the curvilinear nature of the TTC profiles, a 2nd order polynomial was chosen. The second parameter, span, refers to the percentage of points that will be used in each regression step. Although optimization via bootstrapped samples would allow for the ideal span choice, the span choice should reflect the fundamental reaction ability of an individual to adjust their speed or braking behaviors. Accordingly, the span was chosen to align with research from Triggs et al. in a review of reaction times that shows the mean driver response time to an expected braking light stimuli of the leading vehicle is between 1.39–1.45 seconds [89]. The span was chosen to be 1.4 seconds, or 14 data points in a 10 Hz time series. This was represented as a fraction of the overall time series, whose length varied by event. The resulting range of spans, expressed as proportion, was between .05 and .50. Although many researchers support the use of different span percentage until an ideal fit is found, the set length of time for the span allows for both the short and long-time series to be well-understood. Figure 16 shows two very different TTC profiles; the first has two main peaks and lasts almost 20 seconds, whereas the second is a little more than 4 seconds and features a large gap between data points. At first glance, the data from the right appear to be generated by two-separate time series, separated by a vertical displacement.

(a) Three-Way Alignment

(b) Cost Matrix and Alignment Path
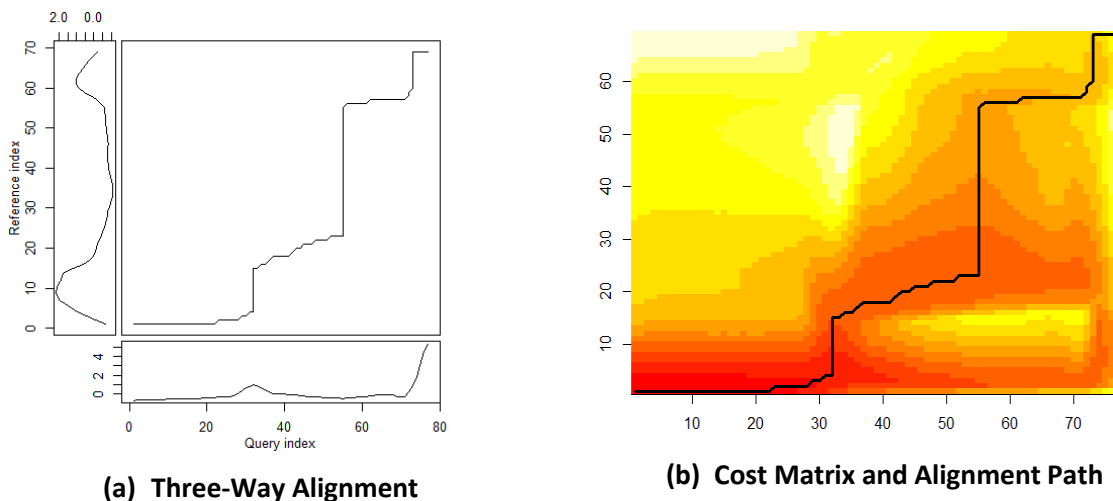
Figure 17. Dynamic Time Warping between Two TTC Curves.

However, in manual observation of the individual points that comprise the TTC calculation, the root cause for the two series is relatively straightforward. For two vehicles in a car-following situation, the random noise in the sensor causes the observed vehicle's speed to increase slightly more than its leading counterpart, the TTC will suddenly decrease. Similarly, if the observed vehicle's speed decreases slightly more than its leading counterpart does, the TTC will suddenly increase as the two vehicles are separating. Thus, the seemingly bizarre result can be well-understood in the context of TTC.

Cluster Analysis and TTC

Past research, such as that by Guo et al. and Quimby et al. [31, 53], acknowledged the potential for methods that can approximate natural groupings within the data set to provide clarity to the research effort. Clustering approaches include k-means clustering and CHAID. An alternative method to do clustering is named dynamic time warping (DTW). Researchers performed an exploratory DTW analysis.

Guo et al.'s analysis applied a k-means clustering approach that uses aggregate single variable observations [90]. This approach is done in two steps: 1) the user determines the preset number of clusters, 2) the algorithm selects a random set of points from the data set corresponding to the predetermined number, 3) the algorithm calculates the internal differences and evaluates the total sum of squares for each cluster, and 4) the algorithm repeats the first three steps by adding new points to each cluster so that the total error within each cluster is minimized for several iterations. In another study, Quimby et al. applied CHAID to divide a set of cases into mutually exclusive groups [53]. The dependent variable in the analysis was five speed groups along a continuous spectrum (lowest speed to highest speed with five levels), and the influence of eight predictor variables (including psychological tests and age).

In this analysis, the DTW method was applied using an open source package [91, 92] to investigate what factors seem to associate with differences in the time series characteristics. DTW is a method by which the dynamic programming algorithm finds the optimum warping path between two time series under a set of constraints. Figure 18 illustrates the warping done by the algorithm, with the purple lines showcasing the DTW distance between the two time series (black and red lines, respectively).
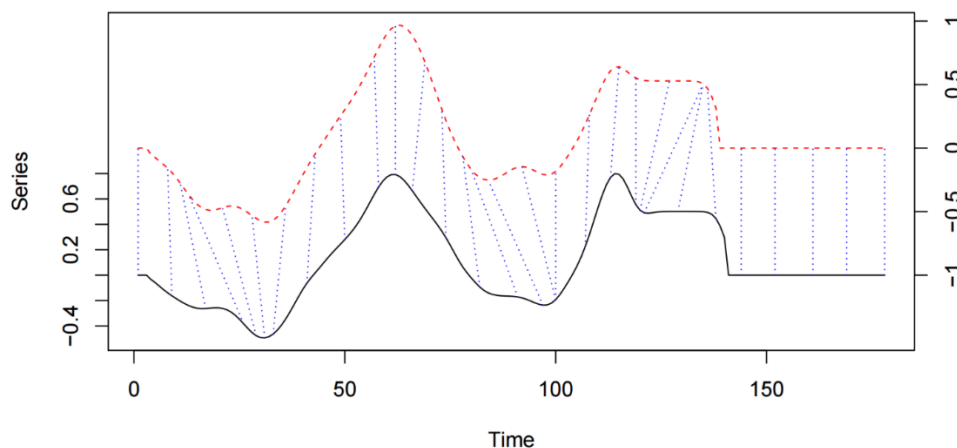
Figure 18. Example of DTW between Two Series [*92*].

DTW applied to naturalistic driving behavior operational data may help determining factors influencing a given operational regime, potentially shedding light on the driver sensitivity factor (k) in microscopic simulation. Another result from this effort is a better understanding of the influential factors as determined by a natural grouping of the data. Final models generated by this project can aid in variable selection by the results from these clustered data.

Beyond the speed profile data, this exploratory analysis will apply this approach to TTC variable to investigate to what extent this variable can be studied in conjunction with speed profile data [*93, 94, 95*].

Time Series Cluster Analysis

To generate clusters, a given distance measure has to be established to evaluate similarity between any two time series. While many potential measures can be used, one of the more common, DTW methods will be applied. This method finds the optimal alignment between two temporal sequences, such that the overall cost of alignment is minimum. This is represented in Figure 17, which demonstrates the alignment between two time series.

Figure 17a shows the three-way alignment plot for the time series. The query is the bottom time series, and the reference is the time series on the left. The cost for the given alignment between any two points on the query and reference index is shown by color intensity in Figure 17b. Red demonstrates the minimum cost path, whereas white would indicate the highest cost. The alignment takes the minimum cost path through the cost matrix on the right. While the DTW calculation provides a powerful way to compare two time series across a temporal stretch or distortion, it is a resource-intensive calculation. There are many methods available to speed up or fine-tune its performance, ranging from providing designated windows for the cost matrix search to adjusting the step parameters. Until more data can be tested and optimal windows found for TTC and speed profiles, a generic application will be able to find an appropriate alignment between the two curves.
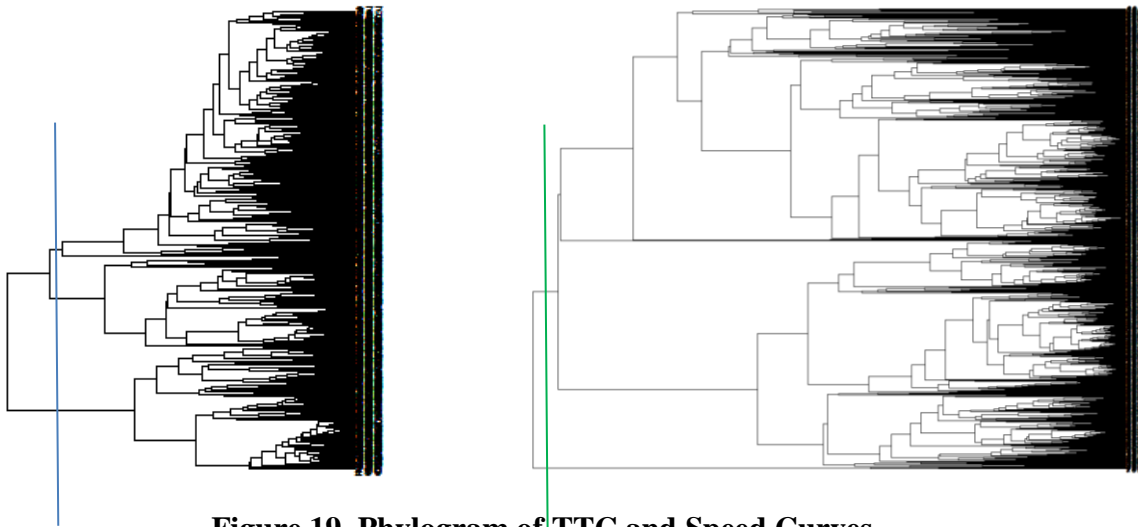
Figure 19. Phylogram of TTC and Speed Curves.

From the DTW distance, Ward's hierarchal agglomerative clustering is applied to clusters based on each of the DTW comparisons. This method has been used before to understand genetic and evolutionary progress across key indicators in different species. When two entities have a minimal pairwise difference, they are merged to the same group until each of these have been appropriately assigned to a cluster. The phylograms are shown in Figure 19.

(a) Cluster 1: Increasing TTC

(b) Cluster 2: Constant with Peak TTC

(c) Cluster 3: Decreasing TTC
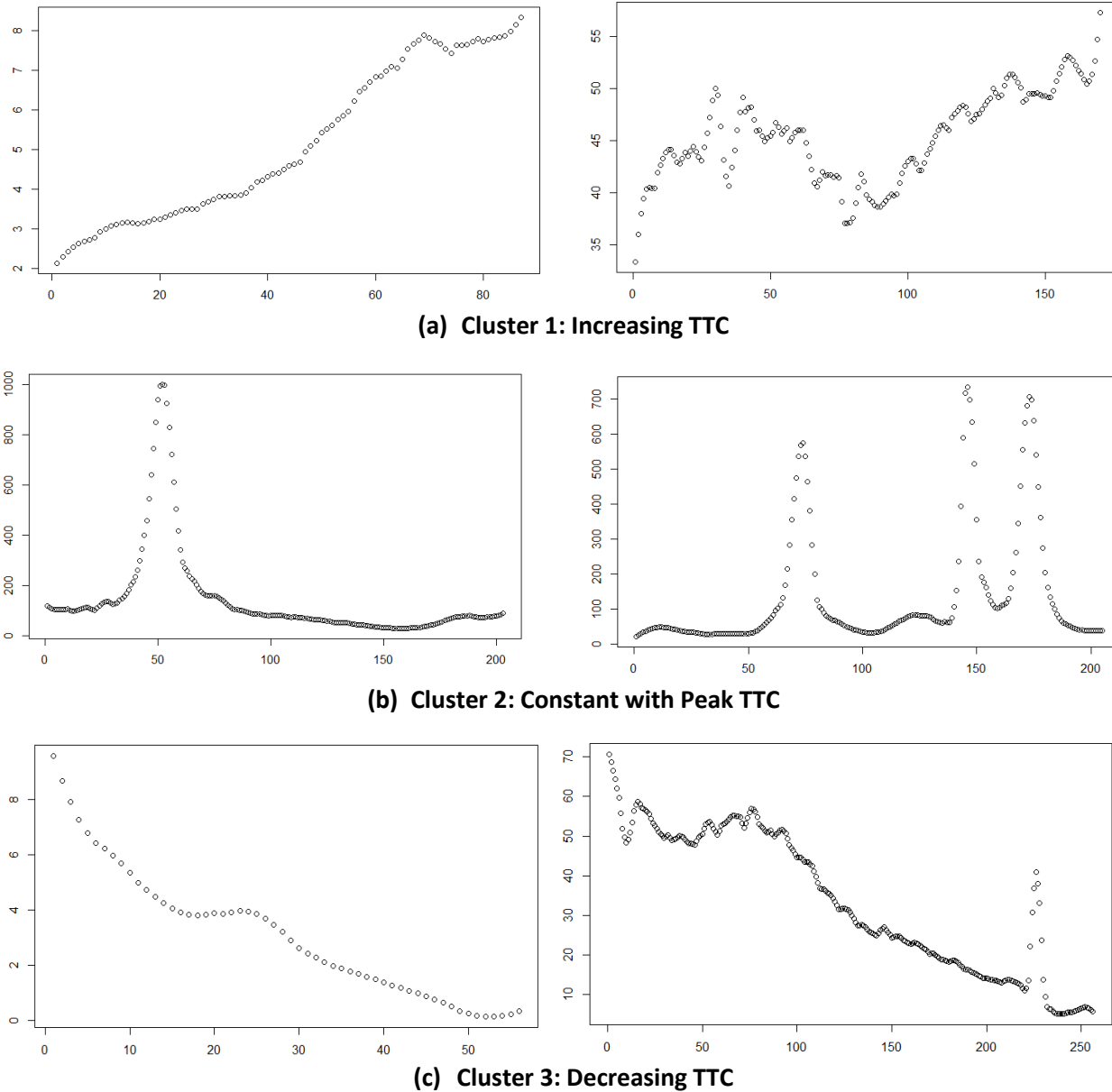
Figure 20. Examples of Clustered TTC Data.

The blue line drawn on Figure 19 (left) shows the cut-off point for TTC to generate three clusters. The phylogram clearly becomes much denser beyond this line with 108, 217, and 79 events in the three clusters defined, respectively. The green line drawn on Figure 19 (right) similarly shows the two speed clusters. These were defined by merging the most similar events by their absolute value DTW distance along the alignment path. The result is three distinct clusters of TTC – increasing, constant with peak(s), and decreasing.

Within each cluster, there are simple examples of the cluster with minimal change in pattern over time. However, there are also multiple examples within each cluster that add more degrees of complexity to the classification dilemma. In Figure 19, there are two time series shown from each cluster of TTCs. The left time series represents the ideal shape from that cluster, whereas the right shows a much noisier or more complicated time series from within the same cluster.

In Figure 20a, for example, the left curve has only a slight variation around seven seconds into the profile. While the overall trend is upward for the right curve, there is much more variation than in Figure 20a with multiple changes up and down.

Similarly, cluster 2 features constant TTC with sharp peaks. These sharp peaks indicate an increase in TTC, brought about by a speed increase of the lead vehicle or sudden braking by the following vehicle. In Figure 20b, the left image maintains a constant with only one major peak, whereas the right image has three. In Figure 20c, the left image shows a relatively smooth descent. Although Figure 20c on the right has a noticeable peak, the overall trend indicates that this time series is in fact decreasing. This occurs frequently in unsupervised time series clustering, with multiple events having characteristics of one dominant cluster and features of others.

(a) Cluster 1: Increasing Speed

(b) Cluster 2: Decreasing Speed
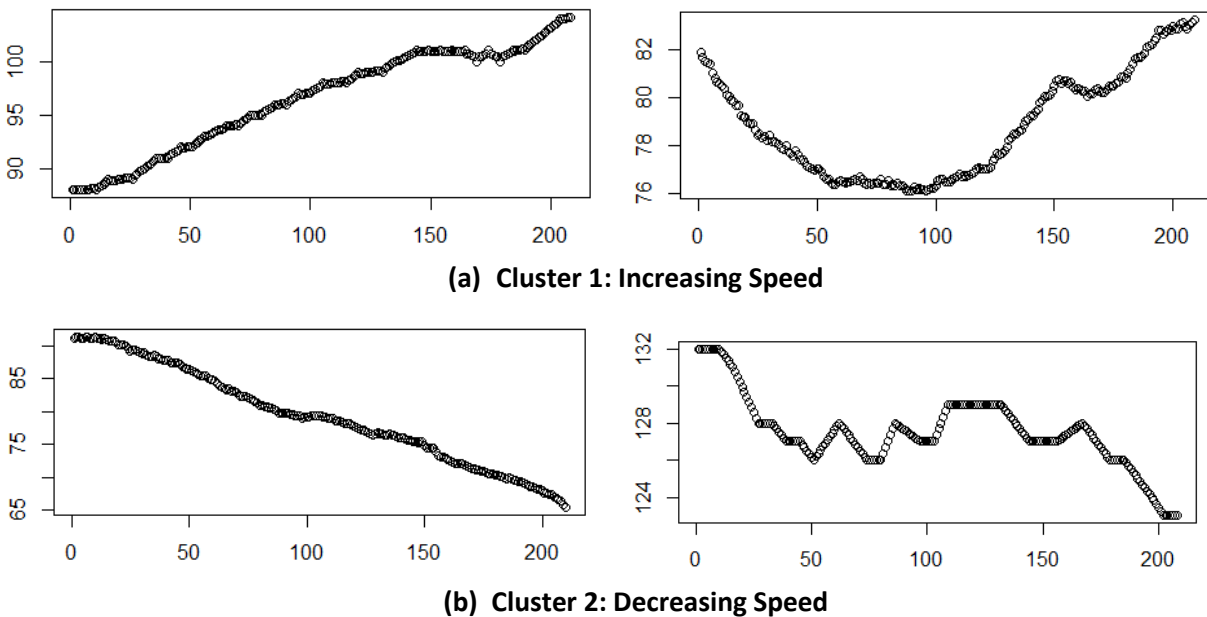
Figure 21. Example of Clustered Speed Data.

Figure 21 shows a similar pattern on the speed data. There are two logical clustered groups for the speed data in increasing and decreasing trends. In Figure 21a (left), the driver increased their speed smoothly over the course of the time series. In Figure 21a (right), the driver first decreased in speed before accelerating in two segments (100–150 and 170–200).

While both are generally increasing, there are semantically different actions taken by these two drivers. Similarly, in the decreasing speed cluster, Figure 21b (left) shows a steady decrease in speed while Figure 21b (right) shows several small variations along the way.

Summary of Findings from Cluster Analysis

Some logical patters were found by performing cluster analysis on both TTC and speed. Although it was possible to separate groups of time series in patterns that are indeed similar, these pattern categorizations still show enough heterogeneity to allow meaningful interpretations.

The TTC analysis also showed that, given how the TTC is defined, discontinuities may occur when the two vehicles following reach the same speed (see Figure 20b), which would add complexity to the analysis. Researchers decided to perform further analyses on the speed data and back-calculate TTC as necessary for interpretation.

Freeway SOC Analysis

As a first step on the formal analysis, researchers investigated what factors associate with the SOC of participants. The SOC was defined as follows. From a filter of 761 events for analysis, researchers broke down all events (each up to 30 seconds) into car-following subevents (fractions of an event when a lead vehicle is identified in the radar track) and non-car-following subevents (fractions where no lead vehicle is present).

Among the non-car-following subevents (i.e., the driver is not following another car), researchers filtered only those that have speed values with no significant changes (i.e., those whose variance in the speed readings did not exceed 4 m^2/s^2). After verifying that these speed profiles did look indeed flat, a subset of 254 subevents were identified. These events were considered samples representative of the SOC for these drivers because they were not following another car and they were not changing their speed. Two subsets were identified from these events: 253 subevents where most variables of interest were available and 77 where it was possible to match with RID elements. The RID provides key information about the context of the driver: speed limits, number of lanes, average width of lanes, etc. Therefore, researchers performed two analyses: 1) an analysis of the subevents that had most variables of interest but were not matched to the RID, and 2) an analysis of the subset of data that had RID matches. These analyses are presented in later sections of this chapter. The next section shows the model specification used for the analyses just described.

Model Specification

The methodology for analysis was mixed-effects model, as described in the prior chapter. This framework allows the inclusion of explanatory factors (i.e., fixed effects) that have been coded from the database and exogenous sources of variability (coded as random effects).

The output of these models directly allows controlling for the variability associated with exogenous factors. Additionally, since the variability from every source of variance is estimated by the model, a comparison between exogenous sources and known explanatory factors is possible in terms of explained variability.

Gaussian Model for SOC

As mentioned, researchers used a mixed model to investigate the variables that associate with variability of SOC in the data. The model parameterizes the mean of the conditional speed as a linear combination of several parameters as shown in Equation 1:

$$SOC_{ijk} = X' \cdot \beta + [\alpha_i + \gamma_{ij}] + \varepsilon_{ijk}$$
Equation 1

Where:

SOC_{ijk}	=	Speed of Choice by the ith driver, j[th] event, k[th] subevent.
X	=	Vector of fixed-effects explanatory factors.
β	=	Vector of fixed-effects coefficients.
α_i	=	Random effect for i[th] participant.
γ_{ij}	=	Random effect for j[th] event from i[th] participant.
ε_{ijk}	=	Random error for k[th] subevent from j[th] event from i[th] participant.

As shown in Equation 1, the model includes fixed and random effects, the former for global explanatory variables, the later for accounting for structures in the data of potentially correlated clusters of data (e.g., all data from one participant may look different as a set, to all data from another participant).

The variance of the conditional mean is estimated as a single parameter independently of the mean. This parameter is simply the variance of the residuals between the raw data and the calibrated model, as shown in Equation 2:

$$V(SOC_{ijk}) = V(\varepsilon_{ijk}) = \sigma_{res}^2$$
Equation 2

Where:

$V(.)$	=	The variance of a random variable.
σ_{res}^2	=	The variance of the model residuals. Other variables as previously defined.

The next section describes the modification of the variance model to account for the expected effect of multiple points per subevent.

Heteroscedasticity in SOC Estimates

The number of free-flowing subevents were not equal for all drivers. In some cases, almost all 30 seconds were free-flowing, whereas in many others, a smaller subset of those segments was free flowing. Since there is no point in including multiple points per subevent that are basically the same number, researchers decided to use the average of such values to have one value per subevent.

However, this decision yielded a data set comprising of averages estimated from different sample sizes. This implies that there is unaccounted heteroscedasticity because it is known that the average of a sample is an efficient estimate of the mean. In other words, it is reasonable to expect that SOC estimates are more accurate when they come from a larger set of points. To account for the implied heteroscedasticity, researchers specified a model with heteroscedastic variance as given in Equation 3:

$$V(\varepsilon_{ijk}) = \sigma_0^2 \cdot n_{ijk}^{-1}$$
<div align="right">**Equation 3**</div>

Where:

σ_0^2	=	The residual variance a datum is expected to have (estimated).
n_{ijk}	=	The number of SOC readings available from each subevent.

Stepwise Model Selection

The components of the vector of predictors are individual variables available in the data set that are incorporated one at a time. To avoid overfitting the model (i.e., trying to estimate more information than is defensible, given the data), researchers informed the model selection with the Akaike Information Criterion (AIC), which is a metric of quality of information, given a statistical model and a data set. The AIC tends to increase when removing a variable contributing a large amount of information from the model fixed effect and when including a variable not contributing much information in addition to the variables already in the model.

Free-flow Speed Choice without Posted Speed Limit Available

After some exploratory analysis of the sites with RID matches, it became apparent that the PSL was probably the most critical variable from the RID. However, as described in the previous section, only 77 subevents were available with this piece of information. Researchers decided to analyze the larger subset where RID data were not available first as described in this subsection.

Researchers performed model selection in the first subset of 251 subevents to investigate what non-age-related factors explain SOC better. This subset of subevents came from originally 150 different drivers and 182 events.

Table 9 shows the estimates from the most parsimonious model resulting from this exercise. The baseline for the model is an average of 53.24 mph for drivers with no traffic violations, driving at LOS A1 at 2-way divided highways with signals in the state of Florida. The residual variability of 5.39 mph indicates that the model can explain the SOC variability within that threshold, after accounting for everything else explicitly accounted for in the estimates. Among those factors explicitly accounted for in the model are the following:

- Decreasing SOC with increasing LOS (a reduction of 5.9 mph for LOS B with respect to LOS A1, everything else equal. The 3.00 mph reduction for LOS A2 was not found statistically significant).
- An increased SOC at freeways and highways with no signals (16.6 mph faster than highways with signals, everything else equal).
- A decreased speed of choice (5.2 mph slower) among drivers from Washington, compared to drivers from Florida, other things equal.
- A decreased SOC at one-way traffic facilities (7.3 mph slower), compared with two-way facilities, other things equal.
- A higher SOC among drivers with at least one traffic violation (3.3 mph faster) compared to drivers with no traffic violations, other things equal.

Table 9. Initial SOC Model without Age Characteristics and PSL (n=254).

Explanatory Variable	Estimate	Std. Error	DF	t-value	p-value	Signif.
Baseline	53.24 mph	3.094 mph	148	17.2105	< 0.0001	***
LOS A2	−3.00 mph	2.558 mph	28	−1.1728	0.2508	
LOS B	−5.88 mph	2.602 mph	28	−2.2578	0.0319	*
Rural Freeway / Highway wo. Signals	+16.56 mph	2.822 mph	28	5.8678	< 0.0001	***
Urban highways	+2.48 mph	10.522 mph	28	0.2359	0.8152	
Washington State	−5.20 mph	1.429 mph	148	−3.6402	0.0004	***
One-way Traffic	−7.25 mph	3.504 mph	28	−2.0702	0.0478	*
At least 1 traffic violation	+3.31 mph	1.404 mph	148	2.3561	0.0198	*
Unaccounted Variability among Participants	±0.001 mph		Notes: Baseline level is defined as drivers with no traffic violations, LOS A1 at 2-way Divided Signalized Rural Highways in Florida.			
Unaccounted Variability among Events per participant	±8.298 mph					
Residual Variability	±5.389 mph					

Significance levels are as follows:
* = Significant at the 0.05 level
** = Significant at the 0.01 level
*** = Significant at the 0.001 level

The model results also indicate three levels of unaccounted variability (i.e., reported as random effects variance) that are of interest to this research:

- The unaccounted variability among participants is the amount of variance in the data that remains unaccounted among participants after all other sources of variability in the model have been discounted (including residuals and variability between events). As indicated in Table 9, virtually no differences between participants remain after accounting for everything else in the model.
- Similarly, ±8.298 mph is the variability resulting from differences between events of a single driver, everything else equal. This is a large amount of unaccounted variability without any more information than the observation that different non-following driving events from the same driver tend to have very different SOCs. Researchers interpret this result as an indication that important sources of variation are absent from this model, such as driver characteristics and driving context elements (e.g., PSL).
- Finally, the residual variability of the model is ±5.389 mph indicating the average expected error of predictions after discounting everything else in the model (including the two estimates of unaccounted variability discussed in the prior two points).

Next, researchers performed a second round of model selection using the model in Table 9 as a starting point but now including age groups and other driver-specific factors (e.g., distraction types and levels, vision conditions self-declared by participants) as potential explanatory variables. The resulting model from this exercise (shown in Table 10) clearly has an improved ability to account for variability in the data.

Although distraction types similar to those defined in past research [70] were used in the model selection process, these variables did not contribute improving the model information quality and were not included in the most parsimonious model shown in Table 10.

Table 10. SOC Model Including Age Characteristics without PSL (n=254).

Explanatory Variable	Estimate	Std. Error	DF	t-value	p-value	Signif.
Baseline	51.81 mph	2.783 mph	145	18.6187	< 0.0001	***
LOS A2	−2.13 mph	2.040 mph	27	−1.0454	0.3051	
LOS B	−5.29 mph	2.082 mph	27	−2.5428	0.017	*
Rural Freeway / Highway wo. Signals	+17.73 mph	2.276 mph	27	7.7891	< 0.0001	***
Urban highways	+3.75 mph	8.775 mph	27	0.4271	0.6727	
Washington State	−5.61 mph	1.162 mph	145	−4.8286	< 0.0001	***
One-way Traffic	−5.52 mph	2.849 mph	27	−1.9390	0.063	#
Select Vision Conditions	−5.80 mph	1.236 mph	145	−4.6947	< 0.0001	***
Driver 24 years of age or younger	+5.71 mph	1.432 mph	145	3.9859	0.0001	***
Driver 45 years of age or older	+3.00 mph	1.450 mph	145	2.0678	0.0404	*
Remaining Variability among Participants	±0.001 mph		Notes: Baseline level is defined as drivers without select visual conditions between 25 and 44 years of age, LOS A1 at 2-way Divided Signalized Rural Highways in Florida.			
Remaining Variability among Events per participant	±6.493 mph					
Residual Variability	±4.969 mph					

Significance levels are as follows:
      # = Significant at the 0.10 level
      * = Significant at the 0.05 level
      ** = Significant at the 0.01 level
      *** = Significant at the 0.001 level

As it can be seen in Table 10, the model includes a statistically significant effect for drivers with a set of vision conditions. Researchers determined this list of conditions to include the following:

- Objects far away are blurry when not wearing corrective lenses (e.g., nearsighted).
- Poor night vision.
- Reading glasses needed.
- Reading glasses needed; Glaucoma; Poor night vision.

Researchers determined this list of conditions by identifying the extreme partial residuals from an intermediate model between those in Table 9 and Table 10 that included the full list of conditions available from the data set. The most parsimonious model resulted from aggregating these conditions under the variable shown in Table 10 since each of these conditions had a very similar deviation in their partial residuals. The three mutually exclusive sets of age groups in Table 10 were determined in a similar way as just described for the select vision conditions. Next appears a brief discussion of the results in the expanded model (Table 10).

The baseline for the model in Table 10 is an average SOC of 51.8 mph for drivers of ages 25 to 44 years, with no visual adverse conditions, driving at LOS A1 on 2-way divided highways with signals in the state of Florida. This baseline SOC is very similar to that from the initial model (53.2 mph). The residual variability in the expanded model reduced to 4.97 mph in the expanded model from to 5.39 mph in the initial model. This reduction corresponds to a 15 percent decrease in residual variance, which means that the model in Table 10 produces predictions 15 percent less imprecise compared to Table 9. Among the factors explicitly accounted for in the expanded model are the following:

- Similar to the initial model, a decrease in SOC relates with an increasing LOS (a reduction of 5.29 mph for LOS B with respect to LOS A1, everything else equal. Also similar to the initial model, the 2.12 mph reduction for LOS A2 was not found statistically significant).
- An increased SOC at freeways and highways with no signals (17.7 mph faster than highways with signals, everything else equal). This estimate is very comparable to the 16.6 mph increase in the initial model.
- Also, like the initial model, a decreased SOC (5.6 mph slower) among drivers from Washington, compared to drivers from Florida, other things equal.
- A decreased SOC at one-way traffic facilities (7.3 mph slower), compared with two-way facilities, other things equal.
- Drivers with vision conditions as explained above, chose speeds significantly lower, other things equal (a reduction of 5.80 mph in SOC for this group of drivers).
- Both younger drivers (17 to 24 years of age) and older drivers (ages 45 to 80 years) tend to choose faster speeds than drivers in the group of reference (between 25 and 44 years of age). On average, younger drivers chose speeds 5.71 mph faster than the group of reference; on average, older drivers chose speeds 3.00 mph faster than the group of reference.
- Interestingly, the higher SOC by drivers with at least one traffic violation compared to no traffic violations in the initial model (3.3 mph faster) was not found significant in the expanded model that included driver age and vision conditions. Although this result may indicate confoundedness between age, vision conditions, and history of traffic violations, researchers assess that it most likely can be explained by a limited statistical power in the model, given the sample size and the amount of unexplained variation. Simply put, a sufficiently large sample size would likely find all these effects statistically significant if this is indeed the explanation.

Similar to the initial model, the expanded model also indicates three levels of unaccounted variability of interest to this research:

- Similar to the initial model, the unaccounted variability among participants remains virtually not existent after accounting for everything else in the model and given the residual variability of the model.
- In contrast with the ±8.298 mph remaining variability between events for a single driver in Table 9, the expanded model indicates a reduced amount of variability between events (±6.493 mph, per Table 10). This is significant reduction of unaccounted variability between events from the same driver in explaining their speeds of choice. Researchers interpret this reduction as direct effect of incorporating the age and vision conditions among the explanatory variables. These factors appear to be additional key sources of variation are absent from this model. Regardless, since the expanded model does not include PSL, researchers anticipate that the amount of

remaining variability between events can still potentially reduce significantly for the subset of data where PSL is available.

As the following step in the analysis, researchers fitted models to the subset of subevents that could be matched to the corresponding RID segments.

Free-flow Speed Choice with Posted Speed Limit Available

As mentioned earlier in this chapter, the PSL could be matched to only 77 subevents from 45 different events and 46 participants. Like the modeling described in previous section, researchers fitted an initial full model without including driver characteristics first and performed stepwise model selection. This exercise resulted in a model with only three explanatory variables: traffic density, PSL, and functional class as shown in Table 11.

Table 11. Initial SOC Model with PSL (n=77).

Explanatory Variable	Estimate	Std. Error	DF	t-value	p-value	Signif.
Baseline	48.69 mph	6.047 mph	42	8.0516	< 0.0001	***
LOS A2	−1.81 mph	2.542 mph	42	−0.7101	0.4816	
PSL	+0.30 mph	0.087 mph	30	3.4236	< 0.0001	***
Minor Collector	−10.22 mph	2.488 mph	42	−4.1092	0.0002	***
Variability among Participants	±0.001 mph		Notes: Baseline level is defined as drivers of all ages on Arterials and Major Collector Freeways and Highways at LOS A1 in Florida and Washington.			
Variability among Events	±4.745 mph					
Residual Variability	±3.926 mph					

Significance levels are as follows:
> # = Significant at the 0.10 level
> * = Significant at the 0.05 level
> ** = Significant at the 0.01 level
> *** = Significant at the 0.001 level

The most salient features of this model are 1) the size of the residual variability (±3.93 mph), which is significantly smaller than the best model without PSL (±4.96 mph), which clearly indicates how important PSL is as an explanatory variable for SOC; and 2) the larger standard error for the model estimates, which implies a reduced statistical power for models on this reduced sample of sites.

Next, researchers performed another round of model selection now considering driver characteristics as potential explanatory variables. The resulting model is very similar to the initial model with PSL and is shown in Table 12. As it can be seen, only one additional term entered the model and its significance and impact on unexplained variability are minimal.

Conspicuously, factors identified on the models in Table 9 and Table 10 as influential on SOC were not found significant contributors on the PSL models (Table 11 and Table 12). This is likely the result of the two salient features of these later models, as described above: reduced statistical power due to reduced sample size and the inclusion of a factor (PSL) that is more influential than those on earlier models.

Regardless, of these salient differences, the variables in common to both sets of models are consistent with each other (i.e., estimates with similar magnitudes and signs). The next section presents a performance comparison between the expanded models described in this section.

Table 12. Expanded SOC Model with PSL (n=77).

Fixed Effect Level	Estimate	Std. Error	DF	t-value	p-value	Signif.
Baseline	47.64 mph	5.962 mph	41	−0.8773	< 0.0001	***
LOS A2	−2.16 mph	2.458 mph	41	−0.8773	0.3854	
PSL	+0.29 mph	0.086 mph	30	3.398567	0.0019	**
Minor Collector	−10.64 mph	2.444 mph	41	−4.355129	0.0001	***
Number of Years Driving	+0.07 mph	0.038 mph	41	1.783977	0.0818	#
Variability among Participants	±0.001 mph		Notes: Baseline level is defined as drivers of all ages on Arterials and Major Collector Freeways and Highways at LOS A1 in Florida and Washington.			
Variability among Events	±4.503 mph					
Residual Variability	±3.957 mph					
Significance levels are as follows: # = Significant at the 0.10 level * = Significant at the 0.05 level ** = Significant at the 0.01 level *** = Significant at the 0.001 level						

The association with regulatory speed limit appears straightforward: higher PSL implies naturally higher SOC. However, SOC tend to be larger than PSL at lower PSLs. SOC and PSL tend to agree at larger PSL as shown in Figure 22.

The blue dashed line is the 1:1 correspondence. The SOC tend to be above the blue dashed line at lower speeds but it converges with speed limit at higher speeds. The symbols and segment lines indicate data from different participants. The same driver would prefer a different free-flow speed for different PSL. There are linear trends drawn in Figure 22 to indicate groups of data from a common participant. Three out of four of these trends clearly indicate an increasing SOC with increasing PSL.
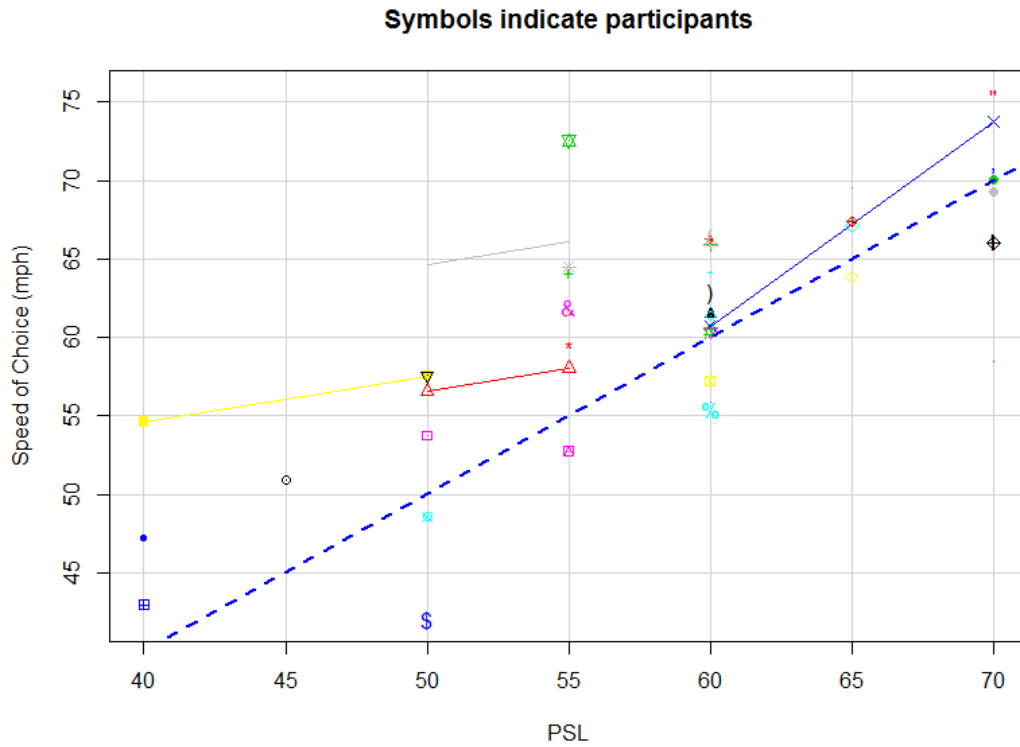
Figure 22. Speed Choice vs. PSL.

Next, researchers prepared a comparative assessment of performance for both SOC expanded models developed (i.e., with PSL and without PSL). Figure 23 displays this graphical assessment. Predictions were estimated for every data point in the data set with the complete set of predictors from each model readily available.

Again, the color-coding indicates groups of data points for different participants. The first two rows in this figure show the model fitted values. That is, the model's estimates for all subevents used to fit the SOC models; the lower two rows show how the SOC predictions compare to actual speeds at events that were not used in the SOC models (i.e., events that did not meet the definition of non-following, free-flowing conditions).

It is not surprising that sections (c) and (d) of Figure 23 clearly show a worse performance than sections (a) and (b), respectively, not only because these are predictions beyond the data used to fill the models, but also because the data set for which the predictions were cast include car-following events as well. Regardless of the anticipated lesser performance, the relative good correspondence between the predictions and the actual speeds is noteworthy.

Summary of Findings

Researchers had two objectives to analyze the SOC using free-flow subevents. One was to identify the most influential factors on SOC. The other was to investigate how accurately the SOC could be predicted to use these predictions in subsequent analyses.

Influential Factors on SOC

The analysis intended to explore what factors are most influential to drivers' SOC. Researchers attempted to merge the time-series data with the RID to include variables known to be influential in the analyses: functional class and PSL. Unfortunately, only a small subset of the time-series data could be matched to the RID. For this reason, researchers performed two analyses: a) one for the larger subset of data that could not be matched to the RID (thus not including PSL as a potential predictor); and b) one for the smaller subset that could be matched to the RID allowing PSL and functional class be included among the potential explanatory variables.
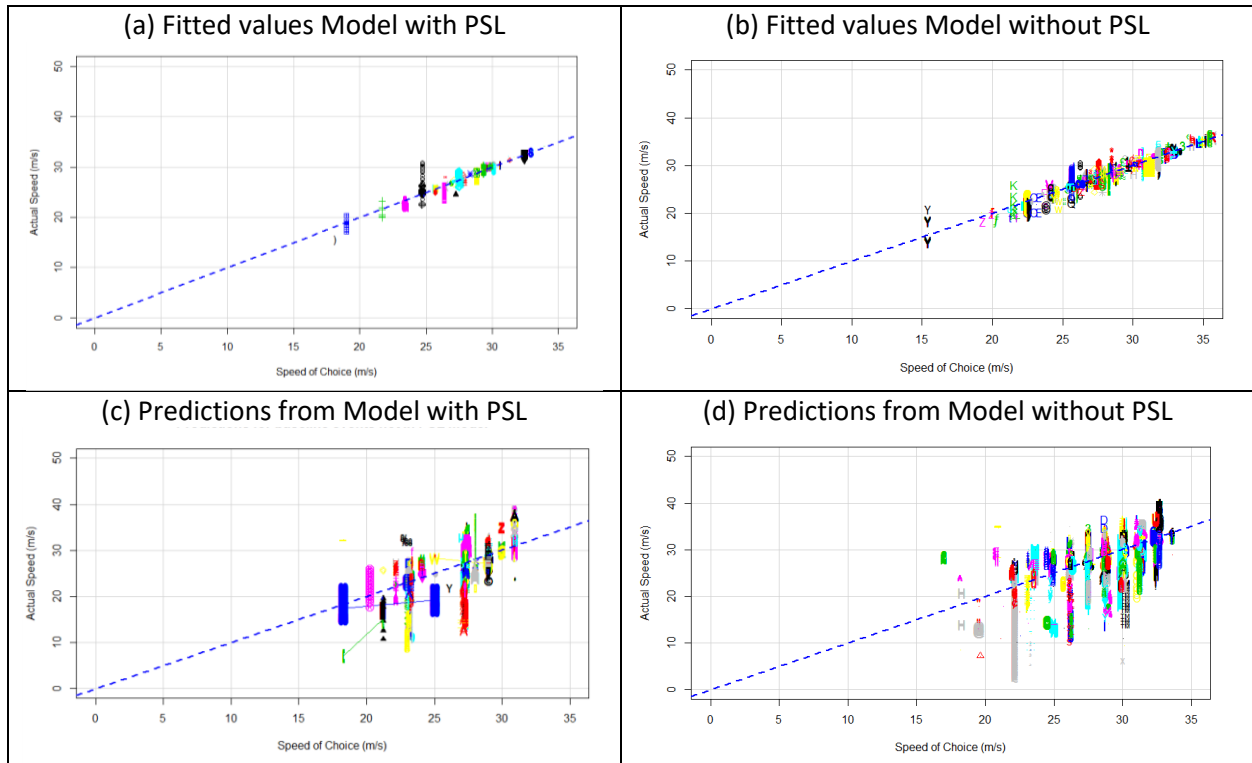
Figure 23. Comparative Performance of SOC Models.

From the analysis of SOC in the larger subset (where PSL was not available), researchers identified six factors that better explain the variability in SOC among drivers represented in this subset. Four of these factors are not related to the driver's characteristics:

- The LOS of the traffic flow relates inversely with SOC.
- Facilities of higher order (i.e., freeways and highways with uninterrupted flow) tend to have larger SOC, most likely because the effect of PSL is confounded with these facilities.
- Drivers in the state of Washington had lower SOC (but this could also indicate that the events from that state occurred on roads with lower PSL).
- One-way facilities also had lower SOC (but confoundedness with PSL could be the underlying cause, like the prior point).

Two of these influential factors did relate to driver characteristics:

- Drivers who are nearsighted, with poor night vision, in need of reading glasses or glaucoma tended to choose lower SOCs than drivers without those conditions.
- Drivers 24 years of age or younger and drivers 45 years of age or older tended to choose higher SOCs than drivers of other ages.

Researchers verified that including the last two terms in the model had a significant impact in the performance of the model: a reduction of the amount of residual variance (i.e., unexplained variability) and a reduction on the variability remaining among events from the same participant after accounting for everything else in the model. Researchers conclude that visual conditions and driver age are significant influential variables in drivers' SOC.

From the analysis of SOC in the smaller subset (where PSL was available from matched RID segments), researchers identified four factors that better explain the variability in SOC among drivers represented in this subset. Three of these factors are not related to the driver's characteristics:

- The PSL was found the most influential variable on SOC. The relationship was found to be of direct proportion (i.e., increasing SOC with increasing PSL).
- The LOS of the traffic flow relates inversely with SOC (but was not found statistically significant).
- Facilities of higher order (i.e., arterials and major collectors) tend to have higher SOCs. This, after accounting for the impact of PSL, suggests that the cause is other characteristics associated with the functional class. Researchers speculate that driver environment and other geometric cues could be the reason for this association with functional class.

Differences between states and between one-way and two-way facilities were not found meaningful after accounting for PSL. This is consistent with the hypothesis that those effects were masking the influence of PSL in the model where PSL was not available.

The only influential factor related to driver characteristics is discussed next—the number of years of driving associated with an increased SOC in the model that included PSL. This finding implies a monotonic relationship between SOC and driving experience. In contrast, the model without PSL suggested a U-shaped relationship: younger and older drivers tend to choose faster speeds. Researchers speculates that the reason for this contrast is the small sample size to which the model with PSL was fitted. A larger sample of data with PSL more representative of younger drivers could potentially yield results that are compatible with the findings from the model with no PSL.

Researchers verified that including the last terms in the PSL model almost had no impact in the performance of the PSL model. Researchers again speculate that sample size may be the reason for significantly lower statistical power, compared to the model without PSL. Researchers conclude from this analysis that PSL, geometry, and traffic characteristics are the most influential factors on SOC, and that driver characteristics may improve performance prediction marginally thereafter. However, the analysis of subevents that did not include PSL suggests that enough statistical power to detect the influence of driver characteristics on SOC probably requires tripling the sample size (from 77 to 254). More research is needed that analyzes SOC for a larger sample of data that includes PSL and other geometric characteristics from the RID.

SOC Prediction Performance

Regardless of the influential factors identified in the previous section, researchers verified graphically that the prediction of SOCs from both models (with and without PSL) seem to be reasonable approximations of the actuals speeds of events. As can be seen in Figure 23(c) and (d), the predictions for events with known PSL tend to be more accurate and precise, whereas the predictions from events without known PSL tend to be more dispersed and biased toward smaller values (i.e., a tendency to under-predict).

Freeway Driver Car-Following Performance

After finalizing the SOC analyses, researchers filtered car-following subevents for analysis. A car-following subevents is defined as the portion of driving that was coded as having a lead vehicle present. The next sections describe an exploratory analysis of the data and the selection and refinement of a dynamic model specification for the formal analysis.

Exploratory Analysis of Driver Car-Following Behavior

Researchers started by examining at the relationships between dynamic variables to gain insights about the characteristics of those relationships and how to best account for them in the formal dynamic model analysis. Figure 24 shows the relationship between pairs of time series from a typical car-following subevent in the data set.
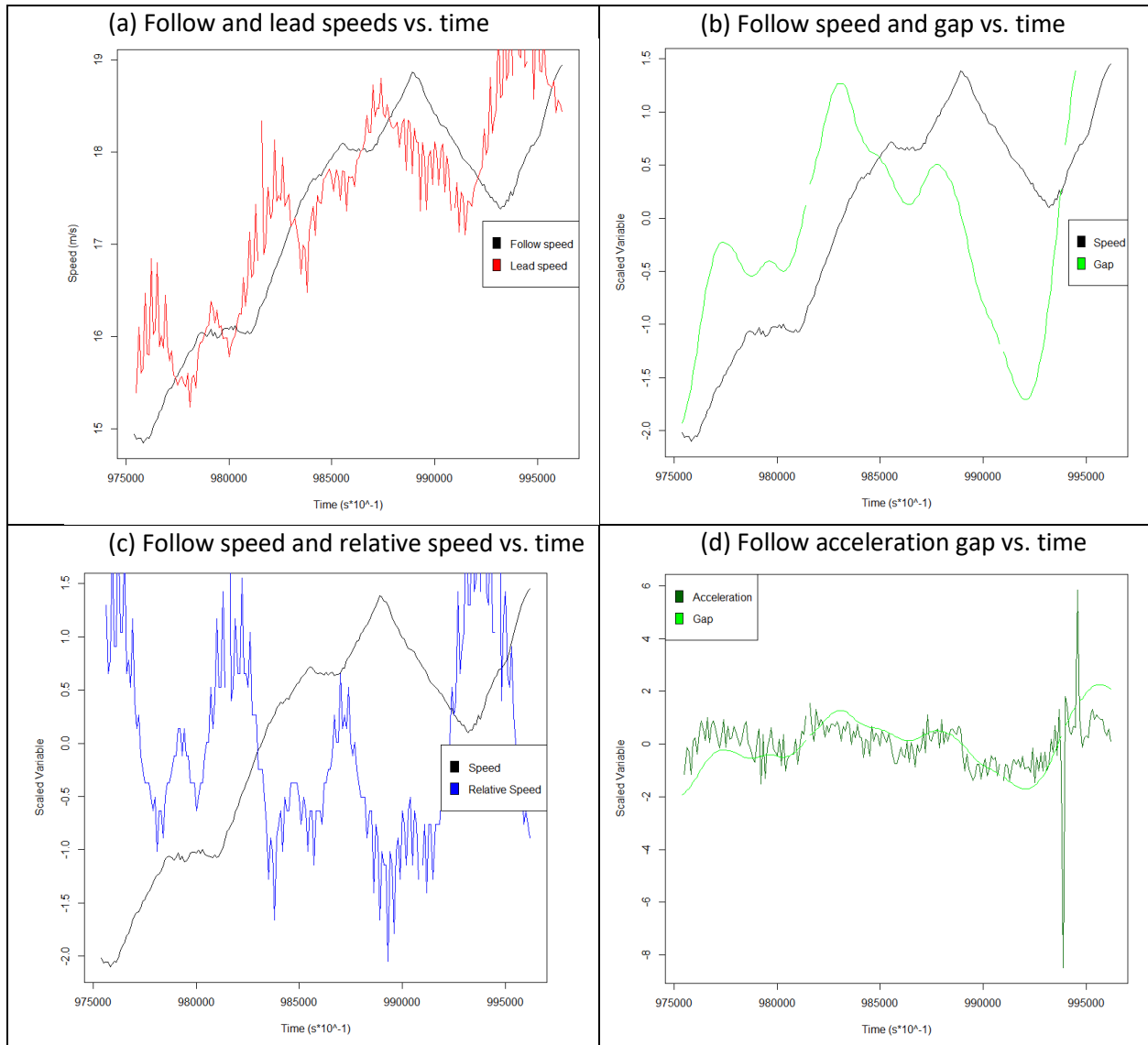
Figure 24. Comparison of Various Pairs of Time Series from a Single Car-Following Subevent.

Figure 24a shows there seems to be a moderate relationship of codependence between follow and lead speed (i.e., clearly codependent but with some degree of variability that seems marginal to the relationship).

Figure 24b shows a cleaner relationship between the two time series (follow speed and gap) and a very clear lead of the gap to the speed (i.e., the gap reaches peak and valley points before the speed does). In a causal analysis framework, this lag suggests that the gap could be an antecedent to the speed (the consequent).

The relationship shown in Figure 24c is the weakest shown in Figure 24. In other words, the two time series (follow speed and relative speed) do not seem to follow each other. However, the relative speed time series seem to oscillate around an approximately flat line, which suggests that there is a tendency

to maintain a constant relative speed and that corrections to deviations tend to follow quickly thereafter.

Finally, Figure 24d shows a rather strong association between follow acceleration and gap. In contrast with Figure 24b, the acceleration seems to have a slight lead on the gap, which indicates that the gap is a consequent to the acceleration. This is expected just from the kinematic relationship between the variables, but both Figure 24b and Figure 24d stress the dynamic nature of the relationships, in some instances leading and in some others lagging.
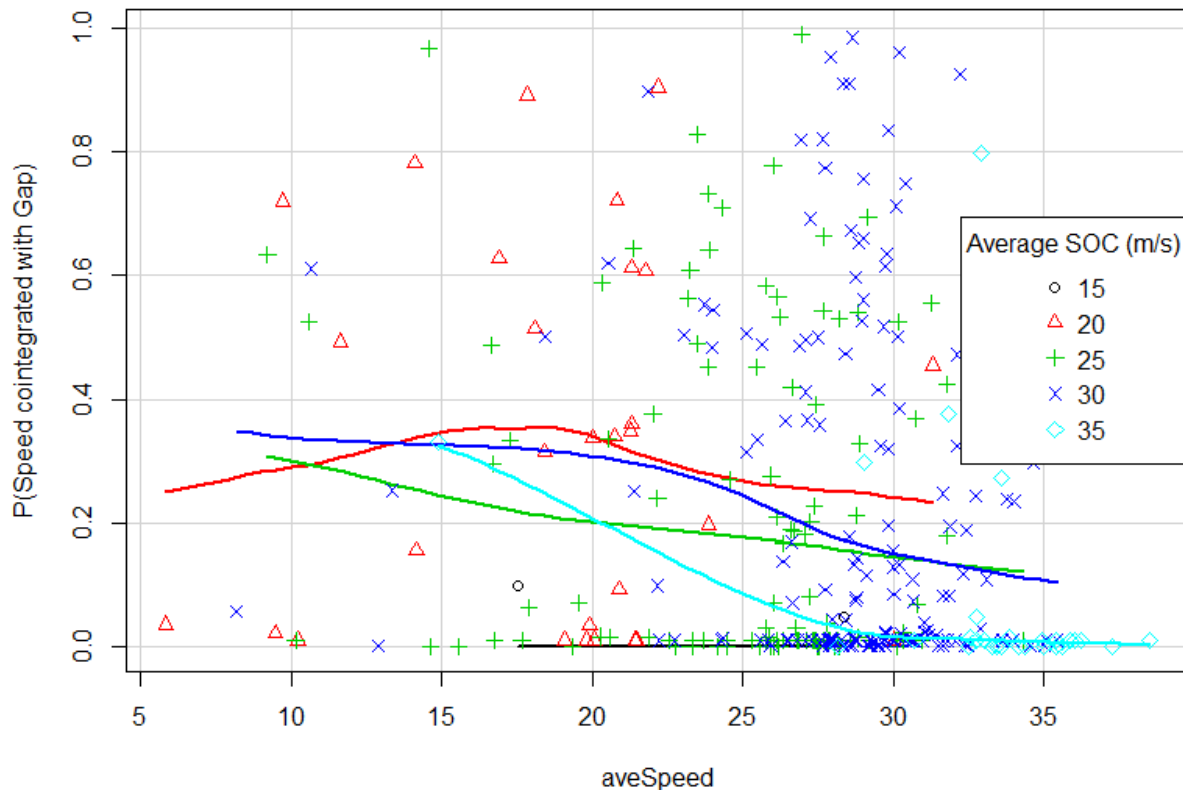
Figure 25. Probability of Speed and Gap Cointegration by Average Speed and SOC.

Given some of the codependence suggested by the visual inspection of the data, researchers decided to run the Granger's causality test, which is a statistical test on the scope and degree of the lead/lag relationship for the car-following speed and gap time series in a subset of randomly selected car-following subevents. Researchers observed results that suggested a causality relationship for lags up to 4 seconds.

Researchers performed an additional test in the relationship between the car-following speed and the gap on the complete data set of car-following subevents. The test performed is ADF, and it yields the

probability of two series being cointegrated [83]. The degree of cointegration indicates how much the two series tend to vary together.

Figure 25 shows an inverse relationship between the probability of cointegration and car-following speed. However, a relationship between P(cointegration) and average SOC is not apparent.
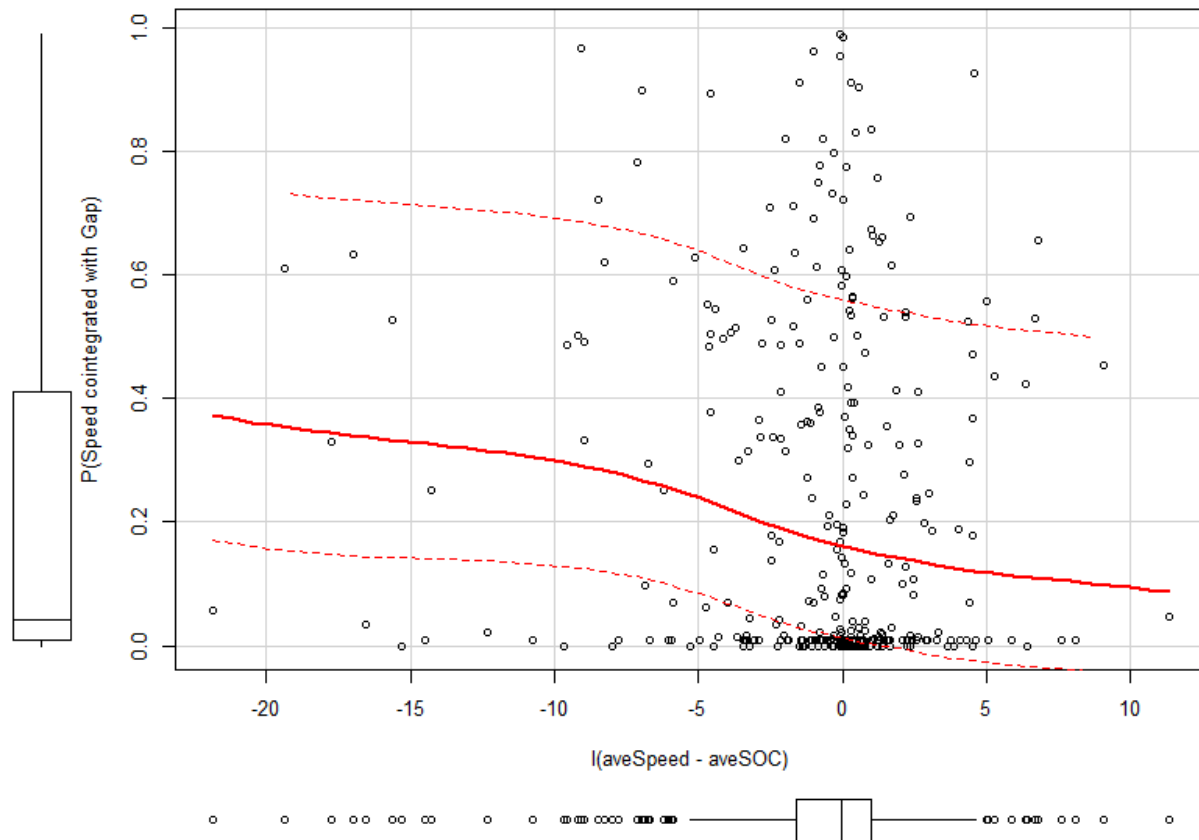
Figure 26. Probability of Speed and Gap Cointegration by Speed Differential from SOC.

The mild trend in Figure 25 toward a reduced likelihood of cointegration may be due to reduced variability expected at higher speeds, which researchers consider likely given that the maximum length of a subevent is 30 s; such a small window of time is less likely to capture the variability of a trip at a higher speed.

Similar to the prior trend, Figure 26 shows a clearer inverse relation between the probability of cointegration and the speed differential between the car-following speed and the SOC. The probability of cointegration seems larger when the differential is negative and large in magnitude. This relationship is potentially useful in the sense that may indicate that drivers could be adjusting their operating speed and gap more actively (thus, more cointegration would be expected between the series) when they are following a lead vehicle but at a speed lower than the SOC.

The higher the ADF probability is, the larger the commonality between the time series, which does not necessarily indicate a causal relationship because there is no direction associated with the probability given. Researchers anticipated treating the underlying assumption of causality when treating leads and lags observed in this exploratory analysis explicitly in a dynamic model, as the following section describes in more detail.

Dynamic Mixed Model for a Kinematic Variable

For the analysis of car-following events, researchers chose to specify a dynamic mixed-effects model for kinematic variables time series. The general form researchers initially decided to use on car-following speed is shown in Equation 4:

$$Speed_{ijk}(t) = X' \cdot \beta + dm_{mixed}\left(Z'_{ijk}(t) \cdot \theta, \varepsilon_{ijk}(t)\right)$$
Equation 4

Where:

$Speed_{ijk}(t)$	=	Instantaneous Speed for the ith driver, j^{th} event, k^{th} subevent, at epoch t.
X	=	Vector of fixed-effects explanatory factors.
β	=	Vector of fixed-effects coefficients.
$Z_{ijk}(t)$	=	Vector of dynamic explanatory factors (fixed and random effects) at epoch t for k^{th} subevent in j^{th} event from i^{th} participant.
θ	=	Vector of dynamic coefficients (fixed and random effects).
$dm_{mixed}(.)$	=	Dynamic model (i.e., time-dependent) including both fixed and random effects.
$\varepsilon_{ijk}(.)$	=	Correlation structure of the residual errors.

One important assumption underlying Equation 4 is that of a stationary process, meaning that the response variable varies dynamically around a long-term average. After a few preliminary runs, researchers determined that this assumption was clearly violated. A time series under this situation is known as an integrated time series. In the arena of time-series analysis, this can be remedied by differentiating the time series recurrently, as necessary, until a stationary time series is obtained for analysis. The analysis from such time series can then be integrated recurrently to obtain estimates and answer questions about the original variable of interest.

Researchers then decided to fit the subsequent models on the acceleration series and determined that the stationary condition was satisfied at the first-derivative level. The next section provides more details about the specification of the acceleration models.

Acceleration Model Specification

Given that the car-following speed is an integrated process, researchers developed statistical dynamic models for the car-following acceleration as the response variable. The general formulation of these models is:

$$Accel_{i,j,k}(t) = X' \cdot \beta + dm_{mixed}\left(Z'_{i,j,k}(t) \cdot \theta, \varepsilon_{ijk}(t)\right)$$
Equation 5

Where:

$Accel_{ijkl}(t)$ =   Instantaneous Acceleration for the ith driver, j[th] event, k[th] subevent, and epoch t. Other variables as previously defined.

As mentioned earlier, the dynamic submodel is defined in a flexible way, so that the dynamic variables may enter the model either as fixed or random effects, depending on the quality of information (per the AIC) and the appropriateness of specifying them as random effects (per the result of a Hausman test). Given the results of the exploratory analysis, researchers determined to define the dynamic submodel to include a monotonic function of a family of lags in the car-following gap time series as shown in Equation 6:

$$dm_{mixed}\left(\boldsymbol{Z'_{ijkl}}(t)\cdot\boldsymbol{\theta}, \varepsilon_{ijk}(t)\right) = f\left[L, gap_{ijk}(t), age, SOC, \boldsymbol{\theta}\right] + \varepsilon_{ijk}(t) \qquad \textbf{Equation 6}$$

Where:

$f[.]$ =   Function specification for the dynamic model.

L =   Lag operator for time series.

$gap_{ijk}(t)$ =   Distance gap between following and lead vehicle at epoch t for k[th] subevent in j[th] event from the ith participant.

age =   Age class of participant, defined as the center of the class originally coded in NDS. Other variables as previously defined.

The definition in Equation 6 specifies a function with parameters that account for gap and age, given that the intent of this research is to review driver characteristics and performance, as well as the differences in performance by age documented in the literature review. When specifying the dynamic model, researchers intend to test the hypothesis that differences may exist in car-following behavior by age group.

Hierarchical Structure and Monotonic Decay in the Dynamic Submodels

Various examinations of multiple car-following subevents strongly suggested that the relationship between car-following gap, speed and acceleration is stronger at short range and decays quickly with increasing gaps. After exploring different approaches to account for this feature appropriately (i.e., strong relationship at short gaps, monotonic decay for increasing gaps), researchers selected a logarithmic transformation as the best performing in preliminary tests.

Additional to the logarithmic decaying influence of gap on acceleration, researchers specified a hierarchical structure similar to that in Equation 1 but also including hierarchical structure in the influence dynamic relationship as well.

As shown in Equation 5 and Equation 6, the vector of coefficients for the dynamic model is estimated as the weights from a linear combination of the components vector **Z** comprised of dynamic variables. Equation 7 shows the composition of vector **Z**.

$$Z_{ijkt} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ L^{\omega_1} \cdot \ln gap_{ijk}(t) \\ L^{\omega_2} \cdot \ln gap_{ijk}(t) \\ L^{\omega_3} \cdot \ln gap_{ijk}(t) \\ L^{\omega_4} \cdot \ln gap_{ijk}(t) \\ L^{\omega_1} \cdot rel\ speed_{ijk}(t) \\ L^{\omega_2} \cdot rel\ speed_{ijk}(t) \\ L^{\omega_3} \cdot rel\ speed_{ijk}(t) \\ L^{\omega_4} \cdot rel\ speed_{ijk}(t) \\ L^{\omega_1} \cdot SOC\ diff_{ijk}(t) \\ L^{\omega_2} \cdot SOC\ diff_{ijk}(t) \\ L^{\omega_3} \cdot SOC\ diff_{ijk}(t) \\ L^{\omega_4} \cdot SOC\ diff_{ijk}(t) \end{bmatrix}$$

Equation 7

Where:

ω_1	=	Fixed superscript for lag operator in *age class* 1 submodel.
ω_2	=	Fixed superscript for lag operator in *age class* 2 submodel.
ω_3	=	Fixed superscript for lag operator in *age class* 3 submodel.
ω_4	=	Fixed superscript for lag operator in *age class* 4 submodel.
$rel\ speed_{ijk}(t)$	=	Relative speed (lead speed minus follow speed) at epoch t, for i^{th} participant, at j^{th} event, at k^{th} subevent.
$SOC\ diff_{ijk}(t)$	=	SOC differential (SOC minus follows speed) at epoch t, for i^{th} participant, at j^{th} event, at k^{th} subevent. All other variables as previously defined.

Only one of the three gap terms and one of the relative speed terms in the vector of dynamic variables are applicable for a given combination of indices i, j, and k (i.e., mutually exclusive submodels). Equation 8 shows the set of coefficients that account for this proposed hierarchical structure.

$$\theta = \begin{bmatrix} \alpha_i \\ \gamma_{ij} \\ \tau_{ijk} \\ \varphi_{age\ class\ 1} \\ \varphi_{age\ class\ 2} \\ \varphi_{age\ class\ 3} \\ \varphi_{age\ class\ 4} \\ \rho_{age\ class\ 1} \\ \rho_{age\ class\ 2} \\ \rho_{age\ class\ 3} \\ \rho_{age\ class\ 4} \\ \zeta_{age\ class\ 1} \\ \zeta_{age\ class\ 2} \\ \zeta_{age\ class\ 3} \\ \zeta_{age\ class\ 4} \end{bmatrix}$$

Equation 8

Where:

α_i	=	Random adjustment to the acceleration baseline for the ith participant.
γ_{ij}	=	Random adjustment to the acceleration baseline for j^{th} event from the ith participant.
τ_{ijk}	=	Random adjustment to the acceleration baseline for k^{th} subevent in j^{th} event from the ith participant.
$\varphi_{age\ class\ 1}$	=	Gap coefficient for age class 1 dynamic submodel.
$\varphi_{age\ class\ 2}$	=	Gap coefficient for age class 2 dynamic submodel.
$\varphi_{age\ class\ 3}$	=	Gap coefficient for age class 3 dynamic submodel.
$\varphi_{age\ class\ 4}$	=	Gap coefficient for age class 4 dynamic submodel.
$\rho_{age\ class\ 1}$	=	Relative speed coefficient for age class 1 dynamic submodel.
$\rho_{age\ class\ 3}$	=	Relative speed coefficient for age class 2 dynamic submodel.
$\rho_{age\ class\ 3}$	=	Relative speed coefficient for age class 3 dynamic submodel.
$\rho_{age\ class\ 4}$	=	Relative speed coefficient for age class 4 dynamic submodel.
$\zeta_{age\ class\ 1}$	=	SOC differential coefficient for age class 1 dynamic submodel.
$\zeta_{age\ class\ 2}$	=	SOC differential coefficient for age class 2 dynamic submodel.
$\zeta_{age\ class\ 3}$	=	SOC differential coefficient for age class 3 dynamic submodel.
$\zeta_{age\ class\ 4}$	=	SOC differential coefficient for age class 4 dynamic submodel. All other variables as previously defined.

Error Structure Submodel

When handling time series data, it is very important to consider explicitly the likely codependence between observations close in time. This need is more critical for situations of higher granularity in the time scale as is the case in this study.

The mixed-effects framework proposed by Pinheiro and Bates [*84*] is compatible with and allows the implementation of time series methods to account for error correlation structures. The general modeling structure permits to account for three types of data features explicitly: 1) variables treated as fixed effects, which are expected to have global effects that are not time-dependent (e.g., facility type, number of lanes, PSL, age groups); 2) variables treated as random effects, which can account for clusters or hierarchical structures in the data; and 3) specific types (i.e., structures) of time-dependency in the errors for a time series at any level of the data set hierarchical structure.

For this particular research, researchers implemented and tested the performance of an error structure at the lowest level of the hierarchical structure in the data. The methods implemented are those widely accepted and used in modeling time series modeling originally proposed by Box et al. and Tiao and Box, [*96, 97*].

The general model framework is known as Auto Regressive Integrated Moving Average (ARIMA) modeling. This error specification accounts for the degree to which a given value in the time series is determined by prior values in the time series. Although researchers found a level 1 integration in the car-following speed time series, researchers controlled for such integration by choosing to analyze the first time derivative of the speed (i.e., the acceleration) instead. Therefore, researchers considered only ARMA for the time structure (without the "I" in ARIMA). Equation 9 shows the general form of the error structure under the ARMA specification.

$$\varepsilon(t) = \sum_{u=1}^{p} \delta_u \cdot L^u \varepsilon(t) + \sum_{v=1}^{q} \vartheta_v \cdot L^v \phi(t) + \phi(t) \qquad \textbf{Equation 9}$$

Where:

$\varepsilon(t)$	=	Acceleration residual at epoch t.
δ_u	=	Coefficient for $L^u \varepsilon(t)$ in the combination of lagged residuals in ARMA model.
ϑ_v	=	Coefficient for $L^v \phi(t)$ in the combination of lagged residual nuances in ARMA model.
$L^u \varepsilon(t)$	=	Lag u of residual time series [i.e., $L^u \varepsilon(t) = \varepsilon(t - u)$].
$L^v \phi(t)$	=	Lag v of residual nuance time series [i.e., $L^v \phi(t) = \phi(t - v)$].
P	=	Largest lag in the autoregressive part of the ARMA model.
q	=	Largest lag in the moving average part of the ARMA model. Other variables as previously defined.

The error structure of the model is such that $E[\phi(t)] = 0$ and $V[\phi(t)] = \sigma_0^2$. The expected variance is estimated along with the rest of parameters in the model.

Discussion on Dynamic Model

A subset of variables defines the dynamic model proposed: car-following speed, car-following acceleration, lead speed, relative speed, and gap. These five are variables of state so that the kinetic state of the car-following process is defined by the current value of these variables.

The model specification is such that the acceleration is the response and other variables are specified as explanatory in the model. However, the dynamic nature of the model and the error correlation structure imply that a future state in the car-following process is a direct result of both past values of the response and levels in the set of explanatory variables. Therefore, all past configuration and state variables are explanatory for subsequent states of the system.

Modeling Process

In the modeling process, researchers considered various competing specifications in the dynamic submodels. The two key time series variables researchers decided to use are the gap between following and lead vehicles and the relative speed between them. The reason for electing these two is that they jointly define the state of the car-following system (two cars). Therefore, these variables should capture the information a driver has available for deciding what the appropriate acceleration should be. Also, results from such a model can be translated in terms of TTC because the state of the system is determined. This feature could then allow comparing the results to past research focused on TTC as a metric. However, researchers imposed the condition that both time series of interest (gap and relative speed) be lagged equally in the dynamic model for a given age sub group, as shown in Equation 7. The rationale for this constraint is the expectation that a driver adjusts their driving based on information gathered simultaneously at some point in the immediate past. Therefore, finding differences in the lags between the dynamic submodels by age would be an indicator of differences in driver performance by age. Furthermore, researchers hypothesize that $\omega_{ag} = E[Reaction\ Time_{ag}]$. In other words, the lag in gap and relative speed should be an estimator of the average reaction time for a given age group.

Researchers fitted a preliminary model and performed stepwise model selection. After testing the performance of various ARMA configurations, researchers determined that an ARMA(p=3, q=1) was most parsimonious and stable in accounting for autocorrelation present in the unadjusted time series.

Initially, researchers fitted models specifying geometric and contextual variables in the fixed effects. The coefficients associated with these variables were yielding small and mildly significant shifts on the baseline acceleration. Levels of distraction among participants were also included in the dynamic model with similar results. For the sake of interpretability, researchers explored specifying these variables as shifts to the dynamic model components, which resulted in significant improvements in model information and reduction of unexplained variability.

To account for the variability associated with the age classes, and while trying to keep the model from degenerating into unstable estimates, researchers tried various thresholds to break the age classes for the dynamic submodels. Various competing models showed promise when having three or four age classes. However, a specification of three age classes was selected to continue exploring the impact of other variables in the dynamic submodels. Informed in the literature review findings, and degree of dissimilarity between the dynamic models, researchers attempted quantifying differences of age between young, young adults, middle-aged, and older drivers. In the final models, researchers used the thresholds for the age classes shown in Table 13.

Researchers calculated the best SOC estimate available for each subevent using the models in Table 10 or Table 12, depending on the available predictors per subevent. Researchers initially attempted using SOC estimates directly in the modeling as a fixed effect, but results were not statistically significant. However, when including the SOC in the dynamic model as part of a differential SOC, results were statistically significant, reduced unexplained variability, and significantly higher quality of information from the model.

Table 13. Age Classes Used in Final Dynamic Model.

Age Class	Age Range
AC 1	17 to 19 years of age
AC 2	20 to 39 years of age
AC 3	40 to 69 years of age
AC 4	Older than 69 years of age

When fitting the initial models, researchers first specified dynamic submodels that accounted independently for relative speed and gap and noted that contextual variables entered the fixed effects that shifted the baseline acceleration accordingly (i.e., by LOS, number of lanes). However, interpretation of the impact of a context as a constant shift in the acceleration is challenging. When including SOC differential terms, all contextual variables dropped from the model and the fit of the models improved significantly.

In general, the amount of variability from the random effects structure remained very small through the modeling process. Researchers considers this an important and positive feature of the models, since it means that the fixed effects and the dynamic submodels explain the acceleration variability very well.

Results

After several rounds of the model selection procedure described above, researchers arrived at the final model shown in Table 14.

Differences in Reaction Time

As it can be seen in Table 14, the vector of Lag parameters (i.e., ω_n) clearly indicate declining performance for older drivers compared to younger drivers. The average perception/reaction time of drivers ages 16 to 20 is estimated to be 1.1 s. Very similarly, drivers ages 21 to 40 are estimated to have a delay in reaction to 1.2 s. Notably, drivers ages 41 to 70 are estimated to experience an additional 0.3 s in reaction time (i.e., average reaction time estimated to be 1.5 s). Finally, an average reaction time of 2.2 s is estimated for drivers 70 years of age or older.

Amount of Unexplained Variability

An important characteristic in the modeling results is the virtually inexistence of variability within the grouping structure. Despite the data representing 221 car-following subevents in 183 events among 145 drivers, the level of these nested groups that accounted for the most variability was that among events given an adjustment per participant has been applied (i.e., variance of 3.10×10^{-06} (mph/s)^2 or standard deviation of 7.87×10^{-04} mph/s).

Table 14. Final Dynamic Model for Car-Following Acceleration (n=29,178).

Omega parameters (i.e., Reaction Time Estimates)	Ages 16–19	Ages 20–39	Ages 40–69	Ages 70 and older		
	1.1 s	1.2 s	1.5 s	2.2 s		
	Estimate	Std. Error	DF	t-value	p-value	Significance
Baseline Acceleration	0.040 mph/s	0.030 mph/s	21727	−1.342852	0.1793	
Ages 16–19 (n=1,335)						
Relative Speed	+1.965 Hz	0.729 Hz	21727	2.695056	0.007	**
Relative Speed:ln(norm.gap)	−0.440 Hz	0.187 Hz	21727	−2.360499	0.0183	*
SOC differential	+0.068 Hz	0.039 Hz	21727	−1.753747	0.0795	#
Ages 20–39 (n=15,847)						
Relative Speed	+0.597 Hz	0.084 Hz	21727	7.072873	<0.0001	***
Relative Speed:ln(norm.gap)	−0.119 Hz	0.023 Hz	21727	−5.177161	<0.0001	***
SOC differential	−0.032 Hz	0.010 Hz	21727	3.384197	0.0007	***
Ages 41–70 (n=10,029)						
Relative Speed	+1.416 Hz	0.130 Hz	21727	10.906901	<0.0001	***
Relative Speed:ln(norm.gap)	−0.312 Hz	0.033 Hz	21727	−9.421193	<0.0001	***
SOC differential	−0.019 Hz	0.009 Hz	21727	2.118484	0.0341	*
Ages 70 and older (n=1,956)						
Relative Speed	+0.675 Hz	0.277 Hz	21727	2.433668	0.015	*
Relative Speed:ln(norm.gap)	−0.130 Hz	0.066 Hz	21727	−1.971608	0.0487	*
SOC differential	−0.013 Hz	0.024 Hz	21727	0.532469	0.5944	
Error Structure						
ARMA Parameters			Notes: Baseline following gap for the model is normalized at 3.28 ft (i.e., 1 m) Significance levels are as follows: # = Significant at the 0.10 level * = Significant at the 0.05 level ** = Significant at the 0.01 level *** = Significant at the 0.001 level			
Delta 1	+0.4757502					
Delta 2	+0.3015964					
Delta 3	+0.1864466					
Theta 1	−0.8370979					
phi 0	1.476 mph/s					
Grouping Structure (within group variance)						
Participants (n=145))	2.02×10^{-10} (mph/s)^2					
Events given a Participant (n=183)	3.10×10^{-06} (mph/s)^2					
Subevents given an Event and a Participant (n=221)	5.72×10^{-09} (mph/s)^2					

The implication is that the dynamic model can explain almost all acceleration variability in the data set except for a residual variability of ±1.476 mph/s as indicated by the phi_0 parameter in the error structure.

Relative Differences in Dynamic Model Coefficients

A preliminary examination of the results shows that young drivers are salient. First, results indicate that younger drivers have increased sensitivity of their acceleration behavior to both the relative speed of the lead vehicle and the distance gap between them and the lead vehicle (per relative differences in the first two coefficient estimates among age-group submodels).

Another interesting characteristic of younger drivers becomes evident when noting that coefficients in the dynamic submodels are generally consistent across age groups except for the coefficient from the SOC differential for drivers 16 to 20 years of age. Even though other age groups were found to reduce their acceleration proportionally to their SOC differential, younger drivers were found to do the opposite.

Relative Speed, Car-Following Gap, and Acceleration

This research found that drivers of ages between 16 and 20 years accelerate on average at a rate of 1.9 mph/s for each 1 mph in lead vehicle relative speed when the car-following gap is the reference 3.28 ft. This rate of acceleration decreases with increasing gap. This effect is estimated at an average decrease of 0.136 mph/s per 1 mph in relative speed for each 2 fold increase in car-following gap (0.136 mph/s =[−0.440 Hz]*[1 mph]*2), or an average decrease of 0.317 mph/s per 1 mph in relative speed for each 5 fold increase in car-following gap (−0.136 mph/s =[−0.440 Hz]*[1 mph]*5).

In contrast, drivers with ages between 21 and 40 years accelerate on average at a slower rate 0.60 mph/s for each 1 mph in lead vehicle relative speed when car-following gap is the reference 3.28 ft). Like all other age groups, this rate of acceleration decreases with increasing gap. For this age group, the average decrease in is 0.037 mph/s per 1 mph in relative speed for each 2 fold increase in car-following gap (−0.037 mph/s =[−0.119 Hz]*[1 mph]*2), or an average decrease of 0.085 mph/s per 1 mph in relative speed for each 5 fold increase in car-following gap (−0.085 mph/s =[−0.119 Hz]*[1 mph]*5).

Drivers with ages between 41 and 70 years on average accelerate at a slower rate (1.416 mph/s for each 1 mph in lead vehicle relative speed when car-following gap is the reference 3.28 ft). Like all other age groups, this rate of acceleration decreases with increasing gap. For this age group, the average decrease is 0.097 mph/s per 1 mph in relative speed for each 2 fold increase in car-following gap (−0.097 mph/s =[−0.312 Hz]*[1 mph]*2), or an average decrease of 0.225 mph/s per 1 mph in relative speed for each 5 fold increase in car-following gap (−0.225 mph/s =[−0.312 Hz]*[1 mph]*5).

Finally, drivers with ages 70 years or older accelerate at a slower rate on average (0.675 mph/s for each 1 mph in lead vehicle relative speed when car-following gap is the reference 3.28 ft). Like all other age groups, this rate of acceleration decreases with increasing gap. For this age group, the average decrease is 0.040 mph/s per 1 mph in relative speed for each 2 fold increase in car-following gap (−0.040 mph/s =[−0.130 Hz]*[1 mph]*2), or an average decrease of 0.080 mph/s per 1 mph in relative speed for each 5 fold increase in car-following gap (−0.080 mph/s =[−0.130 Hz]*[1 mph]*5).

SOC Differential and Acceleration

In addition to the acceleration adjustments associated with gap and relative speed, the model quantified how drivers tend to adjust their speed toward their estimated SOC. For drivers older than 21 years of age, the trend for this effect was found to consistently decrease with increasing age group. The notable

exception to this adjustment toward the SOC was found for drivers of ages between 16 and 20 years who tend to adjust their speed opposite to their estimated SOC, which is counter intuitive. This coefficient is statistically significant at the 0.1 level, which is suggestive evidence against the hypothesis of no effect at all. Regardless, it is estimated that for each 1 mph in SOC differential, younger drivers adjust their speed by 0.068 mph/s away from the SOC (0.068 mph/s =0.068 Hz * 1 mph).

Outside of the group of younger drivers, all other driver age groups were found to adjust their speeds toward the SOC in proportion to the difference between their car-following speed and their estimated SOC (i.e., SOC differential). This adjustment was found to decrease with increasing driver age. Drivers 21–40 years of age adjust their speed by 0.032 mph/s toward their SOC for each mph of SOC differential (0.032 mph/s= 0.032 Hz * 1 mph). Similarly, drivers 41–70 years of age adjust their speed by 0.019 mph/s toward their SOC for each mph of SOC differential (0.019 mph/s= 0.019 Hz * 1 mph). Finally, the model indicates that drivers older than 70 years of age would adjust their speed by 0.013 mph/s toward their SOC for each mph of SOC differential (0.013 mph/s= 0.013 Hz * 1 mph). However consistent with the other age groups with adjustments toward the SOC, this estimated adjustment for older drivers was not found statistically significant, which suggest that older drivers seem to adjust their car-following speed guided by their relative speed and car-following gap only. This is another potential description of speed adjustment by the youngest group, where the SOC differential effect was found barely significant and counter intuitive.

In order to investigate the feasibility of the hypothesis, researchers prepared plots of SOC differential vs. gap by age group. Figure 27 shows these plots. The color code indicates individual car-following subevents.

As it can be seen in this figure, the plots for age group 2 (21–40 years of age) and group 3 (41–70 years of age) indicate a very clear trend to adjusting car-following speed toward the SOC at larger gaps. This is the same general trend in the plot for age group 4 (ages 70 years and older) but with limited number of car-following events. This is probably the explanation of the statistical insignificance of this effect for drivers 70 years of age or older. In the case of drivers between 16–20 years of age, the number of events and variability of SOC differential are even smaller with no clear trend in a very limited range of values, which probably explains the statistical insignificance of the effect for the group of youngest drivers. Researchers believe that the counterintuitive direction of that effect is probably spurious and that given more car-following events for this age group, the coefficient could move to be negative and statistically significant, as for the other four groups.
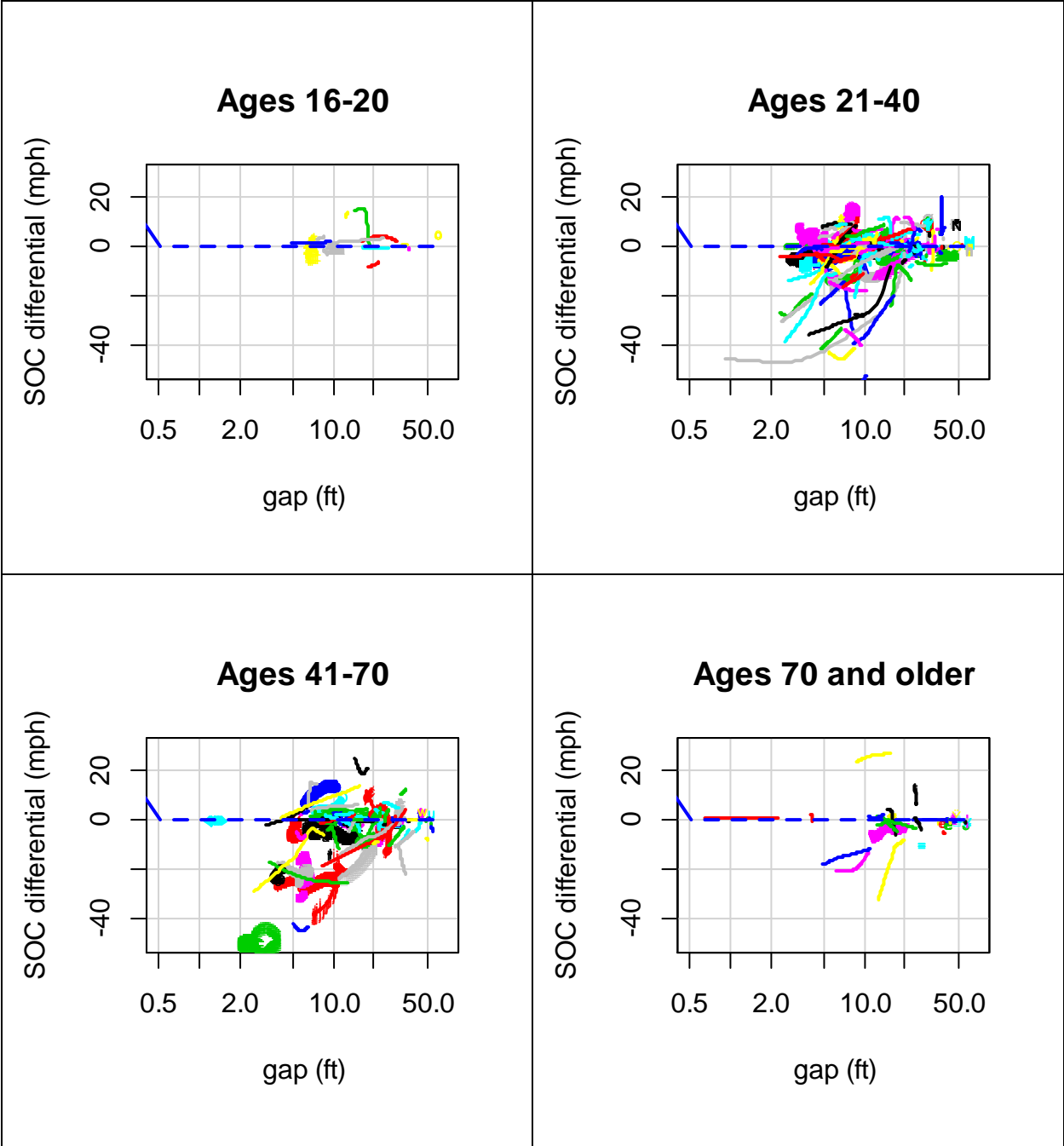
Figure 27. SOC Differential vs. Car-Following Gap by Age Group.

Finally, researchers prepared some plots to demonstrate the fit of the model in Table 14 to the data.
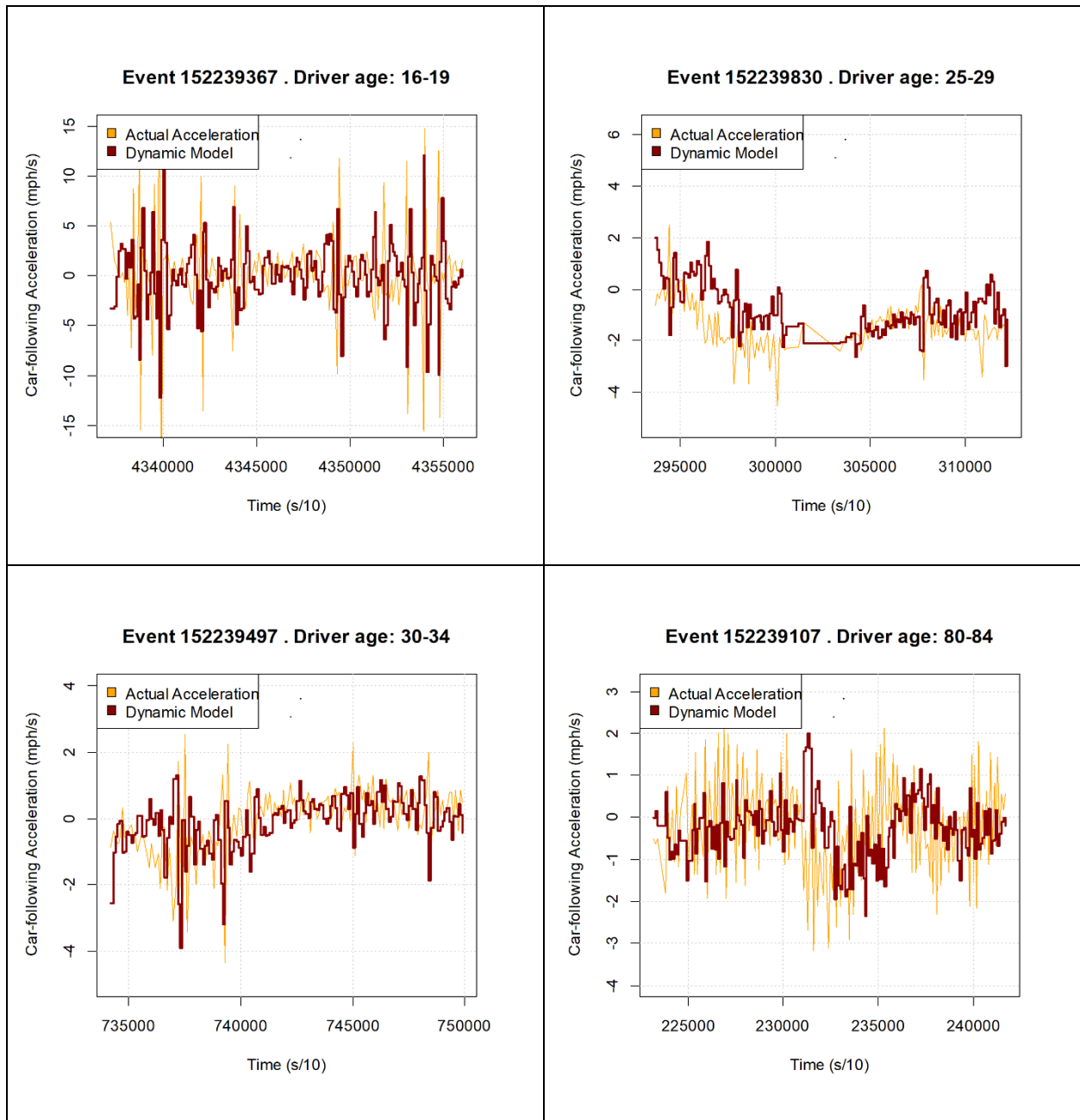
Figure 28. Car-Following Acceleration and Dynamic Model.

Figure 28 shows some sample car-following events. Although the dynamic model may deviate from the actual acceleration signal for short periods, it follows the raw data generally well. Although deviations are minimal, some are telling of remaining heterogeneity within the age groups. For example, the lag between the dynamic model and the actual acceleration observed for event 152239107 in the lower

right quadrant of the figure clearly indicates that this elder driver (age between 80 and 84) has a slightly faster reaction time than estimated for his age group.

Discussion of Results

This study produced a credible dynamic model for car-following accelerations by age groups. The results indicate that younger drivers tend to adjust their speeds more actively in response to their relative speed to the lead vehicle and car-following gap. Their speed adjustment seems to be independent of the SOC differential, as opposed to other age groups that seem to account for this differential when applying accelerations to their driving.

In general, a decaying performance with increasing age was found in the dynamic model: estimated reaction times decayed from 1.1 s down to 2.2 s as the age of the drivers increased from 16 to 19 year-olds to 70 years and older.

Interestingly, the amount of speed adjustment per unit of SOC speed differential decayed with age group too (except for the younger age group, in which case, it is likely that limited data explain a counterintuitive result in this regard). This seems to indicate that drivers are attentive to the dynamic conditions when adjusting their speed regardless of their age but their pre-conceived expectation what the free flow speed should be (i.e., SOC in these analyses) becomes less relevant as they age.

CHAPTER 7: STUDY 2 RAMP SPEED ANALYSIS

In this present study, researchers defined three types of speeding (State I, State II, and State III). Under this scheme, drivers adapt one of the three types of driving state during each trip including both on-ramp (SF) and off-ramp (FS) trips. Naturally, a driver's choice of the speeding state is determined by multiple factors such as the ramp design, direction of travel, driver characteristics, time and day of trip, and other trip details including speed variance, speed mean, and number of trips per driver at each study location. Researchers conducted a time-series-classification analysis and then compared how external and personal factors seem to relate to speeding.

Several methods exist for conducting the time series feature extraction and classification analysis. These include wavelet transformation, Fourier transformation K-nearest neighbor, piecewise linear approximation, support vector machines, artificial NN, etc. [98, 99]. In this study, researchers applied wavelet transformation, DTW, and NN to investigate speeding behavior on freeway ramps.

As the first step, researchers applied a wavelet transformation to reduce the time series data dimension and obtain a relatively balanced times series data. Using a DTW algorithm, researchers then clustered the on-ramp and off-ramp speed traversals into three states of driving and then analyzed that result using NN architecture to uncover variables associated with the three driving clusters.

Time Series Reduction and Matching

Researchers applied the DWT method recursively to reduce the dimension of the speed series and create a relatively balanced set of traversal profiles. Each series was saved together with all levels of wavelet decompositions. After reviewing the resulting reduced sequences of the trip time series, researchers selected the reduced series of the lengths ranging from $20 < Seconds \leq 41$. Figure 29 depicts an example of the SF and FS trips completed by two drivers (D-171263 and D-639928).

All four trips (2 SF and 2 FS) were completed on Ramp #3. After applying the recursive wavelet transformation in two iterations, the FS speed series of the two drivers have been reduced from 126 and 167 seconds to 31 and 41 seconds, respectively. The SF trips, after two iterations, have been reduced from 109 and 113 seconds to 27 and 28 seconds, respectively.

The transformed values represent the trip speeds after they have been transformed twice. For example, if driver D-171263 has the following speeds at the initial four seconds of the SF trip: $S^0_{t,t=1,2,3,4} = (30.5, 30.02, 31.31, 33.7)$ where the superscript 0 indicates the initial level (Figure 29, upper-left graph), this time series will be reduced to a sequence of 3 values after the 1st level wavelet transformation : $S^1_{t,t=1,2,3,} = (30.26\sqrt{2}, 30.66\sqrt{2}, 32.52\sqrt{2}) = (42.79, 43.36, 45.99)$ where $30.26 = \frac{30.5+30.02}{2}$ and so on.
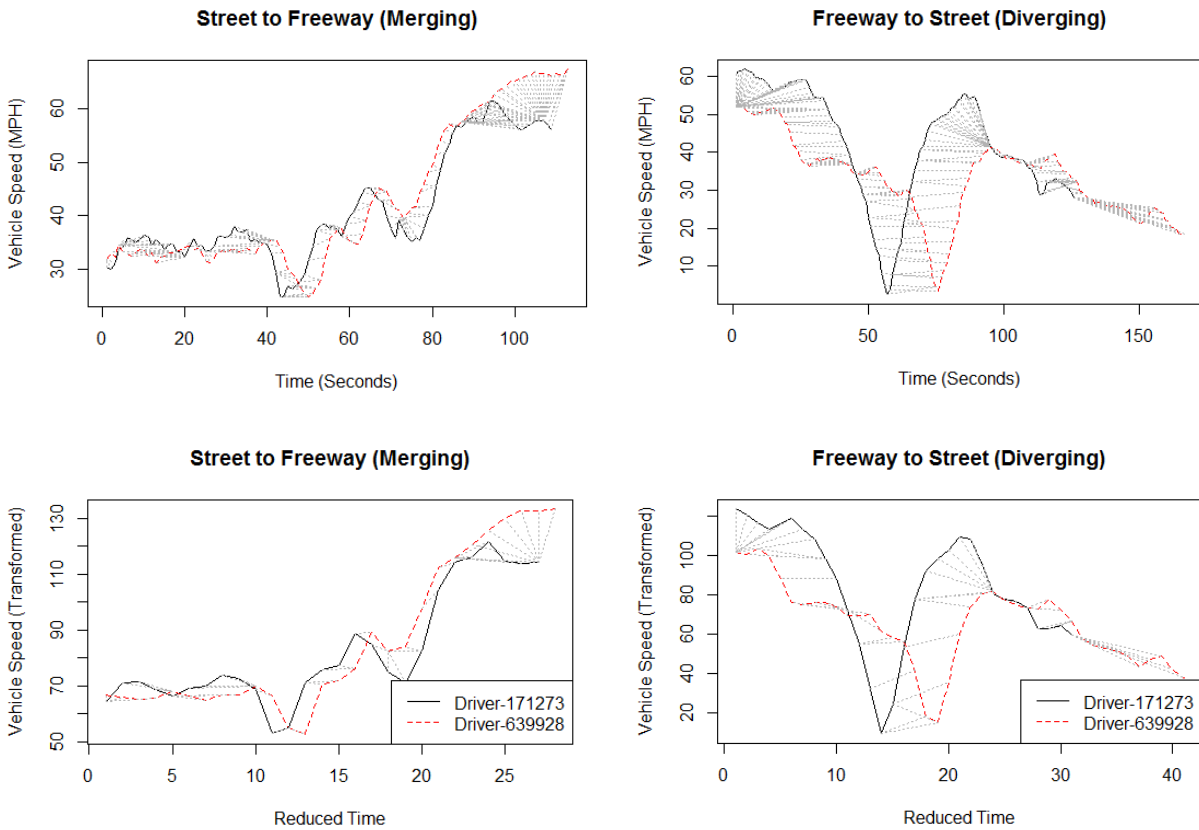
Figure 29. Matching Two Traversals: Original vs. DWT – Transformed Series.

The rationale of applying the coefficient $\sqrt{2}$ is to reduce energy loss [*100*]. An additional iteration yields the second level wavelet transformation: $S^2{}_{t,t=1,2} = \left(43.08\sqrt{2}, 44.68\sqrt{2}\right) = (60.92, 63.19)$ where $43.08 = \frac{42.79+43.36}{2}$, and so on (Figure 29). As can be observed, no information is lost, and the shape of the series is preserved.

Speed Profile Clusters

Researchers ran a cluster analysis on the reduced speed profiles using the PAM algorithm: State I, State II, and State III. The clustering was conducted using the R open software package [*101*]. As shown in Table 15, 242 FS trips and 129 SF trips were assigned to State I, 19 FS trips and 45 SF trips were assigned to State II, and 131 FS trips and 82 SF trips were assigned to State III.

Figure 30 shows the three states of driving recognized by the cluster analysis. The values shown on the *y*-axis do not represent the actual speeds but rather the transformed speeds. Also, the clusters show the individual trip series, meaning that the same driver could be represented in each of the three clusters at different trips. However, over repeated periods, the true natural inclination of the driver will tend to place him or her on one state or another.

Table 15. Summary of Speed Profile Clusters.

Clusters	Freeway to Street		Street to Freeway	
	# of trips	*% of total*	*# of trips*	*% of total*
State I	242	62%	129	50%
State II	19	5%	45	18%
State III	131	33%	82	32%
Total	392	100%	256	100%

At first inspection, Figure 30 shows high volatility on speed series from State I (purple colored). Some of those values often reach zero (coming to a complete stop). Although volatility is still present among speed profiles from State II (green and yellow colored), speeds in this group rarely reach a value of zero. Finally, speed from State III (green and turquoise colored) clearly exhibit smoother transitions from freeway to street or vice versa.
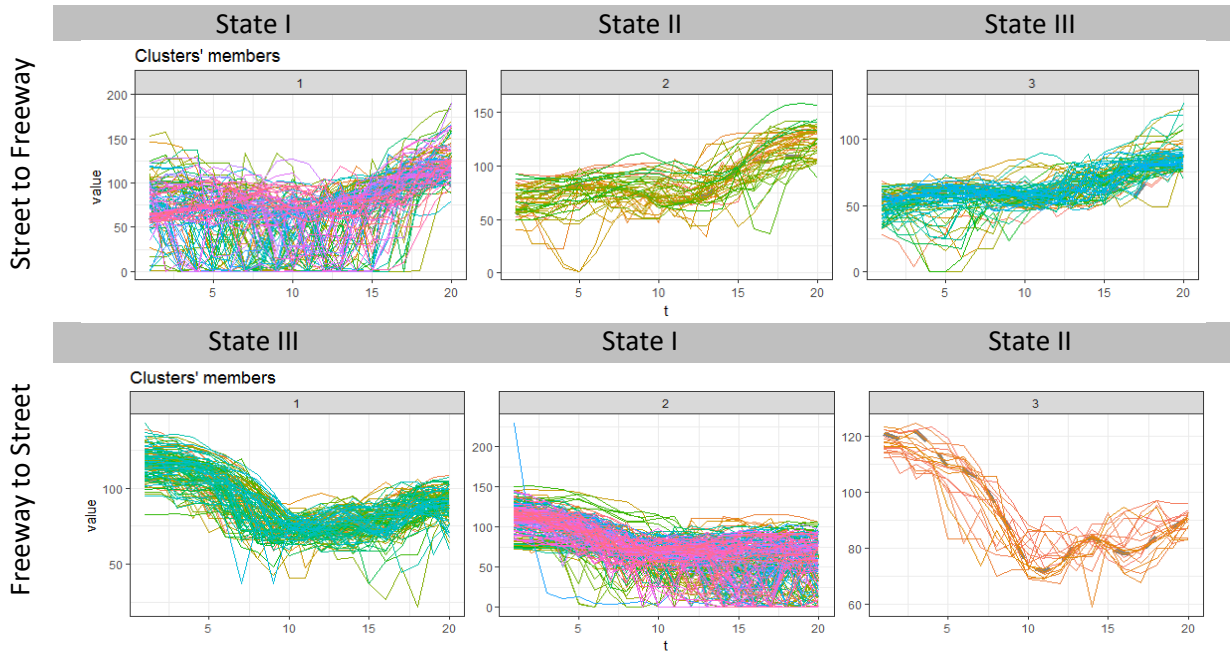
Figure 30. Speed Profile Clusters.

Observing the volatility of each cluster, one explanation might be that at State I drivers are less cautious because there are abrupt changes and sudden stops in the speed profiles. However, their driving behavior could potentially be attributed to congestion too (i.e., the vehicles might stop due to lead vehicles stopping). Likewise, States II and III could indicate either more cautious drivers or less congested conditions. For the following analysis, researchers review and interpret the results under the hypothesis that an increasing state level indicates a more cautious behavior by the drivers. Note that cautious driving refers in this case to the smoothness of the speed profile and it should not be interpreted as safe driving or risky driving necessarily.

Results of Network Analysis

Researchers used 47 input variables to assess their association with the three states of speed on the ramps using the NN architecture. Ramp characteristics, driver characteristics, and trip characteristics are the broad categories that these variables can be grouped into. To conduct the NN analysis, the data were sampled into training and testing data sets. The NN is trained using 70 percent of observations connected through 15 hidden neurons. The testing data (remaining 30 percent of observations) were then predicted using the trained algorithm. The mean squared errors showing the difference between the predicted and observed data were estimated to be 2.5 percent, 2.0 percent, and 3.4 percent, for the States I, II, and III, respectively. This observation indicates a highly accurate prediction of each output.

Researchers used Garson's algorithm [102] to establish relative importance of variables. This algorithm uses the absolute values of connection weights to determine the relative importance of inputs in the network. The importance of each variable is shown in Table 20, and it is defined such that the sum of all contributions will be equal to 100 percent. As shown, the speed variance is the most important predictor of the speeding state response variable with almost 7 percent contribution. Trip weekday, mean speed, maximum speed, trip time, and travel direction each contribute 3 percent or more to the entire network. In general, the importance diagram implies that the trip time, date, and the locations of the trips are the main contributors to the speeding behavior. Therefore, speeding behavior most likely reflects the traffic conditions rather that the driver characteristics.

The connection weights can help to explore the direction and the magnitude of the input effects on the outputs (see Table 16).

Table 16. NN Connection Weights and Importance of Predictors of Speeding Behavior.

Variable Name	Importance (% contribution)	Overall Connection Weights			Variable Name	Importance (% contribution)	Overall Connection Weights		
		State I	*State II*	*State III*			*State I*	*State II*	*State III*
Speed Variance	6.3%	37.04	21.83	1.95	Risks for Fun	2.0%	7.96	4.70	2.93
Trip Week Day	4.0%	7.15	−7.68	−14.00	Difficulty Enjoying Leisure Activities	2.0%	18.04	−2.76	−3.79
Mean Speed	3.9%	−43.98	17.03	21.73	Sudden Lane Change	2.0%	−8.70	0.76	−4.78
Maximum Speed	3.4%	15.22	26.80	−1.95	Losses Things	2.0%	14.02	−8.08	4.92
Trip Time	3.0%	−16.26	−15.34	6.23	Feels Restless	1.9%	5.91	6.06	9.15
Travel Direction	3.0%	−11.41	−26.51	22.14	RPS	1.9%	0.49	−5.24	26.23
Not Wearing Seatbelt	2.9%	−10.77	12.78	8.22	Area Type	1.9%	−3.94	−5.65	−2.37
Trip Year	2.8%	14.48	−14.92	−10.40	Vehicle Classification	1.9%	−12.25	6.91	−7.86
Sleeper Type	2.8%	11.95	15.77	−5.43	Road Rage	1.9%	−0.25	3.95	4.45
Interchange ID	2.7%	−12.12	−4.39	−3.48	Passenger Interaction	1.9%	−1.88	3.17	4.97
Ramp Design	2.5%	3.38	−7.61	0.03	Speeding < 20 mph	1.8%	−9.18	−7.66	15.32
Trip Month	2.4%	−13.20	8.60	−28.88	Passing on Right	1.8%	−0.10	5.43	−0.80
Accelerates at Yellow Light	2.4%	5.56	9.42	−15.27	Driving to Reduce Tension	1.7%	−0.69	14.70	17.59
Age	2.4%	−1.39	−4.44	8.65	Difficulty Awaiting Turn	1.7%	−0.04	11.17	−2.30
Speeding > 20 mph	2.3%	0.82	12.75	2.88	Rolling Stop Sign	1.7%	−8.33	2.50	1.90
Driving Under Influence	2.3%	4.56	5.42	−4.53	Number of Trips per Driver	1.7%	21.35	−10.84	−5.69
Road Racing	2.2%	16.54	10.01	4.81	Speeding for Thrill	1.5%	−1.12	0.20	1.74
Depth Perception	2.2%	−9.09	4.76	−1.36	In a Hurry	1.5%	3.11	4.48	−10.62
Sleep Schedule	2.2%	14.38	19.27	11.42	Running Red Light	1.5%	−14.24	8.60	4.89
Tailgating	2.1%	0.52	11.36	−8.95	Illegal Turns	1.3%	−8.82	0.80	−1.44
Difficulty Organizing	2.1%	−9.44	14.48	−4.12	Gender	1.3%	2.30	3.06	20.53
Easily Distracted	2.1%	28.11	−15.44	−10.07	Barkley Score	1.1%	12.60	0.54	5.69

Figure 31 is a graphic representation of the driving states and some of the most important factors.
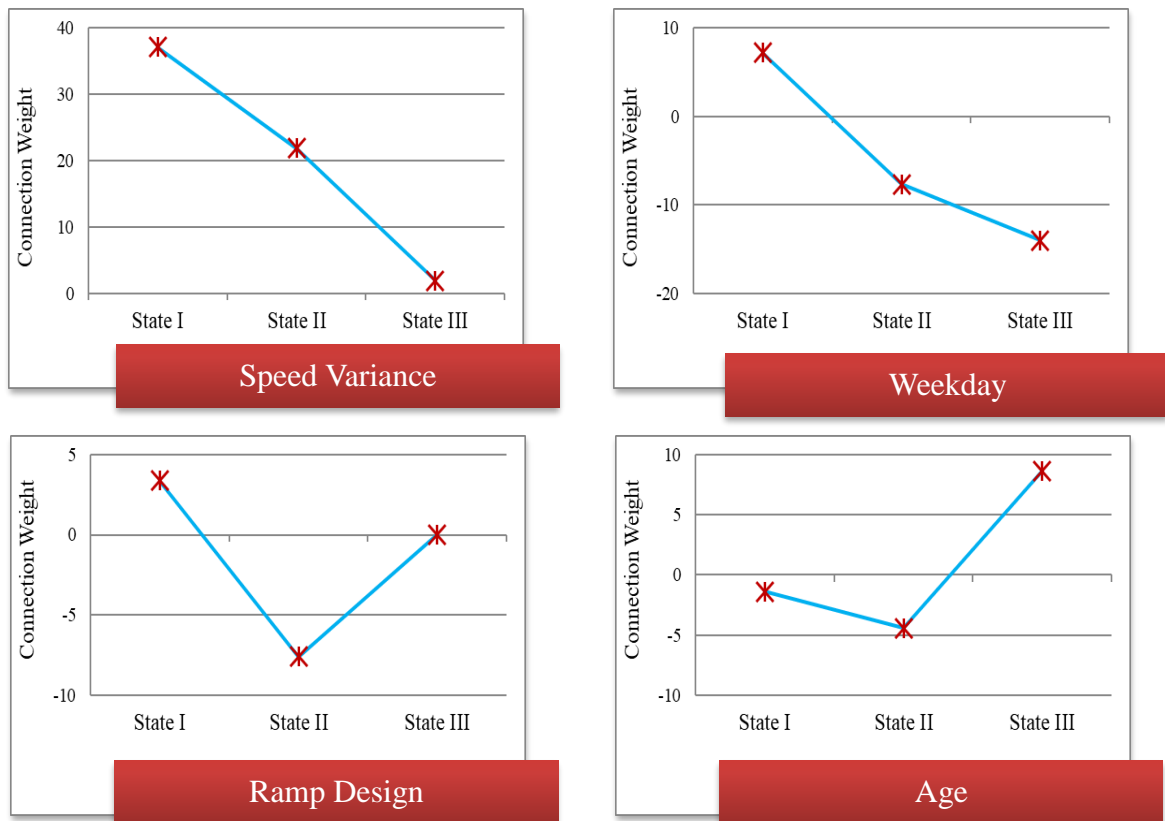
Figure 31. Description of Relationships between the Driving States and Important Variables.

The connection weights for the speed variance indicate that as drivers shift from State I to State III, the magnitude of this effect starts to decrease. The results in Table 16 also indicate that the two other speed related inputs, mean and maximum speeds, are important. In State I, the mean speed has a negative association. This is expected because this state represents some drivers that come to a complete stop. As the drivers move from State I to State III, the association of the mean speed is increasingly positive. This finding can be interpreted together with the inverse association of speed variance and the increasing index of the speeding states: the association of speed variance is smaller in State III compared to State II and that, in turn, is smaller compared to State I. However, the association of mean speed shows the opposite trend (i.e., increasing with increasing speed state index). Although this could imply that speed variance may indicate cautious driving behavior, in the most likely scenario that the driver states associate with traffic conditions, the associations of mean speed and speed variance would simply suggest that when the mean speed is higher, it has also a smaller variance. Again, this scenario is consistent with the characteristics of congested versus uncongested traffic.

The second most important variable group identified in the NN analysis is the time and date of the trips. For State I, the effect of week day is increasing while for the States II and III the effect is decreasing. Note that the week day is in increasing order from Monday to Sunday. This implies that as the week starts to progress more drivers start driving in State I and less drivers drive in State III. If the driving states indeed reflect traffic conditions, this result indicates that during the weekdays from Monday thru Friday drivers encounter more traffic congestion as opposed to the weekends.

Perhaps the most interesting relationship found is the association with the ramp design. The ramp design variable was coded from least sophisticated (or complex) to more sophisticated. An example of a simple ramp is that of a diamond interchange, while a complex ramp is that from a partial cloverleaf interchange. The importance of the influence of ramp design is moderate to high (ranked 11 out of 47 factors in Table 16). The relationship between State III driving and ramp design is very small indicating that drivers are as likely to exhibit State III speeds (higher speed mean and lower variance) regardless of the complexity of the ramp design. In contrast, the effect is large and negative for State II, which implies that drivers become less likely to exhibit driving behavior in this state as the interchange becomes more sophisticated (after discounting the influence of traffic). Finally, the effect is the opposite for State I. A positive weight suggests that drivers become more likely to exhibit driving characteristics within the State I pattern (after discounting the influence of traffic) as the ramps become more challenging to navigate.

Among the driver characteristics, the age, risk perception elements, and sleeping habits are observed to contribute more to the driving states. The driver age is negatively related to State I and II and positively related to State III. This finding implies that older drivers tend to exhibit driving patterns more like State III and less like the State I and State II driving after discounting other influential elements, such as time of day and geometric design, per the discussion above.

The RPS is positively related to State III with a magnitude of the effect being very high. This feature suggests that State III associate with high RPS. For State II, the weight for RPS is negative, implying that this state of driving tends to have drivers with smaller RPS. An examination of the speed-related elements of the risk perception survey (i.e., accelerates at yellow light, speeding for thrill, or speeding over the speed limit) suggests that the State III driving seem to associate with drivers who perceive the risks related to these behaviors poorly. State II tends to have such drivers less represented.

The Barkley's ADHD score was not found to be significantly associated with the driving states although individual elements of the survey were found to be of importance (easily distracted and difficulty organizing).

CHAPTER 8: CONCLUSIONS AND FUTURE DIRECTIONS

This project identified and quantified relationships between traffic OP variables (such as operating speeds), HF characteristics (e.g., driver demographics), and SA variables (crash or near-crash outcome) making use of the recently available SHRP2 databases.

Researchers assembled two data sets from the SHRP2 products to study driver performance and investigate how the operational characteristics may explain SA outcomes.

Study 1

The first database was assembled from close to 800 events from freeway trips in non-curve, uncongested conditions. Besides several exploratory analyses performed initially, researchers performed two main analyses: 1) investigation of factors that affect driver's SOC; and 2) driver performance on car-following situations.

SOC

For the SOC analysis, researchers developed models based on events where the drivers did not change their speed and they were not following another vehicle. These events were hypothesized to be representative of the driver's free flow SOC.

An examination to the potential explanatory variables available in the data, researchers recognized a significant limitation in the small subset for which PSL and other driver environment variables were not available. As a result, researchers elected to perform two sets of analyses: one with the bigger data not-including PSL and one with the smaller data set that included that variable.

In the analysis without PSL, estimation of SOC improved significantly when accounting explicitly for driver characteristics such as the number of previous traffic violations, driver age, and a set of issues with vision (nearsightedness, poor night vision, glaucoma, and use of reading glasses).

For the analysis including PSL, models with PSL were in general better predictors of SOC. Although researchers attempted to incrementally add driver characteristics to the best-fitting model that does not have such variables (similar to the analysis of data without PSL), a single driver-related variable (number of years driving) was found to have a marginal improvement on the prediction of SOC. Researchers speculate that the reason for this reduced impact of HF is explained by the limited sample size with PSL (less than a third of the sample size for SOC without PSL) and the significant power of PSL in explaining SOC.

A comparison on the precision of SOC predictions highlights the relevance of PSL in determining the SOC. The best model from the larger data set that included driver characteristics (but not including PSL) could explain all variability in SOC except for ±6.493 mph among events and ±4.969 mph residual variability. In contrast, the best model including PSL explained SOC variability except for ±4.503 mph among events and ±3.957 mph residual variability. This implies that the unexplained variability among events and overall residual variability were reduced by approximately 1 mph each. However, it is salient that the amount of variability between events remains large comparable to the residual variability in models with and without PSL. This suggests that event-specific factors, unaccounted for these models, have a large impact on the driver's selection of free-flowing speed.

Overall, the factors that explain most variability in SOC were PSL, traffic density, and functional class. Driver characteristics such as driver age and visual acuity were found influential too, but such influence was quantified in a larger data set that did not have PSL, the most relevant predictor. Therefore, researchers could not assess the relative influence of these human-factors related variables and PSL. Only years of driving experience was found marginally influential on SOC in the presence of PSL as a predictor.

Future Work with SOC

Future research should assemble a larger data set with both PSL and driver features to develop a more comprehensive model that simultaneously account for the influence of driver context variables (such as PSL and road geometry) and driver characteristics on SOC. Such future work should investigate what other event-specific variables could contribute to reducing the large amount of unaccounted variability between events that was found even in the best SOC models with PSL as a predictor.

Car-Following Behavior and Human Factors

For the second part of this study, researchers analyzed a set of 221 car-following subevents from 145 drivers to investigate the influence of driver characteristics in car-following performance. Researchers specified a relatively complex dynamic model. Initially, the model focused on following speed and allowed the inclusion of time-invariant covariates as fixed effects that shift the dynamic model (the dynamic model itself allowed for both random and fixed effects). It was determined that the speed could not be modeled as a stationary time series and thus the focus moved to the car-following acceleration.

All time-invariant covariates considered were dropped in the acceleration model during the step-wise model selection, except for the role of some of them in determining the SOC, which entered the dynamic model in a compound variable.

A visual examination of the relationship between the acceleration and other stochastic variables (such as car-following gap and car-following relative speed) suggested a lagged relationship. Meaningfully, this lagged relationship is expected if a causal relationship exists (i.e., a change in acceleration responds to a change in the other variables). Researchers hypothesized that such causal relationship exists and specified the dynamic model to include the lag operator to account for this feature. Furthermore, researchers specified four acceleration submodels, one for each of four mutually exclusive age groups, to quantify the differences in car-following performance that are expected by driver age.

Researchers allowed the degree of lagging between the acceleration and the explanatory time series to vary during the estimation of the dynamic model coefficients, to assess differences in performance between driver age groups. Results showed, as expected, a set of lags that indicated a declining performance with increasing age. The lag in the dynamic acceleration submodel for drivers between 16 and 19 years of age suggests that this group have an average reaction time of 1.1 s. For drivers of ages 20–39, the reaction time estimate was nearly identical (1.2 s). Drivers of ages 40–69 were found to have longer average reaction times (1.5 s). Finally, the reaction time for older drivers was estimated significantly larger at 2.2 s, twice as long as the youngest drivers.

Regardless of the differences found in estimated reaction times, the results from the dynamic model indicate notable differences in the acceleration adjustment that different age groups make in response

to the state of the car-following system. Following behavior was found most sensitive to the car-following relative speed at close range for drivers 16–19 years of age. The second-most sensitive group in terms of sensitivity to relative speed at close range was drivers between 40 and 69 years of age. The remaining two groups were found to have very similar sensitivity to relative speed at close range.

The sensitivity to SOC differential was found, as expected, to indicate an adjustment toward the SOC, except for drivers 16–19 years of age (statistically significant at the 0.10 level). Based on an examination of the raw data, researchers argue that this counterintuitive finding could be due to a spurious happening and that the sensitivity to SOC may be close to zero for this age group. For the other three age groups, the amount of SOC sensitivity, with adjustment toward the estimated SOC, was found to consistently decrease with increasing age.

Future Work with Car-Following Behavior

Future research should examine the car-following model performance against driving situations that were not included in the modeling effort. For this proposed evaluation, more data are needed. With more events (or longer driving events), future work would can validate the differences found between age groups and evaluating the expected distortions in car-following behavior that distracted driving introduces. Ultimately, a larger set of crashes or near-crashes is needed to assess the impact of car-following behavior in the risk of crashes or near crashes.

Study 2

The second study concerns the SOC on freeway ramps. In this study, researchers used the SHRP2 data from Pennsylvania. Researchers used the trip summary, roadway, and driver characteristics to identify the most influential factors affecting the drivers' speed choice on the ramps. The most influential factors were found to be the time and week day of trip, direction of travel, and ramp characteristics. Based on the influence of these factors, drivers were found to adjust their speed more uniformly. In this study, the impact of driver characteristics was not found to be as important, although the results of the NN indicate that the driver characteristics influences the SOC in a certain degree. For example, the group of drivers who drove more consistently (State III) was not found to perceive driving at higher speeds as risky behavior. State I drivers are found to be easily distracted based on their Barkley ADHD scores.

In this study, researchers used rigorous and sophisticated statistical tools to conduct naturalistic data analysis. The results indicate that the SOC is not merely a momentary decision without any input but rather a process that is influenced by number of contextual factors. This process was found especially complex in the case of the car-following dynamic model. As future improvements to this work, researchers envision to increase the sample size of available data to validate the statistical models and computational tools developed in this project, as well as to investigate more nuanced relationships with other driver-related variables of interest, such as distraction level and types. Regardless, the results of this study are beneficial for traffic engineers, designers, and transportation data analysts. The knowledge harnessed about speed choice and car-following behavior could be used to screen for scenarios where the expected speeds are in disharmony with the context, and then a list of appropriate countermeasures to address the speed related SA issues can be developed.

REFERENCES

[1]   I. Van Schagen and F. Sagberg, "The potential benefits of naturalistic driving for road safety research: Theoretical and empirical considerations and challenges for the future," *Procedia – Social and Behavioral Sciences,* no. 48, pp. 692-701, 2012.

[2]   V. Shankar, P. Jovanis, J. Aguero-Valverde and F. Gross, "Analysis of Naturalistic Driving Data: Prospective View on Methodological Paradigms," *Transportation Research Record: Journal of the Transportation Research Board,* no. 2061, pp. 1-8, 2008.

[3]   V. Neale, S. Klauer, R. Knipling, T. Dingus, G. Holbrook and A. Petersen, "The 100 Car Naturalistic Driving Study, Phase 1 – Experimental Design," NHTSA, Washington, DC, 2002.

[4]   T. Dingus, S. Klauer, V. Neale, A. Petersen, S. Lee, J. Sudweeks, M. Perez, J. Hankey, D. Ramsey, S. Gupta and C. Bucher, "The 100-car Naturalistic Driving Study, Phase II – Results of the 100-car Field Experiment," NHTSA, Washington, DC, 2006.

[5]   K. Campbell, *The SHRP 2 Naturalistic Driving Study: Addressing Driver Performance and Behavior in Traffic Safety,* TRB, 2012.

[6]   Virginia Tech Transportation Institute, "InSight SHRP2 NDS," 2017. [Online]. Available: https://insight.shrp2nds.us/home/index. [Accessed 20 July 2017].

[7]   AAA Foundation for Traffic Safety, "Aggressive Driving: Research Update," AAA Foundation for Traffic Safety, Washington, DC, 2009.

[8]   Federal Highway Administration, "Analyzing Driver Behavior to Understand the Factors Contributing to Highway Crashes," [Online]. Available: htpps://www.fhwa.dot.gov/goshrp2. [Accessed 14 March 2014].

[9]   L. Aarts and I. Van Shagen, "Driving Speed and the Risk of Road Crashes," *Accident Analysis and Prevention,* vol. 38, no. 2, pp. 215-224, 2006.

[10]   H. T. Zwahlen, "Advisory Speed Signs and Curve Signs and Their Effect on Driver Eye Scanning and Driving Performance," *Transportation Research Record ,* no. 1111, pp. 110-120, 1987.

[11]   K. Dixon, R. Avelar and C. Monsere, "Oregon Signalized Intersection Safety Performance Functions and the Effect of Speed," in *TRB 95th Annual Meeting Compendium of Papers*, Washington, D.C, 2016.

[12]   L. Friedman, D. Hedeker and E. Richter, "Long-term effects of repealing the National Maximum Speed Limit in the United States," *American Journal of Public Health. Vol. 99,* pp. 1626-1631, 2009.

[13]   M. Firestine, H. McGee and P. Toeg, "Improving Truck Safety at Interchanges," FHWA, Washington, DC, 1989.

[14]   A. McCartt, V. Northrup and R. Retting, "Types ad characteristics of ramp-related motor vehicle crashes on urban interstate roadways in Northern Virginia," *Journal of Safety Research,* vol. 35, no. 1, pp. 107-114, 2004.

[15]   S. Kim, T. Song, N. Rouphail, S. Aghdashi, A. Aramo and G. Gonçalves, "Exploring the association of rear-end crash propensity and micro-scale driver behaviour," *Safety Science,* vol. 89, pp. 44-54, 2016.

[16]   E. Hauer, "Speed and Safety," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2103, pp. 10-17, 2009.

[17]   R. Elvik, P. Christensen and A. Amundsen, "Speed and road accidents. An evaluation of the Power Model," TOI, 2004.

[18]   D. Solomon, "Accidents on main rural highways related to speed, driver, and vehicle," U.S. Department of Commerce, Washington, DC, 1964.

[19]   C. Lave, "Speeding, Coordination, and the 55 mph Limit.," *The American Economic Review,* vol. 75, no. 5, pp. 1159-1164, 1989.

[20]   N. Garber and R. Gadiraju, "Factors Affecting Speed Variance and Its Influence on Accidents," *Transportation Research Record,* vol. 1213, pp. 64-71, 1989.

[21]   H. Behbahani, N. Nadimi and S. Naseralavi, "New time-based surrogate safety measure to assess crash risk in car-following scenarios," *Transportation Letters,* pp. 229-238, 2015.

[22]   S. Gousios, N. Garber and Q. Lui, "Simulation based validation of time to collision as a safety performance surrogate," *Urban Transport ,* pp. 637-648, 2009.

[23] K. Pecheux, J. Bauer and P. McLeod, "Pedestrian Safety Engineering and ITS-Based Countermeasures Program for Reducing Pedestrian Fatalities, Injury Conflicts, and Other Surrogate Measures. Final System Impact Report," Federal Highway Administration, Washington, DC, 2009.

[24] T. Saleem and A. Lorion, "Estimating crash modiication factors using surrogate measures of safety," in *25th CARSP Conference*, Ottawa, Ontario, 2015.

[25] A. Stevanovic, C. Kergaye and J. Haigwood, "Assessment of Surrogate Safety Benefits of an Adaptive Traffic Control System," in *3rd International Conference on Road Safety and Simulation*, Indianapolis, IN, 2011.

[26] K. Wu and P. Jovanis, "Exploratory Modeling with Naturalistic Driving Data," *Accident Analysis and Prevention,* vol. 45, pp. 507-516, 2012.

[27] P. Songchitruksa and A. Tarko, "The extreme value theory approach to safety estimation," *Accident, Analysis, and Prevention,* vol. 38, no. 4, pp. 811-822, 2006.

[28] T. F. Golob, W. W. Recker and V. M. Alvarez, "Freeway safety as a function of traffic flow," *Accident Analysis and Prevention Vol. 36,* pp. pp. 933-946, 2004.

[29] T. Golob, W. Recker and V. Alvarez, "Tool to Evaluate Safety Effects of Changes in Freeway Traffic Flow," *Journal of Transportation Engineering,* vol. 130, no. 2, pp. 222-230, 2004.

[30] B. Cheng, Q. Lin, T. Song, Y. Cui, L. Wang and S. Kuzumaki, "Analysis of Driver Brake Operation in Near-Crash Situation Using Naturalistic Driving Data," *International Journal of Automotive Engineering,* vol. 2, no. 4, pp. 87-94, 2011.

[31] F. Guo, S. Klauer, J. Hankey and T. Dingus, "Near Crashes as Crash Surrogate for Naturalistic Driving Studies," *Transportation Research Record,* vol. 2147, pp. 66-74, 2010.

[32] J. Gibson and L. Crooks, "A Theoretical Field-Analysis of Automobile-Driving," *The American Journal of Psychology,* vol. 51, no. 3, pp. 453-471, 1938.

[33] M. Brackstone and M. McDonald, "Car-following: a historical review," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 2, no. 4, pp. 181-196, 1999.

[34] J. Caird, C. Willness, P. Steel and C. Cialfa, "A meta-analysis of the effects of cell phones on driver performance," *Accident Analysis and Prevention,* vol. 40, no. 4, pp. 1282-1293, 2008.

[35] J. Caird, K. Johnson, C. Willness, M. Asbridge and P. Steel, "A meta-analysis of the effects of texting on driving," *Accident Analysis and Prevention,* vol. 71, pp. 311-318, 2014.

[36] C. Yager, J. Cooper and S. Chrysler, "The Effects of Reading and Writing Text-Based Messages While Driving," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* vol. 51, no. 1, pp. 2196-2200, 2016.

[37] J. Meyer, "Designing In-Vehicle Technologies for Older Drivers," *Technologies for an Aging Population,* vol. 39, no. 1, 2009.

[38] P. Olson and M. Sivak, "Perception-Response Time to Unexpected Roadway Hazards," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 28, no. 1, pp. 91-96, 1986.

[39] N. Reed, N. Kinnear and L. Weaver, "Why do older drivers have more 'failed to look' crashes? A simulator based study," Transport Research Laboratory, London, UK, 2012.

[40] M. Svetina, "The reactioni times of drivers agesd 20 to 80 during a divided attention driving," *Traffic Injury Prevention,* vol. 17, no. 8, 2016.

[41] L. Staplin, K. Lococo, C. Martell and J. Stutts, "Taxonomy of Older Driver Behaviors and Crash Risk," NHTSA, Washington, DC, 2012.

[42] J. Cooper, I. Vladisavljevic, N. Medeiros-Ward, P. Martin and D. Strayer, "An investigation of driver distraction near the tipping point of traffic flow stability," *Human Factors,* vol. 51, no. 2, pp. 261-268, 2009.

[43] P. Cooper and Y. Zheng, "Turning gap acceptance decision-making: impact of driver distraction," *Journal of Safety Research,* vol. 33, pp. 321-335, 2002.

[44] H. Alm and L. Nilsson, "The effects of a mobile telephone task on driver behaviour in a car following situation," *Accident Analysis and Prevention,* vol. 27, no. 5, pp. 707-715, 1995.

[45] J. He, A. Chapparo, B. Nguyen, R. Burge, J. Crandall, B. Chaparro, R. Ni and S. Cao, "Texting while driving: Is speech-based texting less risky than handheld texting?," Eindhoven, The Netherlands, 2013.

[46] K. Brookhuis, G. de Vries and D. de Waard, "The effects of mobile telephoning on dirving performance," *Accident Analysis and Prevention,* vol. 23, no. 4, pp. 309-316, 1991.

[47] K. Thompson, A. Johnson, J. Emerson, J. Dawson, E. Boer and M. Rizzo, "Distracted driving in elderly and middle-aged drivers," *Accident Analysis and Prevention,* pp. 711-717, 2012.

[48] N. Medeiros-Ward, J. Cooper and D. Strayer, "Hierarchical control and driving," *Journal of Experimental Psychology,* vol. 143, pp. 953-958, 2014.

[49] J. Montgomery, K. Kusano and H. Gabler, "Age and Gender Differences in Time to Collision at Braking From the 100-Car Naturalistic Driving Study," *Traffic Injury Prevention,* pp. 15-20, 2014.

[50] A. Heino, H. van der Molen and G. Wilde, "Differences in risk experience between sensation avoiders and sensation seekers," *Personality and Individual Differences,* vol. 20, no. 1, pp. 71-79, 1996.

[51] M. Zuckerman, Sensation Seeking and Risky Behavior:, Washington, DC: American Psychological Association, 2014.

[52] W. Van Winsum and A. Heino, "Choice of time-headway in car-following and the role of time-to-collision information in braking," *Ergonomics,* vol. 39, no. 4, pp. 579-592, 1996.

[53] A. Quimby, G. Maycock, C. Palmer and A. Buttress, "The factors that influence a driver's choice of speed – a questionnaire study," TRL Report, Wokingham, UK, 1999.

[54] D. Massie, K. Campbell and A. Williams, "Traffic accident involvement rates by driver age and gender," *Accident Analysis and Prevention,* vol. 27, no. 1, pp. 73-87, 1995.

[55] R. Avelar, S. Hammond and K. Fitzpatrick, "An Investigation on Monthly and Hourly Trends in Crash Risk for Older Drivers at Freeways and Highways using SHRP2 Data," in *96th Annual Meeting of the Transportation Research Board ,* Washington, DC, 2017.

[56] J. Charlton, J. Oxley, B. Fildes, P. Oxley, S. Newstead, S. Koppel and M. O'Hare, "Characteristics of older drivers who adopt self-regulatory driving behaviours,"

Transportation Research Part F: Traffic Psychology and Behaviour, vol. 9, no. 5, pp. 363-373, 2006.

[57] IIHS, "Teenagers.," November 2016. [Online]. Available: http://www.iihs.org/iihs/topics/t/teenagers/fatalityfacts/teenagers.

[58] T. Seacrist, A. Belwadi, A. Prabahar, S. Chamberlain, J. Megariotis and H. Loeb, "Comparison of crash rates and rear-end striking crashes among novice teens and experienced adults using the SHRP2 Naturalistic Driving Study," *Traffic Injury Prevention,* pp. 48-52, 2016.

[59] J. Stutts, C. Martell and L. Staplin, "Indentifying Behaviors and Situations Associated with Increased Crash Risk for Older Drivers. DOT HS 811 093.," National Highway Traffic Safety Administration, Washington DC, 2009.

[60] S. Kauer, T. Dingus, V. Neale, J. Sudweeks and D. Ramsey, "The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-car Naturalistic Driving Study Data," NHTSA, Washington, DC, 2006.

[61] H. Summala, D. Lamble and M. Laakso, "Driving experience and perception of the lead car's braking when looking at in-car targets," *Accident Analysis and Prevention,* vol. 30, no. 4, pp. 401-407, 1998.

[62] J. Harbluk, Y. Noy and M. Eizenman, "The Impact of Cognitive Distraction on Driver Visual Behaviour and Vehicle Control," Transport Canada, Ontario, Canada, 2002.

[63] J. Greenberg, L. Tijerina, R. Curry, B. Artz, L. Cathey, D. Kochhar, K. Kozak, M. Biommer and P. Gant, "Dirver Distraction: Evaluation with Event Detection Paradigm," *Transportation Research Record,* vol. 1843, pp. 1-9, 2003.

[64] K. Young and M. Regan, "Driver distraction: A review of the literature," in *Distracted Driving*, Sidney, Australia, Australasian College of Road Safety, 2007, pp. 379-405.

[65] National Highway Traffic Safety Administration, "Risky Driving," 2017. [Online]. Available: https://www.nhtsa.gov/risky-driving. [Accessed 26 May 2017].

[66] C. Carney, D. V. McGehee, K. Harland, M. Weiss and M. Raby, "Using Naturalistic Driving Data to Examine Teen Driver Behaviors Present in Motor Vehicle Crashes, 2007-2015.," AAA Foundation for Traffic Safety, Washington DC, 2016.

[67] T. Victor, "Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk. Strategic Highway Research Program (SHRP 2) Report S2-S08A-RW-1," Transportation Research Board, Washington DC, 2015.

[68] IIHS, "Large naturalistic study provides new window on driver distraction.," 10 December 2015. [Online]. Available: http://www.iihs.org/iihs/sr/statusreport/article/50/10/3.

[69] T. Dingus, F. Guo, S. Lee, J. Antin, M. Perez, M. Buchanan-King and J. Hankey, "Driver crash risk factors and prevalence evaluation using naturalistic driving data," PNAS, 2016.

[70] L. Higgins, R. Avelar and S. Chrysler, "Effects of Distraction Type, Dirver Age, and Roadway Environment on Reaction Times – An Analysis Using SHRP-2 NDS Data," Manchester Village, VT, 2017.

[71] F. Guo, S. Klauer, Y. Fang, J. Hankey, J. Antin, M. Perez, S. Lee and T. Dingus, "The effects of age on crash risk associated with driver distraction.," *International Journal of Epidemiology,* pp. 258-265, 2017.

[72] J. Leversen, B. Hopkins and H. Sigmundsson, "Ageing and driving: Examining the effects of visual processing demands," *Transportation Research Part F,* pp. 1-4, 2012.

[73] M. Brackstone, B. Waterson and M. McDonald, "Determinants of following headway in congested traffic," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 12, no. 2, pp. 131-142, 2009.

[74] H. Perrin, P. Martin and B. Hansen, "Modifying Signal Timing During Inclement Weather," *Transportation Research Record,* vol. 1748, 2001.

[75] M. Haglund and L. Aberg, "Speed choice in relation to speed limit and influences from other drivers," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 3, no. 1, pp. 39-51, 2000.

[76] Transportation Research Board, "Strategic Highway Research. Saving Lives Reducing Congestion Improving Quality of Life. Special Report 260," National Accademy Press, Washington, DC, 2001.

[77] A. Comoreanu, "Strictest And Most Lenient States on Speeding and Reckless Driving," 2016. [Online]. [Accessed November 2017].

[78] J. Antin, S. Lee, J. Hankey and T. Dingus, "Design of the In-Vehicle Driving Behavior and Crash Risk Study: In Support of the SHRP 2 Naturalistic Driving Study," Transportation Research Board, Washington, DC, 2011.

[79] B. Jonah, "Sensation Seeking and Risky Driving: A Review and Synthesis of the Literature," *Accident Analysis and Prevention,* vol. 29, no. 5, pp. 651-665, 1997.

[80] I. Thielen, R. Hartmann and D. Soares, "Risk Perception and Speeding," *Cad Saude Publica,* vol. 24, no. 1, pp. 131-139, 2008.

[81] E. Hartmann, F. Baekland and G. Zwilling, "Psychological differences between long and short sleepers," *Archives of Psychiatry,* vol. 25, no. 5, pp. 463-468, 1972.

[82] S. Panwai and H. Dia, "Comparative Evaluation of Microscopic Car-Following Behavior," *IEEE Transactions on Intelligent Transportation Systems,* vol. 6, no. 3, pp. 314-325, 2005.

[83] A. Buck, "Cointegration and Error Correction," 1999. [Online]. Available: https://astro.temple.edu/~buck/notes/cointegration/cointegration.HTM. [Accessed March 2018].

[84] J. Pinheiro and D. M. Bates, Mixed-Effects Models in S and S-PLUS, New York: Springer, 2000.

[85] P. Chaovalit, A. Gangopadhyay, G. Karabatis and Z. Chen, "Discrete wavelet transform-based time series analysis and mining," *ACM Computing Surveys,* vol. 43, no. 2, p. 6, 2011.

[86] D. Percival and T. Walden, Wavelet Methods for Time Series Analysis, Cambridge, UK: University Press, 2000.

[87] Olden J. and D. Jackson, "Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks," *Ecological modellin,* vol. 154, no. 1, pp. 135-150, 2002.

[88] A. Zeileis, C. Kleiber, W. Kraemer and K. Hornik, "Testing and Daing of Structural Changes in Practice," *Computational Statistics and Data Analysis,* vol. 44, pp. 109-123, 2003.

[89] T. Triggs and W. Harris, "Reaction Time of Drivers to Road Stimuli," Monash University, Victoria, Australi, 1982.

[90] F. Guo and F. Youjia, "Individual driver risk assessment using naturalistic driving data," *Accident Analysis and Prevention,* vol. 61, pp. 3-9, 2013.

[91] The R Development Core Team, "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, 2013. [Online]. Available: http://www.R-project.org. [Accessed April 2013].

[92] A. Sarda-Espinosa, "Package 'dtwclust'," 2018.

[93] D. Bates, M. M., B. Bolker and W. S., "Fitting Linear Mixed-Effects Models using lme4," *Journal of Statistical Software,* vol. 67, no. 1, pp. 1-48, 2015.

[94] P. Tormene, T. Giorgino, S. Quaglini and M. Stefanelli, "Matching incomplete time series with dynamic time warping: an algorithm and an application to post-stroke rehabilitation," *Artificial Intelligence in Medicine,* vol. 45, no. 1, pp. 11-34, 2009.

[95] A. Zeilieis, F. Leisch, K. Hornik and C. KLeiber, "An R Package for Testing Structural Change in Linear Regression Models," *Journal of Statistical Software,* vol. 7, no. 2, pp. 1-38, 2002.

[96] G. Box, G. Jenkins, G. Reinsel and G. Ljung, Time Series Analysis: Forecasting and Control, Fifth Edition, John Wiley and Sons, 2015.

[97] G. Tiao and G. Box, "Modeling Multiple Time Series with Applications," *Journal of American Statistical Association,* vol. 76, no. 376, pp. 802-816, 1981.

[98] H. Byun and S. Lee, "Applications of support vector machines for pattern recognition: a surve," Niagara Falls, Ontario, Canada, 2002.

[99] R. Bracewell, The Fourier trasform and its applications. Vol. 3, New York, NY: McGraw-Hill, 1986.

[100] S. Li and A. Jain, "Dynamic Time Warping," in *Encyclopedia of Biometrics*, Springer, 2009.

[101] T. Giorgino, "Computing and Visualizing Dynamic Time Warping Alignments in R. The dtw Package," *Journal of Statistical Software,* vol. 31, no. 7, pp. 1-24, 2009.

[102] G. Garson, "Interpreting neural network connection weights," *Artificial Intelligence Expert,* vol. 6, pp. 47-51, 1991.

[103] American Association of State Highway and Transportation Officials, Highway Safety Manual, Washington, DC: AASHTO, 2010.

[104] K. Green, G. Mcgwin and C. Owsley, "Associations between Visual, Hearing, and Dual Sensory Impairments and History of Motor Vehicle Collision Involvement of Older Drivers.," vol. 61, no. 2.

[105] N. C. f. S. a. Analysis, "Traffic Safety Facts," National Highway Traffic Safety Administration, Washington, DC, 2016.

[106] S. Lyman, "Older Driver Involvements in Police Reported Crashes and Fatal Crashes: Trends and Projections," vol. 8, no. 2, 2002.

[107] J. Cicchino and A. Mccartt, "Critical Older Driver Errors in a National Sample of Serious U.S. Crashes," vol. 80, 2015.

[108] M. Betz and S. Lowenstein, "Driving Patterns of Older Adults: Results from the Second Injury Control and Risk Survey," vol. 58, no. 10, 2010.

[109] D. Mayhew, H. Simpson and S. Ferguson, "Collisions Involving Senior Drivers: High-Risk Conditions and Locations," vol. 7, no. 2, 2006.

[110] K. Ball, C. Owsley, B. Stalvey, D. Roenker, M. Sloane and M. Graves, "Driving Avoidance and Functional Impairment in Older Drivers," vol. 30, no. 3, 1998.

[111] L. Molnar, J. Charlton, D. Eby, S. Bogard, J. Langford, Koppel, S., G. Kolenic, S. Marshall and M. Man-Son-Hing, "Self-Regulation of Driving by Older Adults: Comparison of Self-Report and Objective Driving Data," vol. 20, 2013.

[112] G. Li, E. Braver and L. Chen, "Fragility Versus Excessive Crash Involvement as Determinants of High Death Rates per Vehicle-Mile of Travel Among Older Adults," vol. 35, no. 2, 2003.

[113] J. Cicchino, "Why Have Fatality Rates Among Older Drivers Declined? The Relative Contributions of Changes in Survivability and Crash Involvement," vol. 80, 2015.

[114] Center for Transportation Research and Education, "SHRP 2 - Roadway Information Database," 2009. [Online]. Available: http://www.ctre.iastate.edu/shrp2-rid/. [Accessed 20 July 2016].

[115] W. N. Venables and B. D. Ripley, Modern Applied Statistics with S. Fourth Edition, New York: Springer, 2002.

[116] J. Fox and S. Weisberg, "An {R} Companion to Applied Regression, Second Edition," Sage, Thousand Oaks, CA, 2011.

[118] K. Dixon, K. Fitzpatrick and R. Avelar, "Operational and Safety Trade-Offs, Reducing Freeway Lane and Shoulder Width to Permit an Additional Lane," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2588, pp. 89-97, 2016.

APPENDIX

EXPLORING THE EFFECTS OF IMPORTANT PREDICTORS OF RAMP SPEEDING BEHAVIOR

Bahar Dadashova, Ph.D.[1]

Associate Transportation Researcher

Texas A&M Transportation Institute

3135 TAMU, College Station, TX 77843-3135

Email: b-dadashova@tti.tamu.edu

Phone: 979-845-7415

Fax: 979-845-6006

Karen Dixon, Ph.D., P.E.

Senior Transportation Engineer

Texas A&M Transportation Institute

3135 TAMU, College Station, TX 77843-3135

Email: k-dixon@tti.tamu.edu

Raul Avelar, Ph.D., P.E., PMP

Associate Research Engineer

Texas A&M Transportation Institute

3135 TAMU, College Station, TX 77843-3135

Email: r-avelar@tamu.edu

Phone: 979-862-1651
Fax: 979-845-6006

Word count: 7500 Words (5000 Text + 4 Tables + 6 Figures)

November 15, 2017

[1] Corresponding Author

1 Abstract

2   Traditional measures of speed obtained through traffic observations are not based on detailed
3 information about the related drivers and vehicles. Data from naturalistic studies, such as SHRP2 - NDS,
4 can mitigate this issue by combining the key data on driver, roadway and speeding behavior. The
5 objective of this study is to assess drivers' speeding behaviors on freeway ramps as the function of ramp
6 design, trip summary, and driver characteristics. The data analysis provides insights into various spatial
7 and temporal factors. To conduct the data analysis authors have implemented time series reduction,
8 matching and clustering methods to define a new speeding behavior response variable denoted as driving
9 States. Using the resulting response variable and the three groups of predictors, authors have conducted
10 neural network analysis to identify the most influential predictors and their effects on the speeding
11 behavior of drivers during on-ramp and off-ramp travels. Results of speeding behavior on freeway ramps
12 indicate that the speed choice at these locations is indeed a complex process and is mainly influenced by
13 the temporal and traffic conditions. Personal characteristics of drivers also were found to influence speed
14 choice in these locations.

15

16 **Key words:** Freeway Ramp, Naturalistic Driving, Human Factors, Traffic Conditions, Speed Choice, Time
17 Series Clustering, Discrete Wavelet Transformation, Dynamic Time Warping, Neural Networks.

1   Introduction

2   Freeway crashes frequently occur at on-ramp and off-ramp locations. These crashes account for 18% of
3   all interstate crashes, 17% of injury crashes and 11% of fatal crashes at interchange locations (*1*). A 2004
4   study by McCartt et al. found that about 50% of ramp-related crashes occurred while the vehicles were
5   exiting the freeway, 36% occurred while the vehicles were entering the freeway, and 16% occurred at
6   the midpoints of access roads (*2*). They also observed that 48% of the crashes were run-off-road
7   crashes, and 36% of them were rear-end collision. These findings suggest that speed adjustments that
8   occur at interchange locations, such as freeway-to-ramp transitions, may be associated with an increase
9   in crashes. Kim et al. found that 85% of all freeway rear-end crashes occurred within 2000 feet of the
10   on-ramp gore (*3*). This study found that there was a strong association between rear-end crash rates
11   and deceleration rate. Overall, speed indicators as well as the acceleration rates have been found to be
12   important predictors of highway safety (*4, 5*, *6, 7, 8*).

13   Several researchers have analyzed speeding behavior at ramps as an important roadway design factor,
14   but few of these studies have analyzed the merging and diverging process of drivers as the result of their
15   personal characteristics (*3*, *9*). Speeding behavior is a complex process that can be influenced by several
16   factors: roadway characteristics, traffic conditions, driver characteristics, vehicle dynamics, and weather.
17   Analysis of speeding behavior on freeway ramps might become even more challenging due to the
18   sophisticated design of these sites. The roadway characteristics, speed limits, and other drivers have
19   been found to be among the most influential predictors attributed to speeding behavior (*10*, *11*).
20   Personality characteristics that have been found to be predictive of excessive speeding behavior include
21   conscientiousness, reward sensitivity, sensation seeking tendencies, road rage, inattention, and risk
22   perception (*12, 13, 14, 15, 16*).

23

24   Research Objectives and Methodological Approach

25   The goal of this paper is to evaluate the speeding behavior of drivers during the freeway merging and
26   diverging activities. The secondary objective of this paper is to explore how well the speeding behavior
27   reflects the driver's personal characteristics and surrounding factors. For this purpose, the authors used
28   trip time series data from the Naturalistic Driving Study (NDS) of the Second Strategic Highway Research
29   Program (SHRP2).

30   In the present study, authors hypothesize that there are three types of speeding, labelled as States
31   (State I, State II and State III). Drivers are assumed to adapt one of the three types of driving State during
32   each trip including both on-ramp (street to freeway - SF) and off-ramp (freeway to street - FS) trips.
33   Driver's choice of the speeding State depends on multiple factors such as the ramp design, direction of
34   travel, driver characteristics, time and day of trip, and other trip details including speed variance, speed
35   mean, and number of trips per driver at each study location. The authors used a three-step procedure to
36   conduct time series classification and analyze the impact of external and personal factors on the
37   speeding behavior (Figure 32).

38   Several methods exist for conducting the time series feature extraction and classification analysis. These
39   include wavelet transformation, Fourier transformation K-nearest neighbor, piecewise linear
40   approximation, support vector machines, artificial neural networks and so on (*17, 18, 19, 20, 21*). In this

1 study researchers have applied wavelet transformation, dynamic time warping and neural networks to
2 conduct the data analysis as explained below.

3 As the first step, authors applied wavelet transformation to reduce the time series data dimension to
4 obtain a relatively balanced times series data. Using the Dynamic Time Warping tools, authors then
5 clustered the on-ramp and off-ramp speed traversals into three States of driving behavior. As the final
6 step, authors conducted the joint analysis of driving States using neural networks architecture.
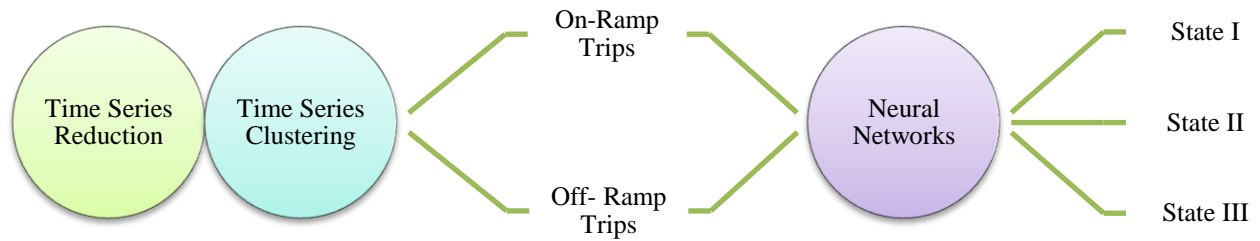
7
8 **Figure 32. Ramp speeding behavior analysis framework**
9

10 The rest of this paper is organized as follows. In the next section the SHRP2 dataset is described briefly.
11 This section is followed by the description of the statistical methods and tools implemented in this
12 study. In the third section the data analysis results are discussed. The paper ends with the conclusions,
13 acknowledgments and references.

14

15 SHRP2 Data Description

16 The SHRP2 – NDS dataset includes data that represent more than 3,500 volunteer drivers age 16 to 98.
17 The data collection duration extended over a four-year study period (2010-2013). The SHRP2 – NDS
18 study includes data from various locations in six states in the United States: Indiana (IN), New York (NY),
19 North Carolina (NC), Washington (WA), Pennsylvania (PA) and Florida (FL). NDS data is complemented
20 by a Roadway Information Database (RID) that contains detailed information for roadways in the study
21 regions. NDS trip data can be linked to the RID roadway data using the Link ID – a unique road segment
22 identifier, variable. The integrated SHRP2 data can provide researchers with an abundance of
23 information such as the driver and roadway characteristics, vehicle speed information over distance and
24 time, and other congestion-related factors including the time and day of participant trips.

25

26 NDS Traversals

27 In this study, the authors used NDS trip information (or *traversals*) collected from ramps at four
28 interchanges on two freeway sections in Altoona, PA. For the analysis of speeding behavior authors
29 considered the state where the enforcement of the speed traffic law was not very strict to diminish the
30 impact of this factor on drivers' speeding behavior (*22*). Data were obtained from four interchanges on
31 two freeways (Figure 33):

32 • Urban freeway: William Penn, Blair County.

1   •   Rural freeway: Bud Shuster Freeway, Blair County.

2   The length of each ramp located on these freeway interchanges is approximately 2 miles long which
3   include locations on the intersecting freeway and street, and the ramp connecting the two. To fulfill the
4   objectives of this study, the authors identified trips where each driver had traversed the same ramp in
5   both directions of travel. These trips are identified as:

6   •   Merging: Street to Freeway (SF)
7   •   Diverging: Freeway to Street (FS)
8

Bud Shuster HW and 17th St (Urban)
• Diamond Interchange

Bud Shuster HW and PA-865 (Urban)
• Trumpet Interchange

William Penn HW and Tunnehill (Rural)
• Partial Cloverleaf Interchange

William Penn HW and West 2nd St (Rural)
• Trumpet Interchange

9
10
11   **Figure 33. Ramp trajectories.**
12

1   For the analysis of speeding behavior, authors used speeds recorded by the Global Position
2   System (GPS) transponder located in each study vehicle. The NDS trip time series provides both GPS
3   speed and the speed recorded by the vehicle's own network. However, since the speeds recorded by the
4   GPS also have the matching latitude and longitude information for every second, the authors elected to
5   use this data source as the primary indicator of speeding behavior.

6   To obtain a consistent study sample, the authors only selected trip data that spanned the entire
7   duration of the target freeway ramps with travel times of 60-70 seconds or greater. This produced 859
8   trips taken by 32 participants. In many cases, drivers travelled the study route multiple times over the
9   study period. The number of trips per driver for most of the drivers ranged from 2 to 70 trips, with the
10   average of 30 trips per driver. Two of the study participants regularly used one of these routes; they had
11   accumulated 341 trips (117 and 224 trips) which accounted for almost 40% of all trips. Such an
12   overrepresentation in the trip numbers can present a bias in the results where the speeding will be
13   analyzed as the function of driver characteristics among other explanatory factors. Therefore, due to the
14   potential for undue and biased influence in the results from these two drivers if all their trips were used,
15   the authors decided to reduce the number of trips by random sampling. For this purpose, authors assigned
16   0 and 1 to each trip randomly and selected the trips that were assigned value 1. As the result 53 and 77
17   trips were left. The final study data set used in this study included 256 SF and 392 FS trips.

18   Next, the authors assembled a trip summary for each trip. The associated data included the year,
19   month, weekday, time bin as well as each trip's maximum speed, mean speed, and speed variance. There
20   is no exact hour for the trips, instead each trip was assigned a three hours long time bins.

21   NDS Driver Characteristics

22   To explore how the individual characteristics of a driver may influence their speeding behavior, the
23   authors used the following driver information obtained through interviews and psychological testing of
24   the SHRP2-NDS participants:

25   • Driver demographics (age and gender).
26   • Barkley's Attention Deficit Hyperactivity Disorder (ADHD) screening.
27   • Risk perception score (RPS).
28   • Sleeping habits.
29   • Depth Perception.
30   Barkley's ADHD Screening Test

31   Individuals with attention deficit disorder (ADD) and ADHD are prone to frequent inattention
32   and distraction while performing tasks. In addition to inattention being a factor that is potentially
33   associated with speeding (*14*), a study by Quinn et al. noted that individuals with ADD and ADHD have a
34   higher likelihood to be ticketed for speeding (*23*). This observation leads to the potential hypothesis that
35   those who score high on Barkley's ADHD screening test may have higher speeding incidences than those
36   who score low on the test. The six items included in the Barkley's ADHD screening are:

37   1. Easily distracted
38   2. Difficulty organizing
39   3. Loses objects
40   4. Quick screen- difficulty waiting turn
41   5. Feels restless
42   6. Difficulty enjoying leisure activities.
43

1 Each of these six items is scored by the participant by using one of the following three answers: Never or
2 Rarely (1), Sometimes (2), and Often (3). The Barkley's ADHD score was then calculated using the
3 answers provided to these items (*24*).

4 Risk Perception Score (RPS)

5   Risk perception is well-documented in the literature to have a strong impact on speeding behavior
6 (*25*). Those with low risk perception tend to have a high perception of their individual driver control. As
7 part of this characteristic, these drivers tend to dismiss risks, exude a high self-confidence (especially
8 about their driving ability), and demonstrate unrealistic optimism (*26*). It is hypothesized that those with
9 low risk perception (which equates to a low RPS score) will have higher number of speeding incidents
10 than those who have a high RPS score. The list of the questions used for the RPS is shown in Table 18.
11 Respondents answered to the questionnaire by assigning No Greater Risk (1), Moderately Greater Risk
12 (2) and Greater risk (3) to each question.

13

14 **Table 17. Elements of Risk Perception questionnaire.**

• Running red light	• Yellow light acceleration
• Risks for fun	• Driving after taking drug or alcohol
• Sudden lane changes	• Road rage
• Running stop sign	• Driving to reduce tension
• Speeding for thrill	• Passenger interaction
• Tailgating	• Racing
• Illegal turn	• Speeding <20> mph over limit.
• In a hurry	• Not wearing safety belt
• Risk of passing on right	• Risk perception score

2- No Greater risk; 4- Moderately Greater Risk; 7- Greater Risk.

15

16 Sleeping Habits

17   Previous research has found personality differences between long and short sleepers. Specifically,
18 short sleepers were found to be efficient persons that handle stress by keeping busy and by denial. Long
19 sleepers, on the other hand, had higher instances of depression and anxiety and scored higher on most
20 pathology tests (*27*). This portion of the evaluation is an exploratory analysis on both sleep schedule (i.e.,
21 whether the participant keeps a regular sleep schedule, Yes or No) and sleeper type (i.e., light, normal, or
22 heavy) to determine if these factors encompass a variety of traits related to speeding propensity.

23 Depth Perception

24   Depth perception is the ability to visually perceive the world in three-dimensional space and is
25 necessary to accurately determine the distance to an object. Depth perception is a "personal"
26 characteristic that directly affects an individual's visual perception, and therefore may be connected to
27 speeding behavior (or lack thereof).

28   Participants were shown a picture of four rings (top, bottom, left and right) and were asked if the
29 bottom ring seems to be floating towards them. If the participant answers yes, they moved to a second
30 picture and were asked which ring seems to be floating toward them. Drivers who cannot see the ring

1 floating towards them received no score. For all others, they were scored in seconds of arc, where the
2 smaller the seconds of arc, the better depth perception the participant has.

3

4 The list of the predictors used in the study is shown in Table 18.

5   **Table 18. Predictors used to analyze driving behaviors.**
6

Predictor Group	Variable Name	Variable Type	Descriptive Statistics			
			Min	*Max*	*Mean*	*St. D.*
Ramp Characteristics	Area Type	Categorical	1 = Rural, 2 = Urban			
	Travel Direction	Categorical	1 = Freeway to Street, 2 = Street to Freeway			
	Interchange ID	Categorical	1 = Bud Shuster HW and 17th St, 2 = Bud Shuster HW and PA-865, 3 = Will Penn HW and Tunnehill St., 4 = Will Penn HW and West 2nd St			
	Ramp Design	Categorical	1 = Diamond; 2 = Trumpet; 3 = Partial Cloverleaf			
Trip Summary	Trip Year	Numerical	1 = 2011; 2 = 2012; 3 = 2013			
	Trip Month	Categorical	1 = January 2 = February, …, 12 = December			
	Trip Week Day	Categorical	1 = Monday, …, 7 = Sunday			
	Trip Time	Categorical	Three hours bin (8 bins)			
	Maximum Speed	Numerical	41	80	61.9	5.8
	Mean Speed	Numerical	5.3	60.4	41.3	6.7
	Speed Variance	Numerical	3.5	28.1	12.8	5.1
	Vehicle Classification	Categorical	1 = Car, 2 = SUV, 3 = Cross-over, 4 = Minivan			
Driver Characteristics	Age	Categorical	1 = 16-19; 2 = 20-24, 3 = 25-29, 4 = 30-34, 5 = 35-39, 6 = 45-49, 7 = 50-54, 8 = 55-59, 9 = 60-64, 10 = 65-69, 11 = 70-74			
	Gender	Categorical	1 = Female, 2 = Male			
	Barkley Score	Numerical	1	6	2.02	1.06

Risk Perception Score	Numerical	1	126	74.05	38.6
Sleeper Type	Categorical	1 = Light; 2= Normal; 3 = Heavy			
Has a Sleep Schedule	Categorical	0 = No; 1 = Yes			
Depth Perception	Categorical	1 = B/400, 2 = L/200, 3 = B/100, 4 = T/70, 5 = T/50, 6 = L/40, 7 = R/30, 8 = L/25, 9, = R/20			

1

2 Statistical Methodology

3 As it was indicated earlier, the spatial distance completed by all traversals is similar between study sites
4 (approximately two miles) since all the ramps included in the study had approximately the same length.
5 However, because of the speed differentials, the trip durations vary hence resulting in unbalanced time
6 series data. Therefore, as an initial step for the time series dimension reduction evaluation, the authors
7 used the Discrete Wavelet Transformation (DWT) method to obtain a relatively balanced time series.

8   The second step of the statistical analysis involves the matching and clustering of the reduced
9 traversal sequences. Time series clustering can be conducted in several ways. In this study, the authors
10 used the Dynamic Time Warping (DTW) algorithm to cluster the reduced time series data into three
11 driving state clusters: State I, State II, and State III. Given that the on-ramp and off-ramp speeds have two
12 different shapes (e.g. when travelling from freeway to street) the time series clustering might become
13 challenging. Therefore, authors have clustered two sets of speed time series data based on the direction of
14 the trips: freeway to street trips (FS) and street to freeway trips (SF).

15   The clusters or driving states are assumed to be the three levels of the speeding State response
16 variable. After identifying the response variables, the authors then used Neural Networks (NN) to
17 evaluate the impact of the following group of factors on the speeding behavior:

18   • Ramp Characteristics.
19   • Driver Characteristic.
20   • Trip Summary.
21

22 Although the clustering is conducted for the on-ramp and off-ramp time series separately, authors have
23 conducted the joint NN analysis using the results from both sets of clusters and have used the travel
24 direction as one of the predictors.

25 Discrete Wavelet Transformation

26 As previously indicated, due to the speed differentials, the length of the series (i.e. number of seconds)
27 included in this study ranged between 70 to 850 seconds. Such a significant inequality among the series
28 lengths could cause a concern. This problem can be dealt with by reducing the time series dimension
29 (length-wise) using the algorithms such as Discrete Wavelet and Discrete Fourier transformation.

1   In this study, the authors used the DWT to reduce the dimension of relatively longer series.
2   Namely, the Haar wavelet is used to conduct the dimension reduction. Haar wavelet allows the time series
3   of length T to be represented in terms of its orthonormal basis by calculating a set of averages and
4   coefficients (usually $\sqrt{2}$ to ensure energy conservation) ($28, 29$). After the first iteration, the time series
5   length reduces to $T/_2$. The resulting time series is referred as the first level of wavelet transformation.
6   The DWT method can be applied recursively until a single coefficient and average is obtained.

7   Dynamic Time Warping

8   To match and cluster the reduced trip time series of varying lengths, the authors used the DTW
9   algorithm. DTW is essentially a point-to-point matching method under some boundary and temporal
10   consistency constraints. DTW was originally developed for speech recognition (30). The algorithm aims
11   to match two time series vectors by warping the time axis iteratively until the optimal alignment is
12   achieved. Optimal alignment in this case is measured by the time warping distance (31).

13   DTW clustering can be conducted based on the distance or shape of the series. In this study the
14   authors have implemented the Partitioning Around the Medoids (PAM) algorithm. The PAM algorithm is
15   based on the principle of finding the sequence that is in the cluster center. The members of the given
16   cluster are then selected based on the cutoff distance from the medoids. This method essentially optimizes
17   the matching by using both shape and distance measures.

18   Neural Networks

19   To describe the relationships between the predictors (inputs) and the speeding state (output), the authors
20   used the multi-layer perception (MLP) NN architecture trained by a backpropagation algorithm. NN
21   methods are known for their ability to deal with a relatively large number of predictors. The NN
22   framework or architecture has three elements: input, hidden layer and output (Figure 34). Input and output
23   refer to the predictors and response variable respectively. Hidden layers are the collection of neurons
24   organized and connected to each other using the arrows that are referred to as the weights. Weights can
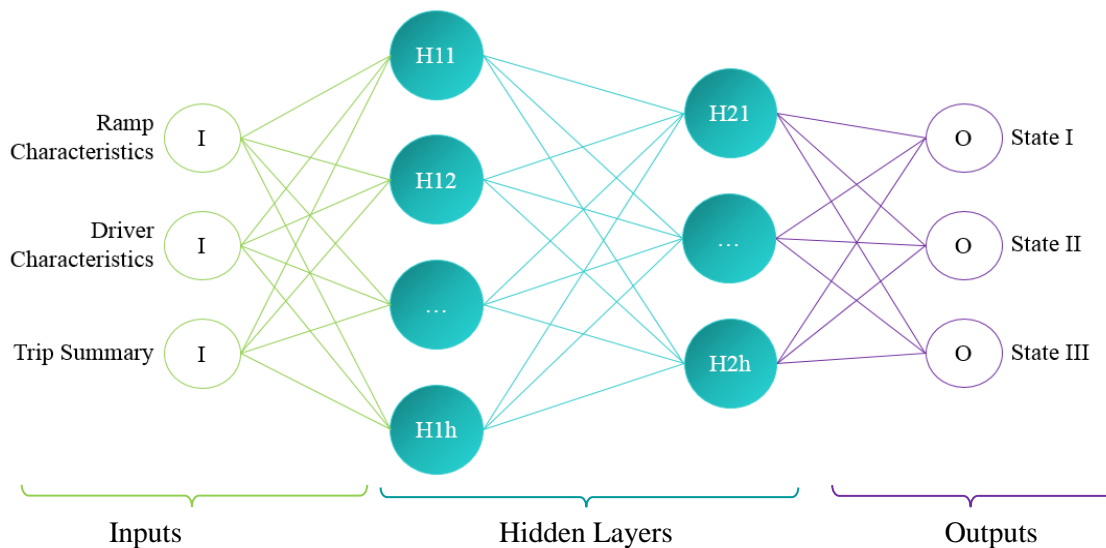25   also be understood as the parameter estimates although they should not be interpreted as such (32).

Figure 34. Neural Networks framework.

1

2   In the NN architecture, the relative contribution of the inputs to the output depends on the magnitude
3   and the direction of the connection weights (*33*). Connection weights are computed using the weights
4   of individual inputs in each hidden layer. Greater connection weight indicates the higher intensity of the
5   association. Negative connection weights represent an inhibitory (reducing) effect while the positive
6   connection weight represent an excitatory (increasing) effect of the neurons on the output. The NN
7   architecture is depicted in Figure 4.

8

9   Result and Discussion

10   Time Series Reduction and Matching

11   To reduce the dimension of the speed series and create a relatively balanced set of traversals, the authors
12   recursively applied the DWT method. Each series was saved together with all levels of wavelet
13   decompositions. After reviewing the resulting reduced sequences of the trip time series, the authors
14   selected the reduced series of the lengths ranging from $20 < Seconds \leq 41$. An example of the SF and
15   FS trips completed by two drivers (D-171263 and D-639928) is depicted in Figure 4.
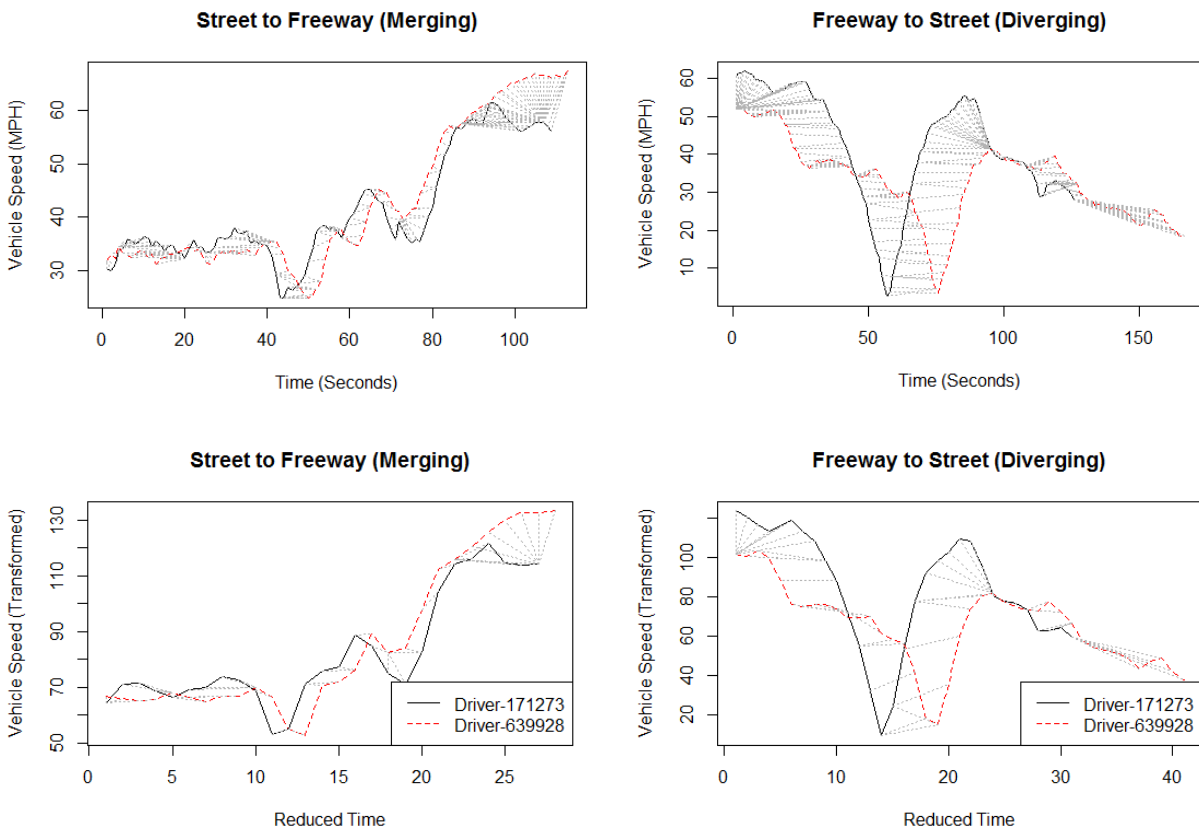

Figure 35. Matching two traversals: Original vs DW – Transformed series.

16   All four trips (2 SF and 2 FS) were completed on Ramp #3. After applying the recursive wavelet
17   transformation in two iterations, the FS speed series of the two drivers have been reduced from 126 and

1   167 seconds to 31 and 41 seconds respectively. The SF trips, after two iterations, have been reduced from
2   109 and 113 seconds to 27 and 28 seconds respectively. The transformed values represent the trip speeds
3   after they have been transformed twice. For example, if driver D-171263 has the following speeds at the
4   initial four seconds of the SF trip: $S^0{}_{t,t=1,2,3,4} = (30.5, 30.02, 31.31, 33.7)$ where the superscript 0
5   indicates the initial level (Figure 35, upper-left graph), this time series will be reduced to a sequence of 3
6   values after the 1st level wavelet transformation : $S^1{}_{t,t=1,2,3,} = (30.26\sqrt{2}, 30.66\sqrt{2}, 32.52\sqrt{2}) =$
7    $(42.79, 43.36, 45.99)$ where $30.26 = \frac{30.5+30.02}{2}$ and so on. The rationale of applying the coefficient $\sqrt{2}$ is
8   to "conserve the energy" (see [31] for more information on this subject). After the next iteration, the
9   second level wavelet transformation for this sequence will be equal to: $S^2{}_{t,t=1,2} =$
10    $(43.08\sqrt{2}, 44.68\sqrt{2}) = (60.92, 63.19)$ where $43.08 = \frac{42.79+43.36}{2}$, and so on (Figure 35 lower left
11   graph). As can be observed, no information is lost, and the shape of the series are not altered after the
12   wavelet transformations.

13   Speed Profile Clusters

14   The authors clustered the reduced speed profiles into three states using the PAM algorithm: State I, State
15   II and State III. The clustering was conducted using the R-CRAN open software (34). As shown in Table
16   19, 242 FS trips and 129 SF trips were assigned to State I, 19 FS trips and 45 SF trips were assigned to
17   State II and 131 FS trips and 82 SF trips were assigned to State III.

18

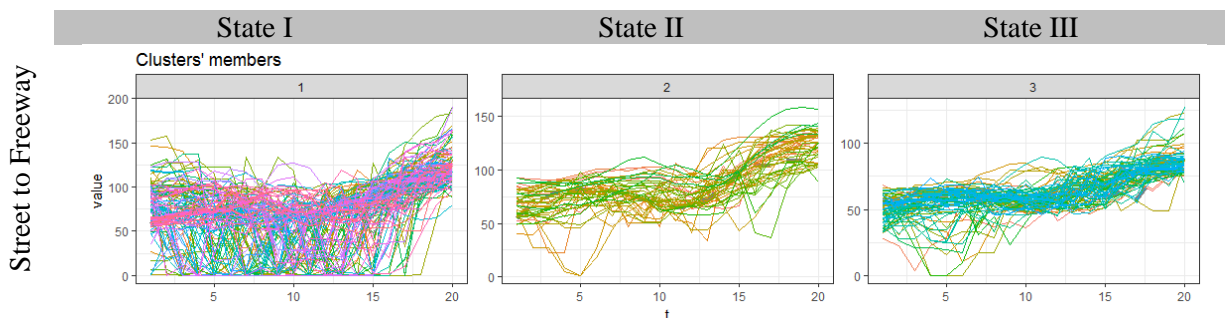19   **Table 19. Summary of speed profile clusters.**

Clusters	Freeway to Street		Street to Freeway	
	# of series	*% of total*	*# of series*	*% of total*
State I	242	62%	129	50%
State II	19	5%	45	18%
State III	131	33%	82	32%
Total	392	100%	256	100%

20

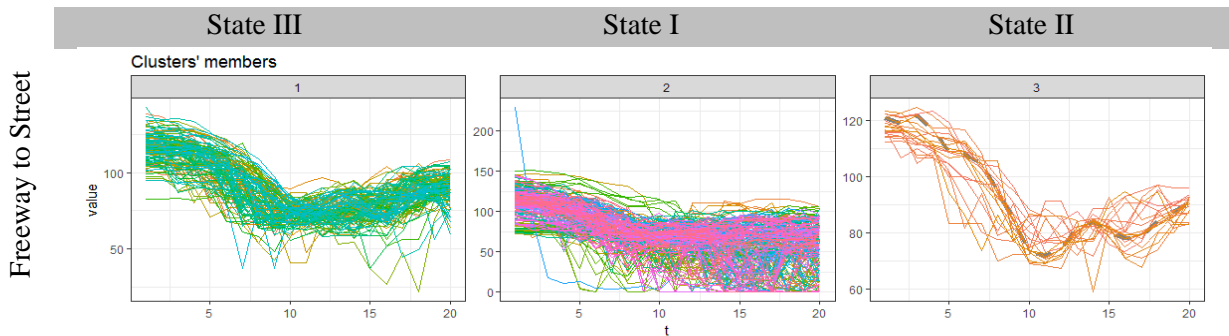21   The speed profile clusters can be visualized in Figure 36.

22

Figure 36. Speed profile clusters.

1

2   Since the time series data were reduced using the wavelets, the values shown on the *y-* axis do not
3   represent the actual speeds but rather the transformed ones. Also, it should be noted that the clusters show
4   the individual trip series, meaning that the same driver could belong to all three clusters. However, it is
5   assumed that over a longer period, the true behavior of the driver will surface as additional trips occur.
6   This is also one of the reasons why all the trips need to be included in the clustering analysis. On the
7   contrary, the analysis might not be able to capture the true nature of the driver's behavior, as there is no
8   way to assure convergence to the "long term behavior".

9   At first inspection, it can be observed that the speed series belonging to State I (purple colored)
10   are highly volatile, some of the values reaching to zero (coming to a complete stop). Speed profiles
11   assigned to State II (green and yellow colored) are still relatively volatile; however, the speeds rarely
12   reach a value of zero. The speed series assigned to State III (green and turquoise colored), on the other
13   hand, exhibit smoother speed transitions when travelling from one location to another.

14   Observing the overall volatility of the series belonging to each cluster, one explanation might be
15   that at State I the drivers are less cautious because there are abrupt changes and sudden stops in the speed
16   profiles. However, their driving behavior can also be attributed to congestion, i.e. the vehicles might be
17   stopping due to congestion. It seems like at States II and III the drivers could be more cautious (or under
18   less congested conditions), although at State II the drivers continue making abrupt speed changes. At
19   State III the drivers seem very cautious because the transition from freeway to street and from street to
20   freeway appears to be quite smooth. Yet, this smooth transition could also be indicating minimal or
21   inexistent congestion or opposing traffic. For the following analysis, the authors review and interpret the
22   results under the hypothesis that an increasing State level indicates a more cautious behavior by the
23   drivers. Note that cautious driving only refers to the speeding pattern of the drivers and should not be
24   interpreted as safe or risky driving.

25   Results of Network Analysis

26   The authors used 47 inputs to assess the three States of speeding behavior on the ramps using the NN
27   architecture. These inputs belong to the following categories: ramp characteristics, driver characteristics,
28   and trip characteristics. To conduct the NN analysis, the data was sampled into training and testing
29   datasets. The NN is trained using 70% of observations connected through 15 hidden neurons. The testing
30   data (remaining 30% of observations) was then predicted using the trained algorithm. The mean squared
31   errors showing the difference between the predicted and observed data was estimated to be 2.5%, 2.0%,
32   and 3.4%, for the States I, II and III respectively. This observation indicates a highly accurate prediction
33   of each output.

1   The relative importance of variables was found by using the Garson's algorithm (*35*). Garson's
2 algorithm uses the absolute values of connection weights to determine the relative importance of inputs in
3 the network. This relative importance was calculated for the entire network and not for each State
4 individually. It shows the individual contributions of each input to the network. The sum of all
5 contributions will be equal to 100%. The importance of each variable is shown in Table 20. As shown, the
6 speed variance is the most important predictor of the speeding State response variable with almost 7%
7 contribution. Trip weekday, mean speed, maximum speed, trip time, and travel direction each contribute
8 3% or more to the entire network. In general, the importance diagram implies that the trip time and date
9 and the locations of the trips are the main contributors to the speeding behavior. Therefore, it can be
10 concluded that the speeding behavior is mainly influenced by the traffic conditions rather that the driver
11 characteristics.

1

Table 20. NN Connection weights and importance of predictors of speeding behavior.

Variable Name	Importance (% contribution)	Overall Connection Weights			Variable Name	Importance (% contribution)	Overall Connection Weights		
		State I	State II	State III			State I	State II	State III
Speed Variance	6.3%	37.04	21.83	1.95	Risks for Fun	2.0%	7.96	4.70	2.93
Trip Week Day	4.0%	7.15	-7.68	-14.00	Difficulty Enjoying Leisure Activities	2.0%	18.04	-2.76	-3.79
Mean Speed	3.9%	-43.98	17.03	21.73	Sudden Lane Change	2.0%	-8.70	0.76	-4.78
Maximum Speed	3.4%	15.22	26.80	-1.95	Losses Things	2.0%	14.02	-8.08	4.92
Trip Time	3.0%	-16.26	-15.34	6.23	Feels Restless	1.9%	5.91	6.06	9.15
Travel Direction	3.0%	-11.41	-26.51	22.14	Risk Perception Score	1.9%	0.49	-5.24	26.23
Not Wearing Seatbelt	2.9%	-10.77	12.78	8.22	Area Type	1.9%	-3.94	-5.65	-2.37
Trip Year	2.8%	14.48	-14.92	-10.40	Vehicle Classification	1.9%	-12.25	6.91	-7.86
Sleeper Type	2.8%	11.95	15.77	-5.43	Road Rage	1.9%	-0.25	3.95	4.45
Interchange ID	2.7%	-12.12	-4.39	-3.48	Passenger Interaction	1.9%	-1.88	3.17	4.97
Ramp Design	2.5%	3.38	-7.61	0.03	Speeding < 20 mph	1.8%	-9.18	-7.66	15.32
Trip Month	2.4%	-13.20	8.60	-28.88	Passing on Right	1.8%	-0.10	5.43	-0.80
Accelerates at Yellow Light	2.4%	5.56	9.42	-15.27	Driving to Reduce Tension	1.7%	-0.69	14.70	17.59
Age	2.4%	-1.39	-4.44	8.65	Difficulty Awaiting Turn	1.7%	-0.04	11.17	-2.30
Speeding > 20 mph	2.3%	0.82	12.75	2.88	Rolling Stop Sign	1.7%	-8.33	2.50	1.90
Driving Under Influence	2.3%	4.56	5.42	-4.53	Number of Trips per Driver	1.7%	21.35	-10.84	-5.69
Road Racing	2.2%	16.54	10.01	4.81	Speeding for Thrill	1.5%	-1.12	0.20	1.74
Depth Perception	2.2%	-9.09	4.76	-1.36	In a Hurry	1.5%	3.11	4.48	-10.62
Sleep Schedule	2.2%	14.38	19.27	11.42	Running Red Light	1.5%	-14.24	8.60	4.89
Tailgating	2.1%	0.52	11.36	-8.95	Illegal Turns	1.3%	-8.82	0.80	-1.44
Difficulty Organizing	2.1%	-9.44	14.48	-4.12	Gender	1.3%	2.30	3.06	20.53
Easily Distracted	2.1%	28.11	-15.44	-10.07	Barkley Score	1.1%	12.60	0.54	5.69

2

1   To explore the direction and the magnitude of the input effects on the outputs, it is helpful to
2   examine the connection weights (Table 20). The connection weights are calculated for each input by
3   using the weights from all hidden neurons (15 in total). The results were obtained for all three outputs
4   which are the speeding states (i.e. three states of speeding behavior).

5   Figure 37 graphically represents the relationships between the driving States and some of the most
6   important factors.

7
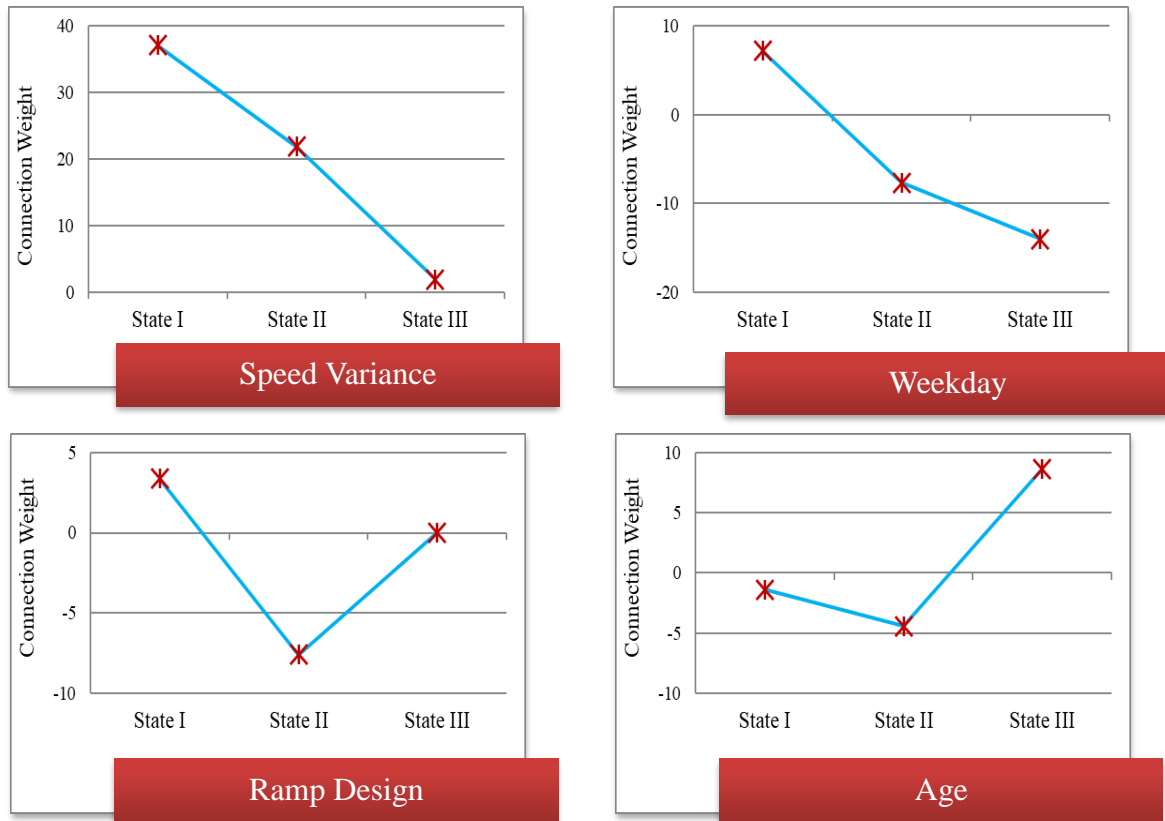
9   **Figure 37. Description of relationships between the driving states and important variables.**
10

11   The magnitude of speed variance effect on speeding behavior is 37.04, 21.83 and 1.95 for the States I, II
12   and III respectively. Connection weights for speed variance indicates that this variable has an increasing
13   impact (positive sign) on the speeding behavior, although as drivers shift from State I to State III, the
14   magnitude of this effect starts to decrease. This finding is consistent with the initial assumption that as
15   the drivers move from State I to State III, they start driving more cautiously thus the magnitude of the
16   effect of this variable is decreasing as the states progress in increasing order. The two other speed
17   related inputs, mean and maximum speeds, are also important. It can be observed that in State I the
18   mean speed is decreasing, this is expected because some drivers come to a complete stop. As the
19   drivers move from State I to State III, the mean speed is increasing. This finding, together with the effect
20   of speed variance on the speeding states, yields very interesting results. As shown, the effect of speed
21   variance is smaller in State III compared to the State II however the effect of mean speed in State III is

observed to be much higher compared to State II. This would imply that the speed variance is perhaps a reliable indicator of cautious driving behavior.

The second most important variable group is the time and date of the trips. For State I, the effect of week day is increasing while for the States II and III the effect is decreasing. Note that the week day is in increasing order from Monday to Sunday. This implies that as the week starts to progress more drivers start driving in State I and less drivers drive in State III. Using the previous assumptions about the driving states, this result can be interpreted as follows: during the weekdays from Monday thru Friday drivers will be driving more cautiously and on the weekends they will drive less so.

Another interesting relationship is presented by the ramp design. The ramp design is ordered from least sophisticated (or complex) to more sophisticated. The importance of the influence of ramp design is moderate to high (ranked 11 out of 47 factors in Table 20). The effect of the ramp design on the State I is positive and its magnitude is the highest, suggesting that as the ramps become more challenging, drivers would tend to adjust their driving to the characteristics of the State I pattern. In State II, the effect is negative implying that less drivers will be driving in this form as the interchanges become more sophisticated. The relationship between State III driving and ramp design is positive although the magnitude of the effect is very small. This could imply that the speeding behavior of State III drivers is not affected by the ramp design significantly.

Among the driver characteristics the age, risk perception elements, and sleeping habits are observed to contribute more to the driving states. The results show that the driver age is negatively related to State I and II and positively related to State III, implying that as the drivers get older their driving patterns become more like the State III driving, and less like the State I and State II driving. However, one can observe that as the drivers get older they also driver less in State II and more in State I. This could indicate that the relationship between age and speeding behavior is in fact U-shaped and not linear. Further analysis will be necessary to explore the driving patterns in terms of young, middle aged and older drivers.

The RPS is positively related to State I and State III driving, although the magnitude for State I is very small implying an almost non-existent impact. The magnitude of the effect for State III is very high, which would imply that State III drivers have very high risk perception. For the State II drivers this effect is negative, implying that the drivers in this group have smaller risk perception. Analyzing the speed-related elements of the risk perception survey (i.e. accelerates at yellow light, speeding for thrill, or speeding over the speed limit), one can observe that the State III drivers seem to perceive the risks related to these behaviors poorly when compared to State I drivers. For example, State III drivers do not perceive accelerates at yellow light as risky as the State I and State II drivers. It can also be observed that State II drivers show highest concern to this behavior. If the State II driving can be interpreted as "moderate" driving, results show that most drivers might assume the accelerating at yellow light as a risky behavior. These results indicate that there might be a U-shaped relationship between risk perception elements and speeding behavior. More risky behaviors (driving under influence, tailgating, sudden lane changes, and passing on right), on the other hand, are perceived as highly risky by State III drivers. Further research will be needed to explore the relationships between the risk perception and speeding behaviors.

The Barkley's ADHD score was not found to be important although individual elements of the survey were found to be quite important (easily distracted and difficulty organizing). It can be observed that the drivers belonging to State II and State III tend to be less prone to be distracted, perhaps because they tend to be more aware of their surroundings while at State I, the drivers are more prone to be distracted. Moreover, the State III drivers are found to have lower Barkley score than the State I drivers.

1 This is consistent with the initial hypothesis that drivers scoring less in Barkley's ADHD test were
2 expected to have less speeding incidences.

3 Conclusions

4 In this paper the authors have conducted a speed profile analysis of a sample of NDS participants at
5 freeway ramps in Pennsylvania. The authors focused the data analysis using three years of trip data from
6 four ramp locations. To analyze the speeding behavior, the authors used 47 predictors indicating the
7 ramp characteristics, driver characteristics, and trip summary.

8   To perform this evaluation, the authors conducted the data analysis in three stages. The authors
9 implemented the data reduction and clustering of the speed time series to distinguish the three types of
10 driving behaviors named as States. Using the States as the three levels of the speeding behaviour, the
11 authors then developed NN architecture to explore the relationships between the 47 inputs and three
12 outputs (one output per State). Overall the results of the NN analysis uncovered interesting relationships
13 between the ramp speeding behavior and the other explanatory. The results suggest that the traffic
14 conditions and the geographic location are the main predictors of a driver's speeding behavior. Based on
15 the influence of these factors, drivers adjust their speed in a more cautious manner.

16   The impact of driver characteristics was not found to be as important. Although the results of the
17 NN analysis did reveal some interesting facts. For example, the analysis of the driver's risk perception
18 survey revealed that the drivers who drove in a more cautious manner and adjusted the speeds more
19 smoothly (State III) did not perceive driving at higher speeds or accelerating as risky behaviors. Although
20 they did perceive some other risky behaviors such as the driving under influence or sudden lane changes
21 as risky behavior. These drivers were also found to be considering both traffic conditions and driver
22 characteristics, it is quite likely that the drivers in State I, who were assumed to be less cautious in the
23 analysis, might have adjusted their speeds and come to a complete stop simply due to traffic conditions.
24 When examining the answers of these drivers to the risk perception questionnaire, they were observed to
25 have higher risk perceptions as far as the speeding-related questions were concerned. They were also
26 observed to be easily distracted based on their Barkley ADHD scores. These results might explain why
27 the drivers observed in this driving State have stopped more times compared to the other clusters; they
28 would be more likely to be distracted and might have had to stop or reducing the speed significantly more
29 often when encountering highly risky situations. As future improvement on this work, the authors would
30 like to evaluate more ramp trips to define and explain speeding behavior of drivers at these locations with
31 more accuracy. This future work will also consider analysing the effects of risk perception elements in
32 batches (e.g. speed related questions) rather than evaluating the impact of each element separately.

33 Acknowledgements

38

1   References

2

[1] Firestine, M., H. McGee, and P. Toeg. Improving Truck Safety at Interchanges. Report No. FHWA-IP-89-024. FHWA, 1989.

[2]. McCartt, A.T., V.S. Northrup, and R.A. Retting. Types and characteristics of ramp-related motor vehicle crashes on urban interstate roadways in Northern Virginia. Journal of Safety Research, Vol. 35(1), 2004, pp.107-114.

[3]. Kim, S. K., T. Song, N.M. Rouphail, S. Aghdashi, A. Amaro, and G. Goncalves. Exploring the association of rear-end crash propensity and micro-scale driver beahviour. Safety Science, Vol. 89, 2016, pp. 45-54.

[4]. Hauer, E. Speed and safety. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2103, 2009, pp. 10-17.

[5]. Elvik, R., P. Christensen, and A. Amundsen. Speed and road accidents. An evaluation of the Power Model. TØI Report 740, 2004.

[6]. Solomon, D. Accidents on main rural highways related to speed, driver, and vehicle. 1964.

[7]. Lave, C.A. Speeding, coordination, and the 55 mph limit. The American Economic Review, Vol. 75(5), 1985, pp.1159-1164.

[8]. Garber, N.J. and r. Gadiraju, R. Factors Affecting Speed Variance and Its Influence on Accidents. Transportation Research Record, Vol. 1213, 1989, pp. 64-71.

[9]. Kondyli, A. and L. Elefteriadou. Modeling driver behavior at freeway-ramp merges. Transportation Research Record: Journal of the Transportation Research Board, (2249), 2011, pp.29-37.

[10]. Haglund, M. and l. Åberg. Speed choice in relation to speed limit and influences from other drivers. Transportation Research Part F: Traffic Psychology and Behavior, Vol. 3(1), 2000, pp. 39-51.

[11]. Aarts, L. and I. Van Schagen. Driving speed and the risk of road crashes: A review. Accident Analysis & Prevention, Vol. 38(2), 2006, pp.215-224.

[12]. Jovanović, D., K. Lipovac, P. Stanojević, and D. Stanojević. The effects of personality traits on driving-related anger and aggressive behavior in traffic among Serbian drivers. Transportation research part F: Traffic Psychology and Behavior, Vol. 14(1), 2011, pp.43-53.

[13]. Scott-Parker, B., B. Watson, M.J. King, and M.K. Hyde. A further exploration of sensation seeking propensity, reward sensitivity, depression, anxiety, and the risky behavior of young novice drivers in a structural equation model. Accident Analysis & Prevention, Vol. 50, 2013, pp.465-471.

[14]. National Highway Traffic Safety Administration. Risky Driving. https://www.nhtsa.gov/risky-driving. Accessed on May 26, 2017.

[15]. Beanland, V., M. Selbom, and A.K. Johnson, Personality domains and traits that predict self-reported aberrant driving behavior s in a southeastern US university sample. Accident Analysis & Prevention, Vol. 72, 2014, pp. 184-192.

[16]. Bailey, S., A. Lennon, and B. Watson. Getting mad may not mean getting even: The influence of drivers' ethical ideologies on driving anger and related behavior. Transportation Research Part F: Traffic Psychology and Behavior, Vol. 36, 2016, pp.104-116.

[17]. H. Byun, H and S.W. Lee. Applications of support vector machines for pattern recognition: a survey. Proc. 1st Int. Workshop, SVM Pattern Recog. With Support Vector Mach., Niagara Falls, ON, Canada(2002), pp. 213-236.

[18]. Bracewell, R.N. The Fourier transform and its applications. Vol. 31999, 1986, New York: McGraw-Hill.

[19]. Li, Z., X. Jin and X. Zhao. Drunk driving detection based on classification of multivariate time series. Journal of safety research,Vol. 54, 2015, pp.61-e29.

[20] Liang, Y., L. M. Reyes, and J. D. Lee. Real-time detection of driver cognitive distraction using support vector machines." IEEE transactions on intelligent transportation systems, Vol. 8 (2), 2007, pp. 340-350.

[21]. Delen, D., R. Sharda, and M. Bessonov. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accident Analysis & Prevention, Vol. 38(3), 2006, pp.434-444.

[22]. Strictest and Most Lenient States on Speeding and Reckless Driving. *https://wallethub.com/edu/strictest-and-most-lenient-states-on-speeding/14211/*

[23]. Quinn, P. Road safety: Overcoming driving distractions. Additude-Strategies and Support for ADHD and LD. http://www.addiltudemag.com/adhd/article/904.html

[24]. Antin, J. F. Design of the in-vehicle driving behavior and crash risk study: in support of the SHRP 2 naturalistic driving study. Transportation Research Board, 2011

[25]. Jonah, B.A. Sensation seeking and risky driving: A review and synthesis of the literature. Accident Analysis & Prevention, Vol. 29, No. 5, 1997, pp. 651-665.

[26]. Thielen, I.P., R.C. Hartmann, and D.P. Soares. Risk perception and speeding. Cad Saude Publica, Vol. 24(1), 2008, pp. 131-139,

[27]. Hartmann, E., F. Baekland and G.R. Zwilling. Psychological differences between long and short sleepers. Archives of General Psychiatry, Vol. 25, No. 5, 1972, pp. 463-468.

[28]. Chaovalit, P., A. Gangopadhyay, G. Karabatis and Z. Chen. Discrete wavelet transform-based time series analysis and mining. ACM Computing Surveys (CSUR), Vol. 43(2), 2011, p.6.

[29]. Percival, D. B. and A. T. Walden. *Wavelet Methods for Time Series Analysis*, Cambridge University Press, 2000.

[30]. Sakoe,H. and S. Chiba. *Dynamic programming algorithm optimization for spoken word recognition*. IEEE Trans. on Acoustic, Speech, and Signal Processing, Vol. 26, 1978, pp. 43-49.

[31]. Li, S., and A. Jain. Dynamic Time Warping. Chapter in Encyclopedia of Biometrics Springer, 2009.

[32]. Karlaftis, M.G. and E.I. Vlahogianni. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. Transportation Research Part C: Emerging Technologies, Vol. 19(3), 2011, pp.387-399.

[33]. Olden, J.D. and D. A. Jackson. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecological modelling, Vol. 154(1), 2002, pp.135-150.

[34]. Giorgino, T. *Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package.* Journal of Statistical Software, Vol. 31(7), 2009, pp. 1-24.

[35]. Garson, G.D. Interpreting neural network connection weights. Artificial Intelligence Expert, Vol. 6, 1991, pp. 47-51.