

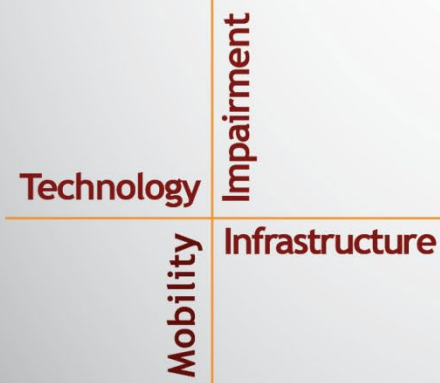
NSTSCCE

National Surface Transportation
Safety Center for Excellence

Analyzing Intersection Gap Acceptance Behavior with Naturalistic Driving Data

Yingfeng (Eric) Li • Haiyan Hao • Ron Gibbons •
Alejandra Medina

Submitted: September 14, 2022



Housed at the Virginia Tech Transportation Institute
3500 Transportation Research Plaza • Blacksburg, Virginia 24061

ACKNOWLEDGMENTS

The authors of this report would like to acknowledge the support of the stakeholders of the National Surface Transportation Safety Center for Excellence (NSTSCE): Tom Dingus from the Virginia Tech Transportation Institute, John Capp from General Motors Corporation, Chris Hayes from Travelers Insurance, Terri Hallquist from the Federal Motor Carrier Safety Administration, and Cathy McGhee from the Virginia Department of Transportation and the Virginia Transportation Research Council.

The NSTSCE stakeholders have jointly funded this research for the purpose of developing and disseminating advanced transportation safety techniques and innovations.

EXECUTIVE SUMMARY

Safety at unsignalized intersections continues to be a major concern for transportation agencies and roadway users. To improve intersection safety, this project conducted a comprehensive study of gap acceptance behaviors at unsignalized intersections using the second Strategic Highway Research Program (SHRP 2) naturalistic driving study (NDS) data. The project team conducted a number of analyses to achieve the research objectives. First, the team identified the critical gaps for a number of common scenarios using three widely accepted methods: binary logistic regression, maximum likelihood method, and probability equilibrium method. The team then went on to develop a complete understanding of the factors affecting gap acceptance decisions using logistic regression and machine learning techniques. Finally, researchers further investigated drivers' longitudinal and lateral acceleration behaviors during turning after accepting a gap and factors affecting their turning behaviors.

For the purpose of this project, the team collected 1,170 accepted and rejected gaps/lags based on 466 NDS trips at 60 unsignalized T-intersections in Washington state and North Carolina. To collect complete information about the involved trips, drivers, and intersections, the project team utilized a number of data sources, including time series data measuring vehicle kinematics for the analyzed trips, forward-facing and rear-view videos for the analyzed trips, driver demographic and driving history data, the SHRP 2 Roadway Information Database, and satellite images.

The following summarizes the major findings of this project:

- **Critical gap determination.** The study showed an overall critical gap of 5.3 seconds for right-turning trips and 6.2 seconds for left-turning trips. For right-turning vehicles, the presence of leading and/or following vehicles resulted in a 0.2-second longer critical gap (i.e., 5.4 versus 5.2 seconds). For left-turning vehicles, the presence of leading and/or following vehicles corresponded with a 0.4-second shorter critical gap (i.e., 5.9 versus 6.3 seconds). In addition, the presence of a two-way, one-lane, left-turn lane also resulted in a shorter critical gap for left-turning vehicles by 0.2 seconds (i.e., 6.1 versus 6.3 seconds).
- **Factors affecting gap acceptance decisions.** This analysis used three methods (logistic regression, decision tree, and random forest) to identify a comprehensive set of factors that affect gap acceptance decisions at intersections. The results showed a large number of factors that could have affected drivers' gap acceptance decisions. For example, being a gap instead of a lag, presence of leading and/or following vehicles, higher volume, intersection being unskewed, and increased number of through lanes all increased the probability of a driver accepting a given gap for multiple scenarios. On the other hand, adverse weather conditions reduced the probability of acceptance for a given gap. In addition, the modeling results suggested that a waiting time of longer than 9 seconds tended to increase the probability of gap acceptance.
- **Driver turning behaviors and impact factors.** During this project, the research team used Gaussian process regression to develop continuous vehicle kinematics profiles for

different turning behaviors, and multiple regression to identify factors affecting how drivers turn after accepting a gap. The maximum longitudinal acceleration for left-turning vehicles was approximately 2.0 m/s^2 (with an average acceleration rate of 1.25 m/s^2), higher than the average maximum longitudinal acceleration rate (i.e., 1.8 m/s^2 with an average rate of 1.53 m/s^2). Overall, both left- and right-turning vehicles initially accelerated quickly after they accepted a gap, and then reduced to a lower but prolonged acceleration rate while turning to reach a desired speed. The maximum lateral acceleration rate for left-turning vehicles was approximately 2.5 m/s^2 (with an average rate of 1.66 m/s^2), also slightly higher than that for right-turning vehicles (i.e., 2.4 m/s^2 with an average rate of 1.47 m/s^2). Although both lateral acceleration profiles resembled a parabolic curve, the peak value for the left-turning profile was reached later in the turning process than that for the right-turning vehicles.

Traditionally, gap acceptance has largely been studied using observational data collected from outside of vehicles based on video or manual timing at intersections. By using observational data, existing research has studied driver gap judgements from an external point of view, without taking into account the role of driver behavior. In addition, most previous studies have been based on a limited number of intersections. Depending on local traffic and driver characteristics, studies frequently found significantly different critical gap values at different intersections. The availability of the SHRP 2 NDS data has provided a unique opportunity to study driver gap acceptance behavior from a different angle. Naturalistic driving data can help provide a better understanding of this problem by allowing the study of driver behavior in an unobtrusive way. The detailed driver, trip, and vehicle kinematical data made possible by use of the NDS database was fully utilized in this study to provide new insights into the gap acceptance behaviors at unsignalized intersections.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION.....	1
LITERATURE REVIEW	2
<i>Factors Affecting Gap Acceptance Behavior</i>	2
<i>Critical Gaps and Critical Gap Determination</i>	4
<i>Critical Gap Values</i>	6
<i>Literature Review Discussion</i>	8
PROJECT OBJECTIVES	9
CHAPTER 2. METHODOLOGY AND DATA COLLECTION	11
DATA COLLECTION.....	11
<i>SHRP 2 Naturalistic Driving Data Overview</i>	11
<i>SHRP 2 Data Request and Processing</i>	12
<i>Data Collection and Preprocessing</i>	14
<i>Summary of Independent Variables and Collected Data</i>	16
METHODOLOGY	19
<i>Critical Gap Determination</i>	20
<i>Methods for Modeling Gap Acceptance Decision</i>	23
<i>Turning Behavior Modeling</i>	25
CHAPTER 3. ANALYSIS RESULTS AND DISCUSSION	29
DATA OVERVIEW.....	29
CRITICAL GAP DETERMINATION	30
FACTORS AFFECTING GAP ACCEPTANCE DECISION	34
<i>Binary Logistic Regression Analysis Results</i>	34
<i>Decision Tree Modeling Results</i>	40
<i>Random Forest Modeling Results</i>	49
<i>Comparison of Model Performance</i>	51
DRIVER BEHAVIOR AFTER ACCEPTING GAPS	51
<i>Vehicle Kinematical Characteristics during Turning</i>	51
<i>Factors Affecting Vehicle Kinematics After Accepting Gaps</i>	60
CHAPTER 4. SUMMARY AND CONCLUSION.....	65
REFERENCES.....	69

LIST OF FIGURES

Figure 1. Graphic. Data categories collected in the SHRP 2 project.⁽⁵²⁾	11
Figure 2. Diagram. Components of the SHRP 2 DAS.⁽⁵³⁾	12
Figure 3. Maps. The selected intersections in Raleigh, North Carolina (left), and Seattle, Washington (right).	13
Figure 4. Graph. Vehicle kinematics profile for a sample left-turn trip.	15
Figure 5. Video stills. Example front- and rear-view video images.	15
Figure 6. Flow chart. Decision tree scheme.	24
Figure 7. Flow chart. Conceptual design of the random forest approach.	25
Figure 8. Graph. Probability and cumulative distributions of accepted and rejected gaps.	29
Figure 9. Graph. Probability and cumulative distributions of accepted and rejected gaps between 2.0–12.0 s.	30
Figure 10. Graph. Critical gap sizes by scenario (maximum likelihood method).	31
Figure 11. Graph. Critical gap sizes by scenario (binary logistic regression).	32
Figure 12. Graph. Critical gap sizes by scenario (probability equilibrium method).	32
Figure 13. Chart. Comparison of critical gap sizes determined by different methods.	34
Figure 14. Scatter plot. Accepted gaps by waiting time.	38
Figure 15. Scatter plot. Rejected gaps by waiting time.	39
Figure 16. Flow chart. Decision tree for all 785 gaps.	44
Figure 17. Flow chart. Decision tree for 423 gaps for right-turn trips.	45
Figure 18. Flow chart. Decision tree for 362 gaps for left-turn trips.	45
Figure 19. Flow chart. Decision tree for 243 gaps for left-turn trips with TWLTL.	46
Figure 20. Flow chart. Decision tree for 119 gaps for left-turn trips without TWLTL.	46
Figure 21. Flow chart. Decision tree for 87 gaps for right-turn trips with leading vehicle.	47
Figure 22. Flow chart. Decision tree for 336 gaps for right-turn trips without leading vehicle.	47
Figure 23. Flow chart. Decision tree for 75 gaps for left-turn trips with leading vehicle.	48
Figure 24. Flow chart. Decision tree for 287 gaps for left-turn trips without leading vehicle.	48
Figure 25. Graphs. Comparison of acceleration profiles between left- and right-turning trips.	52
Figure 26. Graphs. Gaussian process regression profiles for all left-turn trips.	53
Figure 27. Graphs. Gaussian process regression profiles for all right-turning trips.	54
Figure 28. Graphs. Acceleration profiles for left-turning trips (with versus without TWLTL).	54
Figure 29. Graphs. Gaussian process regression profiles for left turn with TWLTL.	55
Figure 30. Graphs. Gaussian process regression profiles for left turn without TWLTL.	55
Figure 31. Graphs. Acceleration profiles for left-turning trips (with versus without leading vehicles).	55

Figure 32. Graphs. Gaussian process regression profiles for left turn with leading vehicles..... 56

Figure 33. Graphs. Gaussian process regression profiles for left turn without leading vehicles..... 56

Figure 34. Graphs. Acceleration profiles for left-turning trips (by age group)..... 57

Figure 35. Graphs. Gaussian process regression profiles for left-turn trips for young drivers. 57

Figure 36. Graphs. Gaussian process regression profiles for left-turn trips for mid-age drivers..... 57

Figure 37. Graphs. Gaussian process regression profiles for left-turn trips for older drivers..... 58

Figure 38. Graphs. Acceleration profiles for right turn (with versus without leading vehicles)..... 58

Figure 39. Graphs. Gaussian process regression profiles for right-turn trips with leading vehicle.. 59

Figure 40. Graphs. Gaussian process regression profiles for right-turn trips without leading vehicle. 59

Figure 41. Graphs. Acceleration profiles for right turn (with versus without leading vehicles)..... 59

Figure 42. Graphs. Gaussian process regression profiles for right-turn trips for younger drivers.. 60

Figure 43. Graphs. Gaussian process regression profiles for right-turn trips for mid-age drivers. . 60

Figure 44. Graphs. Gaussian process regression profiles for right-turn trips for older drivers..... 60

LIST OF TABLES

Table 1. <i>HCM</i> critical gaps (s) for two-way-stop-control intersections. ⁽³⁾	7
Table 2. AASHTO Green Book critical gaps (s) for unsignalized intersections. ⁽⁴⁹⁾	7
Table 3. Determined critical gaps in previous studies.	7
Table 4. Description of final independent variables used in analysis.....	18
Table 5. Summary of the collected data by key variables.	19
Table 6. Basic statistics measures of collected gaps.....	29
Table 7. Basic statistics of the collected gaps/lags between 2 and 12 seconds.	30
Table 8. Critical gap sizes by maximum likelihood method.....	31
Table 9. Critical gap sizes by binary logistic regression.....	31
Table 10. Critical gap sizes by probability equilibrium method.	32
Table 11. Comparison of critical gaps determined by different methods.....	33
Table 12. Overall model performance summary.....	35
Table 13. Model fit statistics for all logistic regression models.	35
Table 14. Significant variables and ORs – all movements.....	36
Table 15. Significant variables and ORs – right turn.....	36
Table 16. Significant variables and ORs – left turn.....	36
Table 17. Significant variables and ORs – left turn with TWLTL.	36
Table 18. Significant variables and ORs – left turn without TWLTL.	37
Table 19. Significant variables and ORs – right turn with leading vehicle.....	37
Table 20. Significant variables and ORs – right turn without leading vehicle.	37
Table 21. Significant variables and ORs – left turn with leading vehicle.	37
Table 22. Significant variables and ORs – left turn without leading vehicle.	37
Table 23. Model information and important features of decision tree models.....	41
Table 24. Model information and important features of random forest models.	49
Table 25. Basic statistics measures of gaussian process modeling profiles.	52
Table 26. Summary of correlation directions for factors affecting turning behavior.	62
Table 27. Linear regression model for average speed.	63
Table 28. Linear regression model for maximum longitudinal acceleration.	63
Table 29. Linear regression model for standard error of longitudinal acceleration.	63
Table 30. Linear regression model for maximum lateral acceleration.	64
Table 31. Linear regression model for standard error of lateral acceleration.	64

LIST OF ABBREVIATIONS AND SYMBOLS

DOT	department of transportation
SHRP 2	Second Strategic Highway Research Program
NDS	naturalistic driving study
RID	Roadway Information Database
TWLTL	two-way left-turn lane
HCM	Highway Capacity Manual
AASHTO	American Association of State Highway Transportation Officials
DAS	data acquisition system
AADT	annual average daily traffic
MLM	maximum likelihood method
BLR	binary logistic regression
AIC	Akaike information criterion
SC	Schwarz criterion

CHAPTER 1. INTRODUCTION

In 2017, there were 8,438 fatal crashes that occurred at intersections, about 45% of which occurred during nighttime, dusk, and dawn.⁽¹⁾ While intersections represent a small section of the total roadway network, these crashes at intersections account for about a quarter of all fatal crashes. Across the nation, unsignalized intersections comprise 90% of all intersections and are involved in about 70% of fatalities occurring at intersections.^(1,2) Recognizing the safety risks at such locations, unsignalized intersection safety has been a focal topic for decades among the traveling public, traffic engineers, and the research community. Currently, many state departments of transportation (DOTs) identify intersections as a safety emphasis area in their State Highway Safety Plans and Strategic Highway Safety Plans.

Even though drivers disregarding stop signs have been widely considered a major contributing factor for crashes at unsignalized rural intersections, an equally important severe crash-related problem at these locations is misjudgment of gaps. To improve safety at unsignalized intersections, a thorough investigation of the factors causing drivers to misjudge gaps remains critical to introducing effective countermeasures. Such investigations should look at different scenarios, accounting for factors such as roadway configuration (e.g., pavement markings, sight triangles, and roadway alignment), lighting condition (e.g., daytime, nighttime with light, and nighttime without light), traffic condition (e.g., traffic pattern and availability of sufficient gaps), and driver characteristics (e.g., age and driving history).

Intersections are risky to navigate by nature due to the conflicting entering/exiting traffic. At a conventional four-leg intersection of two two-lane, two-way roadways, for example, there can be 32 conflict points between different traffic movements, including 8 merging points, 8 diverging points, and 16 crossing points. Traffic signals are one approach used to regulate traffic and mitigate potential crash risks at intersections. However, partly due to their associated costs, traffic signals are typically used only when a number of warranting conditions are met. In many cases, intersections on roadways with relatively low traffic volumes remain unsignalized.

Many such unsignalized intersections have relatively high speed limits, where gap misjudgment could result in severe crashes. Factors such as safe gap availability due to traffic patterns, visibility, available sight distance, time of day, and driver characteristics can all affect the identification and selection of a perceived safe intersection gap. Traditionally, gap acceptance has been mainly studied using observational data based on videotaping or manual timing at intersections. Using such observational data, existing research has tended to study driver gap judgements from an external point of view, with limited consideration of driver factors (e.g., driving history and driver age). Naturalistic driving data, on the other hand, comprises detailed driver information that may provide a better understanding of driver factors in an unobtrusive way. The Second Strategic Highway Research Program (SHRP 2) Naturalistic Driving Study (NDS) data contains synchronized forward-view, rear-view, and face-view video data for thousands of drivers in a naturalistic driving environment. The associated Roadway Information Database (RID) contains detailed roadway and traffic information for the NDS sites. The derived eye-glance data further allow researchers to understand drivers' visual behavior while undertaking different driving maneuvers.

This report describes an attempt to fully understand gap acceptance behavior at unsignalized intersections using the SHPR 2 NDS data.

LITERATURE REVIEW

Gap acceptance at unsignalized intersections has long been an important research topic. A gap is the temporal interval between two consecutive vehicles traveling through a subject intersection. A lag is the remaining proportion of a gap that is available for a vehicle approaching the intersection to perform the desired maneuver. For right-turning vehicles, the gap is defined only based on vehicles in the near lane, while for left-turn and through maneuvers, a vehicle in any lane could define the start or end of a gap.⁽³⁾

When drivers traveling on a minor road arrive at the stop line of an unsignalized intersection, they need to decide when to enter the intersection based on the right of way hierarchy and the availability and distribution of gaps. This decision-making process involves the identification and acceptance of safe gaps. Intersection gap acceptance behaviors are affected by a number of factors, such as intersection sight distance, queue length, delay, and capacity at unsignalized intersections.^(4,5) To date, a fair amount of research has been conducted to understand driver gap acceptance behaviors.

Factors Affecting Gap Acceptance Behavior

As noted, a variety of factors can affect driver gap acceptance behaviors at intersections. These are further enumerated below.

- Roadway and traffic-related factors:
 - **Median and storage lane.** The presence of a median or a storage lane in the major road provides space to accommodate left-turn vehicles, converting the one-stop left-turn maneuver into two separate stop-controlled maneuvers. Such a lane is hence associated with increased roadway safety, capacity and smaller critical gap value.^(3,6) In comparison with undivided major roads, a continuous two-way left-turn lane (TWLTL) could result in up to 35% reduction in total crashes, 30% reduction in delay, and 30% increase in capacity.⁽⁷⁾
 - **Sign control.** Some studies found that gap acceptance behavior was affected by the type of traffic control at the unsignalized intersection (e.g., yield versus stop, stop versus stop with flashing beacons).^(8, 9)
 - **Sight distance.** Studies found that sight obstructions at intersections could increase critical gap values.⁽¹⁰⁾
 - **Intersection geometry.** Studies identified significant differences in gap acceptance behavior among intersections with different geometric layouts.⁽¹¹⁾ Al-Mhairat investigated three roundabouts and three T-intersections and found statistically significant differences in determined critical gaps for these two types of intersections.⁽¹²⁾
 - **Major road traffic volume.** Studies found that minor-road drivers tended to accept shorter gaps at intersections where heavy traffic existed on the intersecting major road or during peak hours.⁽¹³⁾

- **Major road traffic speed.** Studies found conflicting results on the influence of major-road traffic speed on gap acceptance. Some studies suggested a positive correlation between the increased major stream speeds and reduced accepted gap sizes while others revealed a negative correlation.⁽¹⁰⁾ One study found that major stream speed was not an important factor based on statistical analysis but was identified as the most influential factor affecting gap acceptance decisions based on driver reports.⁽¹⁴⁾ Another study further added that spatial gap sizes played a more critical role than temporal gap sizes, suggesting that drivers were better at perceiving the spatial gaps than the temporal gaps in gap acceptance behavior.⁽¹⁵⁾
- **Driver-related factors:**
 - **Age.** Many studies identified younger drivers as tending to be more aggressive and accepting of shorter gaps at unsignalized intersections.^(10,12,13,16-19) Researchers considered the longer critical gaps for elder drivers to be a result of diminished visual capabilities, such as depth and motion perception and slow and less accurate judgement.
 - **Gender.** Some research identified significantly shorter accepted gaps for male drivers compared to female drivers.^(6,10,20-22) Others showed no significant differences between these two groups.^(23,24)
 - **Emotional state.** Drivers' emotional state could also influence gap acceptance behavior. One study, for example, found a positive correlation between increased stress levels and increased probability of accepting shorter gaps.⁽²⁵⁾
 - **Driving Experience.** A study conducted on a driving simulator found that novice drivers were more likely to accept a gap falling within the decision "dilemma zone" and were more likely to perform inconsistently.⁽²⁶⁾
 - **Waiting time.** Research indicated that drivers tended to accept shorter gaps as the waiting time increased,^(6,33,27-29) which could be due to reduced driver patience or better estimation of the upcoming gaps over the increased waiting time.
- **Environmental factors:**
 - **Time of day.** Studies found that gaps during off-peak periods were larger and more stable, making it easier for drivers to merge or cross, resulting in longer accepted gaps during off-peak periods than during peak periods.^(6,12,16) In addition, drivers' gap acceptance behavior varied between daytime and nighttime. This difference was found to be significant for elder drivers.⁽¹⁸⁾
 - **Rain intensity.** Some studies revealed that the accepted gap sizes increased with increased rain intensity.⁽²⁹⁾
- **Other factors:**
 - **Presence of passengers.** Some studies found that drivers accepted significantly shorter gaps if there were passengers in their vehicles,^(13,16) while others showed that drivers tended to drive more safely and accepted longer gaps when there were passengers.⁽²²⁾
 - **Presence of following vehicles.** Drivers waiting for left turns with vehicles queuing behind them tended to accept significantly shorter gaps than those without, while the difference was not significant for right-turn drivers.^(13,22)

- **Approaching vehicle type/size.** Studies found that drivers were more likely to accept shorter gaps in front of small vehicles, such as motorcycles and passenger cars, compared to trucks and buses.⁽³⁰⁾
- **Subject vehicle type/size.** For minor-road vehicles, field studies demonstrated that heavy vehicles required longer gaps than smaller vehicles to enter a major road.^(4,24,31,32) In general, a single-unit truck appeared to require critical gaps of 1–2 seconds longer than a passenger car, and a combination truck appeared to require critical gaps approximately 1–2 seconds longer than a single-unit truck.⁽¹⁵⁾ In addition, trucks, SUVs and commercial vehicles tended to be more aggressive when accepting gaps than passenger cars and vans.⁽¹⁶⁾

Critical Gaps and Critical Gap Determination

As discussed previously, gap acceptance decisions are affected by a wide variety of factors. To fully account for such factors, gap acceptance responses can be analyzed as a statistical model of all significant factors using different methods.^(8,33,34) As accepted gaps and rejected gaps are naturally contrary, some studies also used machine learning methods, such as support vector machines, random forests, decision trees, and artificial neural networks, for classifying and predicting gap acceptance decisions.⁽³⁵⁻³⁷⁾ Such approaches are particularly suitable for predicting driver behavior given specific roadway, traffic, environmental, and driver characteristics. They can be, however, difficult to implement in practice during intersection design and operations analysis due to their data-demanding and microscopic nature.

Critical gaps are an important concept in understanding gap acceptance behaviors at unsignalized intersections and for conducting intersection-related operations and safety analyses. Critical gaps refer to the gap size thresholds that are larger than those individual groups of drivers would likely accept under given roadway/traffic settings. Well-accepted critical gap values for different scenarios are required for a wide range of engineering applications relevant to intersections, such as capacity analyses, roadway geometric design, and traffic control analyses.

The determination of critical gaps has been an international research focus among individuals with an interest in traffic operations and safety. Using observational data, a critical gap is a value between the shortest gap subject drivers are observed accepting and the longest gap they reject. In an ideal setting, a large number of random gaps would be available on the major street while a continuous queue of vehicles from the minor street could be studied for their gap acceptance decisions. Under this scenario, the critical gap may be considered as the gap during which exactly one vehicle entered the major street stream based on a regression curve between the gap sizes and the number of consecutive vehicles that entered each gap.^(5,38)

In reality, such traffic settings are rarely available during observational studies. Researchers frequently need to deal with data anomalies collected in the field, such as lags, very short gaps, and very long gaps. Over the years, the research community has used a large number of methods for the accurate determination of critical gaps. These methods each have different pros and cons and vary in assumptions, level of sophistication, and ease of use. The following briefly summarizes a sample of critical gap determination methods.^(5,39,40)

- **Gap acceptance curve**, which plots the percentage of accepted gaps against gap values, and the critical gap is the gap accepted by a predetermined percentage of drivers (e.g., 50% or 85%).⁽⁴¹⁾
- **Greenshield method**. This method is based on the histogram of the frequencies of rejected and accepted gaps. A critical gap is determined as the gap corresponding to an equal number of drivers that accepted and rejected it.⁽³¹⁾ This method is straightforward to use but is prone to impacts of sample sizes or gap distributions.
- **Ashworth method**. The Ashworth method. This method fits the critical gaps and the accepted gap values into normal distributions and determines the average critical gap as a function of the mean and variance of the accepted gaps.⁽⁴²⁾ This method is considered easy to use and relatively accurate during suitable traffic conditions, but is dependent on the conflicting traffic volume and could be inaccurate for certain traffic and headway patterns.
- **Raff method**. This method uses the intersection of the cumulative distribution of accepted gaps and the inverse curve of the cumulative distribution of rejected gaps as the critical gap. The method makes no assumption of data distribution and is straightforward to use. However, it is susceptible to the impacts of gap data composition and a large proportion of long gaps in the data would result in increased critical gap value.⁽²⁷⁾ The results of the Raff method may be improved by the exclusion of long accepted gaps.^(23,43)
- **Logit method**. Also referred to as the logistic regression method, this method models the gap acceptance and rejection decisions as a binomial response of gap size. The critical gap is determined as the gap size corresponding to a certain level of acceptance probability (e.g., 50% or 85%). This method tends to produce lower critical gap values, particularly when both gap and lag data is used.
- **Maximum likelihood method**. This method assumes that the critical gaps for a given driver population follow a lognormal distribution and the mean critical gap is selected as the general critical gap for the population.⁽⁴⁴⁾ The critical gap is determined with distribution coefficients that maximize the likelihood that the critical gap for an individual is higher than the highest rejected gap and lower than the actual accepted gap. A later study⁽⁴⁵⁾ modified the maximum likelihood method to analyze data with only accepted gaps or rejected gaps. The maximum likelihood method can be less reliable when analyzing data from an inhomogeneous population.
- **Probability equilibrium**. This method was developed to estimate the critical gaps based on the cumulative distributions of accepted and rejected gaps.⁽⁴⁶⁾ The method does not make pre-assumptions about data distribution and requires less computation than the maximum likelihood method. With simpler calculations, this method also is able to produce results similar to the maximum likelihood method in many cases.
- **Hewitt's method**. This method assumes that the accepted and rejected gaps follow the normal distribution, and estimates the average value and variance of critical gaps using

the cumulative distribution function of major stream gaps and an iterative process.⁽²⁸⁾ However, the method could yield unreasonably low critical gap values, particularly when applied on nonhomogeneous traffic conditions.

In addition to the estimation methods, the estimated critical gap values are also influenced by the selection and preprocessing of the collected data. The common considerations when researchers are processing the collected gap data include the following.

- **Inclusion or exclusion of large gaps.** Many studies indicated that gaps larger than 12 seconds were generally accepted by all minor-road drivers. These gaps could not provide meaningful information about gap acceptance behavior but bias the estimation results, and thus should be excluded from the analysis.^(3,11,27,47)
- **Use maximum rejected gaps or all rejected gaps.** Tupper compared three methods with the variation of using the maximum rejected gaps and all rejected gaps. It was found that the use of only maximum rejected gaps gave lower critical gaps for some methods, such as Raff's method,⁽²⁷⁾ but resulted in higher critical gap value for other methods, such as the probability equilibrium method.^(27,46)
- **Use of gaps and lags.** Some previous studies used only lags to determine critical gaps. Studies⁽⁴⁸⁾ suggested that estimations with only lag information could be less accurate than those utilizing all gaps, possibly due to the limited information that lags represent and potential accuracy issues associated with measuring lags. Some studies, on the other hand, showed that there were no statistically significant differences between using lag and gap data.⁽³¹⁾

Due to methodological and data factors involved, different methods frequently perform differently for different datasets.^(16,20,48) For example, a previous National Cooperative Highway Research Program study⁽²⁴⁾ used both the logit method and Raff method. With the gap acceptance decisions for 6,243 right-turn movements and 3,526 left-turn movements observed, the logit method gave a critical gap (at a 50% probability of acceptance) of 6.5 seconds for right-turn passenger cars, and 8.2 seconds for left-turn passenger cars, while the results of the Raff method were about 0.2 seconds shorter. Brilon et al.⁽³⁹⁾ provided a comprehensive review of various estimation techniques of critical gaps. With a series of simulations performed, the author concluded that the maximum likelihood method and Hewitt's method gave the most truthful estimations.

Critical Gap Values

The *Highway Capacity Manual (HCM)*⁽³⁾ and American Association of State Highway Transportation Officials (AASHTO) Green Book⁽⁴⁹⁾ include critical gaps for different movements at unsignalized intersections (Table 1 and Table 2) for roadway design and capacity analysis purposes. Note that the critical gaps specified in the Green Book are more conservative than those in the *HCM*.

Table 1. HCM critical gaps (s) for two-way-stop-control intersections.⁽³⁾

Vehicle Movement	Base Critical Gap 2-Lane Major Street	Base Critical Gap 4-Lane Major Street
Left turn from major	4.1	4.1
Right turn from minor	6.2	6.9
Through move from minor	6.5	6.5
Left turn from minor	7.1	7.5

Table 2. AASHTO Green Book critical gaps (s) for unsignalized intersections.⁽⁴⁹⁾

Vehicle Movement	Gap for 2-Lane Highway Major Street	For Each Additional Lane
Right turn from minor		
Passenger car	6.5	+0
Single-unit truck	8.5	
Tractor-trailer	10.5	
Left turn from minor		
Passenger car	7.5	+0.5
Single-unit truck	9.5	+0.7
Tractor-trailer	11.5	+0.7

In addition to the values outlined in the manuals, Table 3 summarizes the critical gap values determined by previous studies.

Table 3. Determined critical gaps in previous studies.

Study Site		Critical Gap	
		Type	Value (s)
Illinois, Missouri, and Pennsylvania in U.S. ⁽⁴⁾	13 stop-controlled intersections, 5 three-leg intersections, and 8 four-leg intersections	Passenger car, right turn	6.3
		Passenger car, left turn	8
		Single-unit truck, right turn	8.4
		Single-unit truck, left turn	9.8
		Combination truck, right turn	10.7
		Combination truck, left turn	10
Amman, Jordan ⁽¹²⁾	3 loop intersections and 3 three-leg intersections	Peak hours	3.1–3.3
		Non-peak hours	5.0–5.5
Multi-state, U.S. ⁽¹³⁾	8 stop-controlled T-intersections.	Right turn	5
		Left turn	4.25
Kerala, India ⁽¹⁷⁾	2 T-intersections	Right turn	2.8–5.1
San Francisco, U.S. ⁽²¹⁾	5 intersections	N/A	5.6–7.6
Ahmedabad, India ⁽²⁴⁾	A four-leg uncontrolled intersection at suburban area	2 wheeler	2.3–4.8
		3 wheeler	2.65–5.0
		Car	2.25–4.65
Massachusetts and Oregon in U.S. ⁽²⁷⁾	60 sites	Right turn	4.5–24.7
		Left turn	3.5–14.7
Christiansburg, U.S. ⁽²⁹⁾	A four-leg signalized intersection	Left turn, first lane	6.0–7.2
		Left turn, second lane	8.1–9.2

Study Site		Critical Gap	
		Type	Value (s)
Pennsylvania, U.S. ⁽³¹⁾	6 intersections with different traffic patterns, each has two 2-lane roads	Passenger car, left turn	5.00–8.08
		Passenger car, right turn	4.50–7.49
		Passenger car, through	6.00–7.80
		5-axle truck, left turn	7.25–8.27
		5-axle truck, right turn	8.25–12.50
		Less-than-5-axle truck, left turn	10.25–11.16
		Less-than-5-axle truck, right turn	6.25–13.17
Warangal, India ⁽³²⁾	2 All-way-stop-controlled (AWSC) intersections	2 wheeler	2.60–4.84
		3 wheeler	2.27–7.08
		4 wheeler	2.8–8.59
		LCV	3.4–9.9
		Tractor	3.85–11.48
		Bus	4.2–8.6
Dalian, China ⁽⁴³⁾	A four-leg two-lane roundabout	N/A	2.65–3.2
Johor, Malaysia ⁽⁵⁰⁾	2 T-intersections	Right turn	10.0–15.0
		Left turn	6.0–10.0
India ⁽⁵¹⁾	3 uncontrolled T-intersections	N/A	2.90–4.55

Literature Review Discussion

Intersection gap acceptance behavior has been a research focus for decades for purposes relevant to intersection safety, operations, and design. To date, there have been a wide range of methods used for gap acceptance analysis and the determination of critical gaps. The literature revealed that critical gaps determined by previous studies varied significantly based on a number of factors, such as location, method used, and intersection type. Even at the same intersections, critical gap values varied with factors such as vehicle type, vehicle movement, and time of day. In general, the literature review showed that critical gaps were shorter in developing countries, at intersections with mixed traffic patterns and/or higher traffic volumes, for smaller subject vehicles, and during peak hours.

Traditionally, gap acceptance has been mainly studied using observational data based on video or manual timing at intersections. By using observational data, existing research studied driver gap judgements from an external point of view, without taking into account the role of the driver's behavior in the process of gap judgement. In addition, most previous studies were based on a limited number of intersections. Depending on local traffic and driver characteristics, studies frequently found significantly different critical gap values at different intersections (see Table 3). The results clearly suggested that simple average critical gaps among the limited intersections or overall gaps determined with combined data from individual intersections would not realistically produce a well-received critical gap value suitable for different locations.

The availability of the SHRP 2 NDS data has provided a unique opportunity to study driver gap acceptance behavior from a different angle. Naturalistic driving data can help provide a better understand this problem by examining driver behavior in an unobtrusive way. The SHRP 2 NDS data contain synchronized forward-view, rear-view, and face-view video data for thousands of drivers at a massive number of intersections in a naturalistic driving environment. The associated RID contains detailed roadway and traffic information for the NDS sites.

PROJECT OBJECTIVES

This project used SHRP 2 NDS data to gain a comprehensive understanding of driver gap acceptance behaviors at unsignalized intersections. In particular, this project included the following objectives:

1. Identify critical gaps for common traffic/roadway scenarios at unsignalized intersections.
2. Understand significant factors affecting driver gap acceptance behaviors at unsignalized intersections.
3. Investigate driving turning behaviors at unsignalized intersections after a gap acceptance.

CHAPTER 2. METHODOLOGY AND DATA COLLECTION

DATA COLLECTION

SHRP 2 Naturalistic Driving Data Overview

SHRP 2 launched the NDS project between 2010 and 2013 to investigate driver behavior and performance relevant to safety under real-world scenarios.⁽⁵²⁾ With more than 3,500 participating drivers recruited from all age groups, the study collected a massive amount of naturalistic driving and related data (Figure 1) from six sites across the country: Bloomington, Indiana; central Pennsylvania; Tampa Bay, Florida; Buffalo, New York; Durham, North Carolina; and Seattle, Washington.

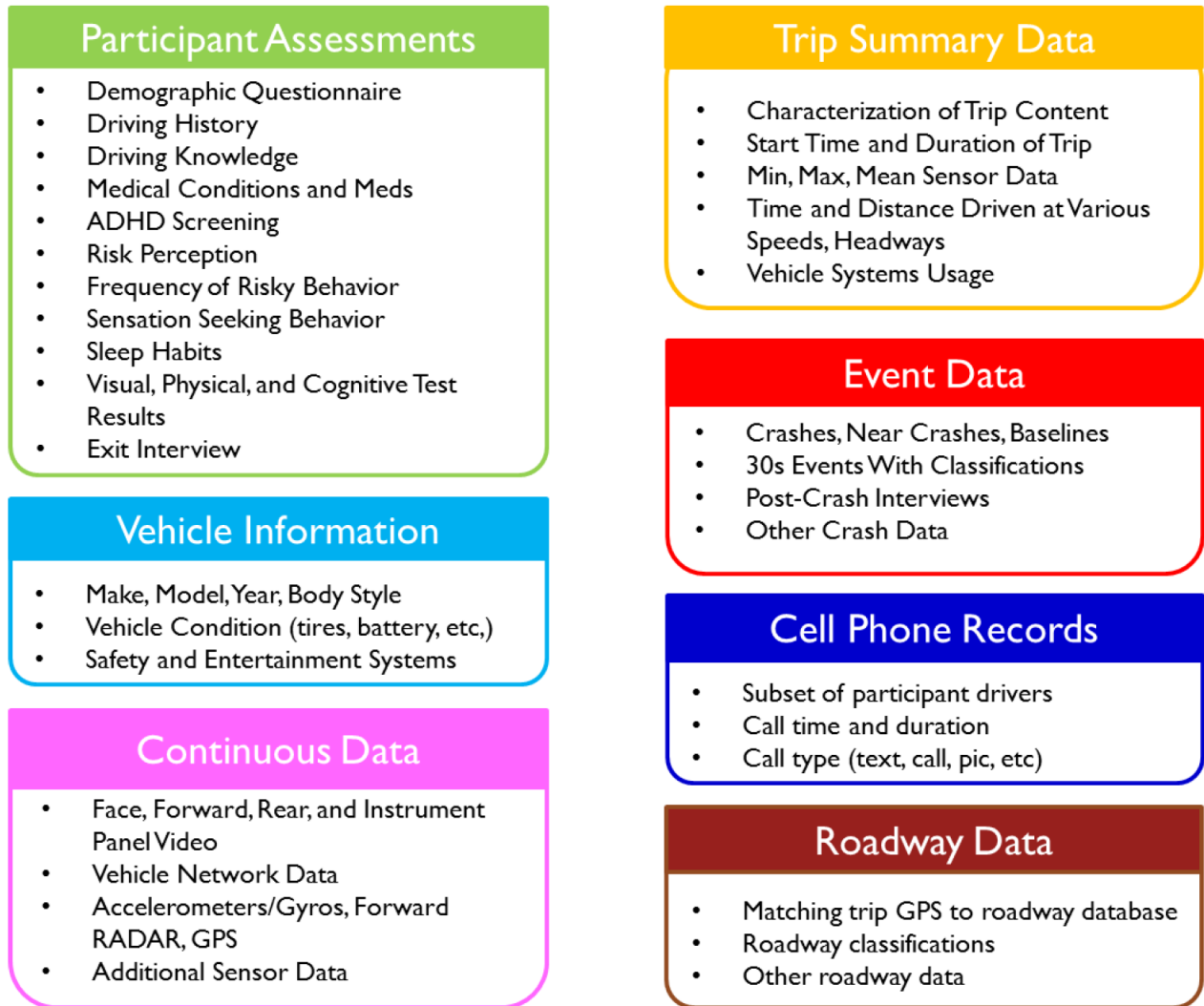


Figure 1. Graphic. Data categories collected in the SHRP 2 project.⁽⁵²⁾

The SHRP 2 NDS used an onboard data acquisition system (DAS) for vehicle kinematic and driver behavior data collection. The DAS consisted of a forward radar, four video cameras (observing driver's face and hand, passenger side, forward and rear roadway), accelerometers,

GPS receivers, computer vision lane-tracking capability, and data storage equipment (Figure 2).⁽⁵³⁾

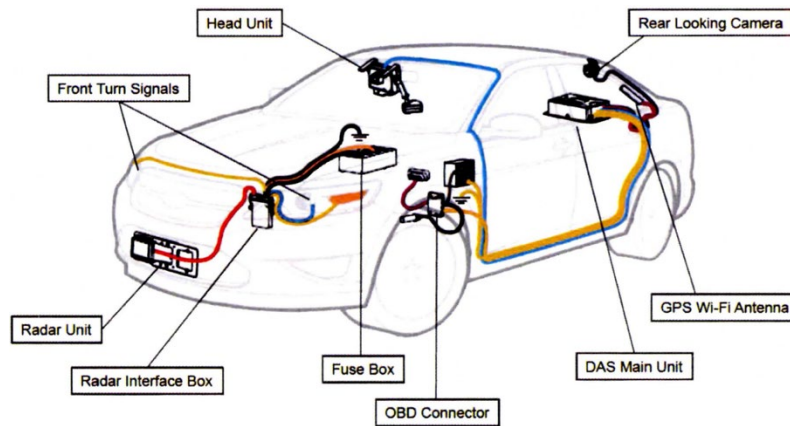


Figure 2. Diagram. Components of the SHRP 2 DAS.⁽⁵³⁾

The final SHRP 2 NDS database contained information about more than 5 million trips and 41,000 events, with additional events and data added as data processing continued.⁽⁵⁴⁾ SHRP 2 sampled three types of events: crashes, near-crashes, and baseline events. Crashes and near-crashes in the NDS database are also collectively referred to as safety-critical events. In addition, the SHRP 2 project also developed the RID with relatively detailed traffic and roadway information for the six participating sites. The database incorporated both data originated at state DOTs and the data collected by instrumented vehicles. The linkage between the SHRP 2 driving and roadway data provided researchers the opportunity to effectively identify driving data on particular roadway segments of interest.⁽⁵⁵⁾

SHRP 2 Data Request and Processing

For this study, the research team selected unsignalized intersections at two NDS sites: Seattle, Washington, and Raleigh, North Carolina, and then requested NDS data at the selected intersections from the SHRP 2 data team. The researchers initially selected 113 intersections based on the RID and using satellite images at Esri® ArcGIS™ and Google® Maps™. All intersections met the following criteria:

- **Unsignalized T-intersections with the minor road being the third leg.** The use of T-intersections reduced the number of conflicting traffic streams and therefore increased the number of useful gaps, thus simplifying the data collection effort.
- **Major roads at the intersections are not divided.** The existence of wide medians would provide space that would convert typical left-turning movements into two-phase maneuvers. Wide medians also significantly increase the area of intersections and therefore time required to traverse them, which would inevitably affect gap acceptance behavior.
- **Intersections are not closely located from other nearby intersections.** Closely located intersections would affect the traffic/headway characteristics at the study intersections.

For this project, the team only selected intersections from which the nearest intersections, including driveways, were at least 500 feet away.

- **Major roads at intersections had an annual average daily traffic (AADT) rate greater than 10,000.** Such intersections would allow the research team to collect a sufficient number of both rejected and accepted gaps.
- **Intersections are not in interchange areas.** Such intersections would not involve traffic coming off freeways, which might exhibit different driving behavior.

The research team initially requested 1,000 NDS trip segments entering the 113 intersections from the minor approaches from the SHRP 2 team. Unfortunately, a number of intersections did not have trips meeting the previously discussed criteria. The 1,000 trips were all extracted from 64 of the originally selected intersections (Figure 3). All subject vehicles were passenger vehicles, as the SHRP 2 NDS database only included passenger vehicle participants.

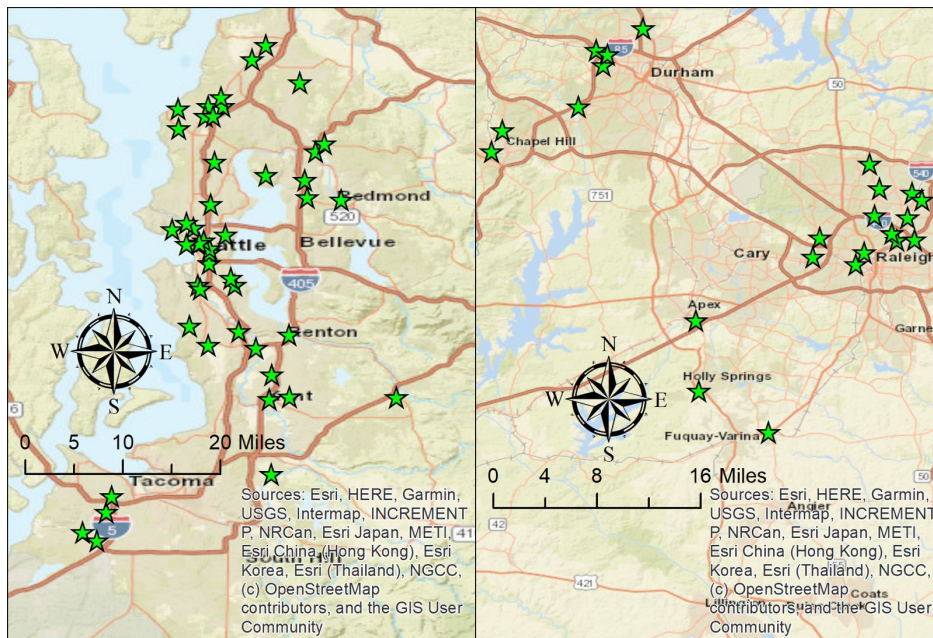


Figure 3. Maps. The selected intersections in Raleigh, North Carolina (left), and Seattle, Washington (right).

For each requested trip, the research team requested the following SHRP 2 data for further analysis:

- Time series data, which is detailed vehicle kinematics data collected at a frequency of 10 Hz coupled with GPS speed and location data collected at a 1-Hz frequency.
- Forward-facing, rear-view, and face-view videos providing continuous video recordings of the outside and inside of the subject vehicle during each trip.
- Driver demographic and driving history data, which included information about driver age, driving age, and past violation/crash history for the subject drivers of the trips selected.

Data Collection and Preprocessing

Video Data Collection

The research team collected and processed the video-related data in a Virginia Tech Transportation Institute data reduction lab. The researchers carefully reviewed each obtained trip video and extracted the following time-related variables to describe the drivers' gap acceptance behavior:

- **Queue start time.** The timestamp at which the subject vehicle started queuing in line and waiting to turn on the minor road (i.e., after arriving at a full stop on the minor approach). If there was no vehicle presented in front of the subject vehicle when approaching the intersection, this variable was coded as NA.
- **Intersection arrival time.** The timestamp at which the subject driver arrived at the intersection, which was referred to as the time when the subject vehicle stopped at the stop sign or stop line as observed from the forward-facing videos. In the cases that the subject driver did not stop completely at the stop sign/line, the intersection arrival time was determined as the time when the subject vehicle's front bumper began to cross the stop line/sign and/or visually began observing the major-road traffic conditions as shown from the face-view videos.
- **Maneuver start time.** The timestamp at which the driver finished the gap observation and started the turning maneuver. A variety of indicators were used to determine this timestamp accurately, including the time when the subject vehicle's speed increased from 0 mph as seen in speed profile, the beginning time of a positive peak in the longitudinal acceleration profile, or the beginning of a positive (for right-turn trips) or a negative (for left-turn trips) peak in the lateral acceleration profile (Figure 4). In addition, the subject drivers' facial expression or hand movements were also considered when determining this information.
- **Major road vehicle arrival time.** The timestamps at which the major road vehicles arrived the intersection. This variable was used to calculate the gap/lag size. The researcher recorded the vehicle arrival time according to the forward-facing and rear-view videos. Figure 5 shows examples of the forward-facing and rear-view camera images at a study intersection. In this case, the subject driver rejected three consecutive gaps and then accepted the fourth one. It should be noted that the estimation of the major road vehicle arrival time based on rear-view videos for long accepted gaps (mostly for accepted gaps that were larger than 10 seconds) could be less accurate due to the limited resolution of the rear-view videos.
- **Wait time.** Researchers calculated the wait time for each accepted and rejected gap as the time difference between the *major road car arrival time* (for a rejected gap) or *maneuver start time* (for an accepted gap) and the *intersection arrival time* or *queue start time* (if not NA).

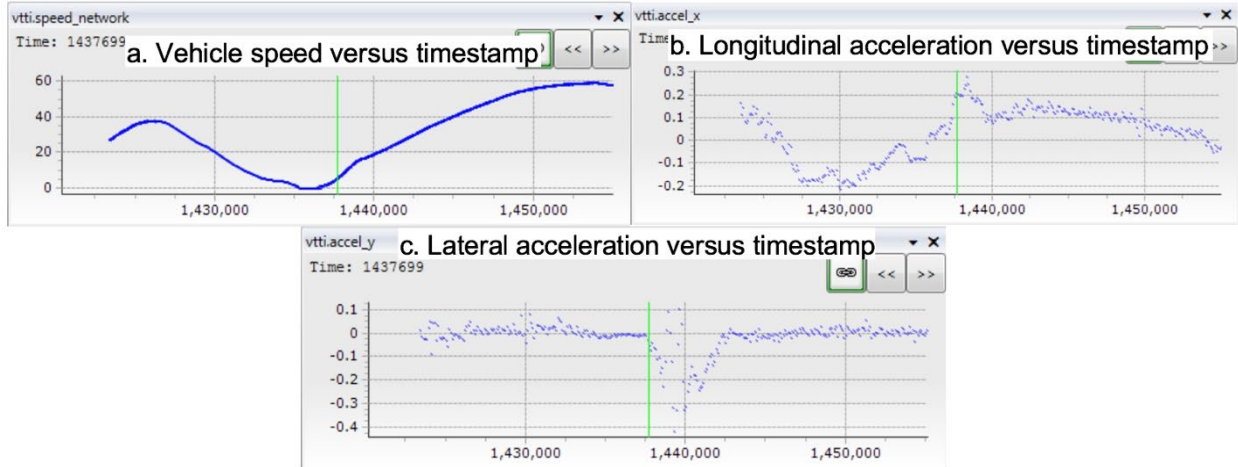


Figure 4. Graph. Vehicle kinematics profile for a sample left-turn trip.

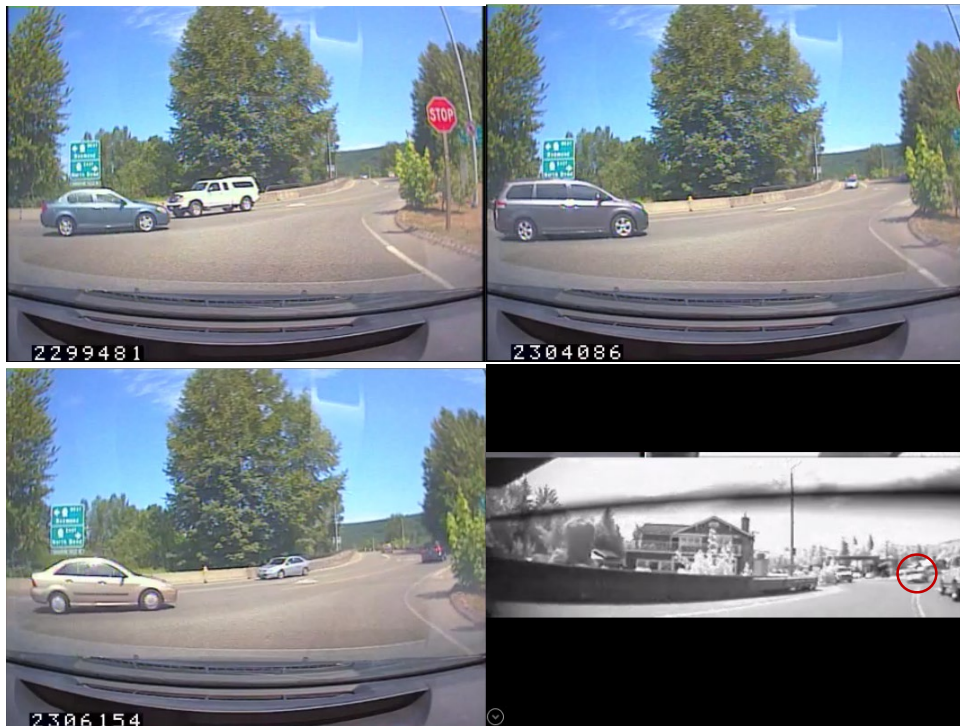


Figure 5. Video stills. Example front- and rear-view video images.

In addition to the above listed time-related variables, the researchers also collected the following trip-related variables from the trip video files: lighting condition, driver distraction, the presence of leading and/or following vehicles, and weather condition. Note that the project team deleted all trips involving severe adverse weather conditions that significantly reduced visibility (e.g., raining, snowing, or heavy fog). The team, however, kept a number of cases with minor weather conditions such as light rain or fog that did not significantly affect the video quality.

Among the original 1,000 trips, the research team was able to collect gap information for 466 trips at 64 intersections, resulting in 270 accepted gaps and 900 rejected gaps. Note that the gap

data included 466 lags, which is determined as the time difference between the subject vehicle's intersection arrival time and the arrival time for the first vehicle on the major road. Many trips were not used during this process due to one or more of the following reasons:

- No gap data collected;
- Trips occurred during severe adverse weather conditions; and
- Trip videos could not be processed due to low quality.

Intersection Data Collection

For each selected intersection, the research team collected a large number of variables depicting the major- and minor-road geometry, traffic control, speed limit, and AADT information from a combination of sources, including the SHRP 2 RID database, Google Maps, and the satellite images in ArcGIS.

Participant Data Collection

The SHRP 2 data team delivered the participant data associated with each trip. Note that all personally identifiable information was viewed and processed in secure VTTI data reduction labs.

Time Series Data

For each trip, the researchers coded a start and an end timestamp corresponding to the start and end of the turning maneuver. The researchers calculated the following variables to describe subject drivers' turning behavior:

- Average speed
- Maximum longitudinal and lateral accelerations
- Standard deviation of longitudinal and lateral accelerations

This data collection effort was for the purpose of understanding drivers' turning behaviors after they accepted a gap. The researchers collected vehicle kinematics data for 433 (out of 466) trips. The delivered time series data for 33 trips were found not to include the turning period and were therefore excluded.

Summary of Independent Variables and Collected Data

This study involved a large number of independent variables collected from different sources. To reduce unnecessary variables and avoid potential issues during the modeling efforts, the project team conducted a preliminary collinearity analysis using the Pearson correlation test. When used together, highly correlated independent variables may result in erratic modeling results, such as increasing the standard errors of the estimated variables or underestimating significant variables. The collinearity analysis suggested that the following variable pairs had a high correlation:

- **Age group and driving age.** The Pearson’s correlation coefficient of these two variables was $R = 0.860$ with $p < 0.0001$, suggesting a significant positive correlation between them. The research team therefore only used age group in this study.
- **Number of crashes within past 3 years and highest crash severity.** The correlation test gave a coefficient of $R = 0.955$ with $p < 0.0001$. The project team only used highest crash severity (of crashes in the past 3 years) in this study.
- **Major road number of through lanes – left approach and major road number of through lanes – right approach.** The Pearson’s correlation test showed $R = 0.974$ with $p < 0.0001$. The project team used the larger number of the two values as “major road number of through lanes.”
- **Major road AADT – right approach and major road AADT – left approach.** The correlation test identified a significant positive correlation between these two variables, with $R = 0.873$ and $p < 0.0001$. The project team used a single variable (major road AADT) with the average of both right and left AADTs to replace the two variables.
- **Major road sidewalk – right approach and major road sidewalk – left approach.** The correlation test gave $R = 0.998$ with $p < 0.0001$, suggesting that these two variable values were highly correlated. The project team therefore used a single variable (major road sidewalk) instead of the two redundant variables. For the new variable, a sidewalk was considered present as long as it was present on either major approach.

Table 4 is a complete list of the final independent variables used in this study.

Table 4. Description of final independent variables used in analysis.

Variable	Variable Type and Values
Gap variable	
Gap indicator	Binary (1 = gap, 0 = lag)
Gap size	Continuous in seconds
Trip variable	
Wait time	Continuous in seconds
Left-turn indicator	Binary (1 = left, 0 = right)
Leading-vehicle presence	Binary (1 = yes, 0 = no)
Distraction involved	Binary (1 = distracted, 0 = not distracted)
Environmental variable	
Lighting condition	Binary (0 = daylight, 1 = other conditions)
Weather condition	Binary (1 = adverse weather, 0 = no adverse weather)
Average speed	Continuous in km/h
Maximum longitudinal acceleration rate during turning	Continuous in g
Longitudinal acceleration standard deviation during turning	Continuous in g
Maximum lateral acceleration rate during turning	Continuous in g
Lateral acceleration standard deviation during turning	Continuous in g
Driver-related variable	
Age group	Categorical (young or 16–24, mid-age or 25–64, older or > 64)
Annual miles traveled	Ordinal (1 ≤ 10,000, 2 = 10,000-15,000, 3 = 15,000-20,000, 4 ≥ 20000).
Number of violations in past 3 years	Binary (0 = 0, 1 ≥ 1)
Max crash severity	Ordinal (0 = not applicable, 1 = property damage, 2 = injury)
Intersection geometry variable	
Major road AADT	Continuous
Major road number of through lanes	Continuous
Major road right-turn-only lane on left approach	Binary (1 = yes, 0 = no)
Major road right-turn-only lane on right approach	Binary (1 = yes, 0 = no)
Major road two-way left turn lane (TWLTL)	Binary (1 = yes, 0 = no)
Major road crosswalk on left approach	Binary (1 = yes, 0 = no)
Major road crosswalk on right approach	Binary (1 = yes, 0 = no)
Minor road crosswalk	Binary (1 = yes, 0 = no)
Major road sidewalk on left approach	Binary (1 = yes, 0 = no)
Major road sidewalk on right approach	Binary (1 = yes, 0 = no)
Major road street parking on left approach	Binary (1 = yes, 0 = no)
Major road street parking on right approach	Binary (1 = yes, 0 = no)
Number of lanes on minor road approach	Continuous
Minor road street parking right	Binary (1 = yes, 0 = no)
Minor road street parking left	Binary (1 = yes, 0 = no)
Intersection alignment	Categorical (right, skew left, skew right)
Presence of street light	Binary (1 = yes, 0 = no)
Major road alignment	Categorical (curve outside, curve inside, tangent)

Table 5 describes the data collected using frequency counts by key analysis variables.

Table 5. Summary of the collected data by key variables.

Count Type	Observation	Count
Drivers by age	< 25	51
	25–64	45
	> 64	54
Gaps by age	< 25	565
	25–64	242
	> 64	363
Gaps by intersection	Max	271
	Mean	18
Gaps by turning type	Left	556
	Right	614

METHODOLOGY

Following the project objectives, the research team used a multi-pronged approach to fully understand driver gap acceptance behaviors at unsignalized intersections. The approach involved the following analyses:

- Developing gap acceptance decision models and determining critical gaps. These two analyses were designed to fully understand factors affecting driver gap acceptance behavior and the critical gaps under different traffic/roadway settings:
 - **Determination of critical gaps.** This analysis used three methods to determine the critical gaps based on the SHRP 2 NDS data, including maximum likelihood (for critical gap determination), binary logistic regression, and probability equilibrium. These three methods each have advantages and disadvantages but are widely used by the research community for critical gap determination.
 - **Modeling gap acceptance decisions.** This analysis modeled drivers' gap acceptance decisions with consideration of all significant independent variables. To develop a comprehensive understanding of the gap acceptance decisions, the project team utilized three methods, including binary logistic regression, decision tree, and random forest.

For both analyses, the project team considered the following scenarios:

- All collected trips
- Right-turn trips
- Left-turn trips
- Left-turn trips with TWLTL on the major road
- Left-turn trips without TWLTL on the major road
- Right-turn trips with leading vehicle
- Right-turn trips without leading vehicle
- Left-turn trips with leading vehicle
- Left-turn trips without leading vehicle

- **Understanding turning behaviors after accepting a gap.** The study of critical gaps and the modeling of gap acceptance decisions were intended to answer the question of when drivers enter the major road traffic stream. It can be equally important to understand *how* drivers enter the major road traffic stream. Major vehicle kinematical metrics such as speed and lateral and longitudinal acceleration rates at strategic points along turning paths after drivers accept a gap provide critical information on how drivers accelerate to merge into the major road traffic stream.

This information can be valuable for microscopic simulation algorithms and tools. The constructed vehicle kinematical profiles can also serve as the typical/baseline turning scenarios for safety analyses of driver behaviors at intersections. In addition, such information may be used for automated vehicle technology development to enable more consistent vehicle behavior at intersections. Vehicle behaviors that are more consistent with human driver expectations will undoubtedly improve safety in a traffic environment with vehicles of mixed levels of automation and nonmotorized vehicle roadway users.

During this analysis, the project team used the Gaussian process regression method to compile the key vehicle kinematic data along the driver's turning path by scenario, including longitudinal and lateral acceleration rates at strategic temporal/spatial points. In addition, the team used multiple linear regression to identify significant correlations between the aforementioned independent variables and drivers' choices on speed selection and acceleration rates when turning after a gap. In addition to those previously listed, this analysis also considered the following scenarios:

- Right-turn trips for young drivers
- Right-turn trips for middle-aged drivers
- Right-turn trips for elderly drivers
- Left-turn trips for young drivers
- Left-turn trips for elderly drivers
- Left-turn trips for elderly drivers

The following sections provide more details on the aforementioned methods.

Critical Gap Determination

This section describes the three methods used for determining the critical gap in this study: maximum likelihood, binary logistic regression, and probability equilibrium.

Maximum Likelihood Method

The maximum likelihood method (MLM) (not to be confused with the maximum likelihood estimation method for regression model fitting) is a widely accepted critical gap estimation method. It considers drivers' gap acceptance behavior to be consistent but heterogeneous. Each driver has their own fixed critical gap and will accept all gaps larger than that gap while rejecting smaller gaps. The critical gaps among the driver population are assumed to follow the lognormal distribution. For subject driver i , MLM has an accepted gap size a_i and a maximum rejected gap size r_i . The probability of an individual driver i having a critical gap between r_i and a_i is

$F(a_i) - F(r_i)$. The MLM's goal is to find the parameters (mean, μ , and variance, σ^2) of the lognormal distribution that can maximize the summed probabilities over all investigated drivers:⁽⁴⁴⁾

$$\text{maximize: } \prod_{i=1}^n (F(a_i) - F(r_i))$$

or:

$$\text{maximize: } \sum_{i=1}^n \text{Ln}(F(a_i) - F(r_i))$$

The optimization process gives the μ and σ^2 of the critical gap's lognormal distribution. With these parameters known, the mean and variance of the critical gap can be derived as:⁽⁴⁴⁾

$$E(t_c) = e^{\mu+0.5\sigma^2}$$

$$D(t_c) = E(t_c)^2(e^{\sigma^2} - 1)$$

The critical gap was determined with accepted and maximum rejected gap size for observed drivers. Hence, this method is "driver-based." The optimization process was usually achieved via an iterative process. Troutbeck⁽⁵⁶⁾ previously recommended a spreadsheet approach to simplify the calculation process. This study adopted the spreadsheet approach but improved it slightly to change the "driver-based" calculation to "trip-based," considering that a driver can appear in multiple investigated intersections and may behave very differently under different exposure conditions.

Binary Logistic Regression

Binary logistic regression (BLR) models the log probabilities of gap acceptance as a response of gap size:⁽⁵⁷⁾

$$\text{logit}(P_{\text{accept}}) = \log\left(\frac{P_{\text{accept}}}{1 - P_{\text{accept}}}\right) = \beta_0 + \beta \times \text{Gap Size}$$

or:

$$P_{\text{accept}} = \frac{1}{1 + e^{-(\beta_0 + \beta \times \text{Gap Size})}}$$

where:

P_{accept} is the probability of the subject driver accepting the gap;

β_0 is the intercept; and

β is the estimated parameter for factor *Gap Size*.

Critical gap is the gap size that will be equally accepted and rejected by drivers; i.e., $P_{accept} = P_{reject}$, or $P_{accept} = 1 - P_{accept}$. Thus, the critical gap determined using BLR is:

$$\beta_0 + \beta_1 \times \text{Critical Gap} = 0$$

$$\text{Critical gap} = \frac{-\beta_0}{\beta_1}$$

The researchers used SAS[®] to perform the BLR modeling.

Probability Equilibrium Method

This method is based on the probability equilibrium between the rejected and accepted gaps.⁽⁴⁶⁾ Compared to the maximum likelihood method, it does not make assumptions about critical gap distribution. It is not “driver-based” and uses all collected rejected gaps. The equilibrium is established from the cumulative distribution of the rejected and accepted gaps as:

$$\begin{cases} F(t) = F_r(t) \cdot F(t) + F_a(t) \cdot (1 - F(t)) \\ 1 - F(t) = (1 - F_r(t)) \cdot F(t) + (1 - F_a(t)) \cdot (1 - F(t)) \end{cases}$$

where:

$F_a(t)$ is the observed probability of a gap of length t to be accepted from the cumulative distribution function of the accepted gaps;

$F_r(t)$ is the observed probability of a gap of length t to be rejected from the cumulative distribution function of the rejected gaps;

$F(t)$ is the cumulative distribution function of the critical gap to be estimated.

The $F(t)$ of the critical gaps can be solved with $F_a(t)$ and $F_r(t)$:

$$F(t) = \frac{F_a(t)}{F_a(t) + 1 - F_r(t)} = 1 - \frac{1 - F_r(t)}{F_a(t) + 1 - F_r(t)}$$

In this method, all observed accepted and rejected gap lengths were recorded in order from smallest to largest. For each observed gap with length t_i , the critical gap $F(t_i)$ was calculated individually. The probability of the critical gap falling between t_{i-1} and t_i was estimated as $F(t_i) - F(t_{i-1})$. Then the estimated critical gap was the weighted sum of observed gap lengths and their associated probabilities:

$$E(t_c) = \sum_{i=1}^n \left(\frac{(t_{i-1} + t_i)}{2} \times (F(t_i) - F(t_{i-1})) \right)$$

$$D(t_c) = \sum_{i=1}^n \left(\frac{(t_{i-1} + t_i)}{2} \times (F(t_i) - F(t_{i-1}))^2 \right) - E(t_c)^2$$

Methods for Modeling Gap Acceptance Decision

In addition to the gap size, there are many other variables that affect drivers' gap acceptance decisions. This section describes the three methods (i.e., binary logistic regression, decision tree, and random forest) used to model the binary gap acceptance decision—the decision to accept or reject the gap—with a variety of collected independent variables in this study.

Binary Logistic Regression and Odds Ratio

Similar to the critical gap determination, binary logistic regression was also used for modeling gap acceptance decisions as a binary response. This analysis, however, considered not only gap sizes but also a variety of other independent variables in the modeling process:

$$P_{accept} = \frac{1}{1 + e^{-(\beta_0 + \beta_i \times X_i)}}$$

where:

- P_{accept} is the probability of the subject driver accepting the gap;
- β_0 is the intercept; and
- β_i is the estimated parameter for explanatory factor X_i .

This analysis also used SAS to perform the modeling. In addition to the estimated variable coefficients, SAS also estimated the associated odds ratios (ORs) to measure the association between independent variables and responses. The OR quantifies the changes in odds of accepting a particular gap for a significant variable in the model. The ORs were calculated with variable coefficients (β_i) as follows:

$$OR = e^{\beta_i}$$

Decision Tree and Random Forest

The decision tree approach, also referred to as classification and regression tree, is a popular supervised machine learning method that is capable of both classification and regression tasks.⁽⁵⁸⁾ Compared to logistic regression, decision tree is a non-parametric method and makes no assumption about data distribution. Compared to other non-parametric machine learning methods, the decision tree method is able to handle both numerical and categorical variables. The resulting tree structure provides direct visualization of the model, making it simple to interpret. In addition, the decision tree machine learning technique is designed to mirror the human decision-making process, therefore making it ideal for modeling drivers' gap acceptance decisions in this study. Figure 6 illustrates the concept of a decision tree.

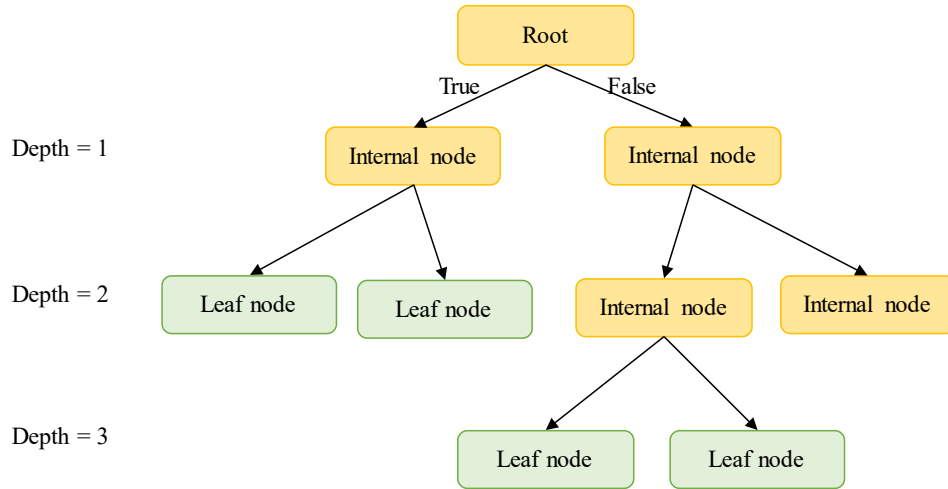


Figure 6. Flow chart. Decision tree scheme.

During a decision tree growing process, all training observations with different classes are input to the root node. The algorithm then selects a feature to split the observations in the parent node into different groups, or children nodes. Usually, the algorithm measures the impurity of the parent node and the weighted sum of children nodes, and selects the feature for splitting as that which can reduce the impurity to the greatest extent. This process runs in a top-down recursive manner until all observations in the node are of the same class (i.e., the impurity reaches its lowest). Researchers typically use Gini index or entropy to measure the node impurities:

$$Gini\ index = 1 - \sum [p(x)^2]$$

$$Entropy = - \sum p(x) \log_2(p(x))$$

where $p(x)$ is the proportion of observations with label x in the node.

Without control, a decision tree can grow very deep until the impurity of the node can no longer decrease. However, such a tree is often over-complicated and suffers severely from overfitting, which means that the model can closely fit the training data but has very low prediction power for new observations. In order to avoid overfitting, pruning is used to reduce the size of the decision tree; this process involves removing branches that contribute little to the overall model. For this purpose, researchers use a number of techniques, such as using the minimum observations on a node to be split, the minimum impurity change for a node to split, or the maximum depth of the tree.

In addition to the potential overfitting problem, a slight change in the training dataset or predictors for a decision tree can result in great variance in the output. The random forest method is developed based on the decision tree technique but is less susceptible to both overfitting and result variation at a cost of reduced interpretability.⁽⁵⁹⁾ When developing a random forest model, an algorithm creates multiple datasets using the bootstrap technique, where each dataset randomly takes a portion of the training data to develop an individual decision tree. The

individually and randomly developed decisions trees therefore form the so-called forest decision trees. The random forest approach outputs the averages or the majority votes of the estimation results based on the multiple decision trees it develops during the process (Figure 7).

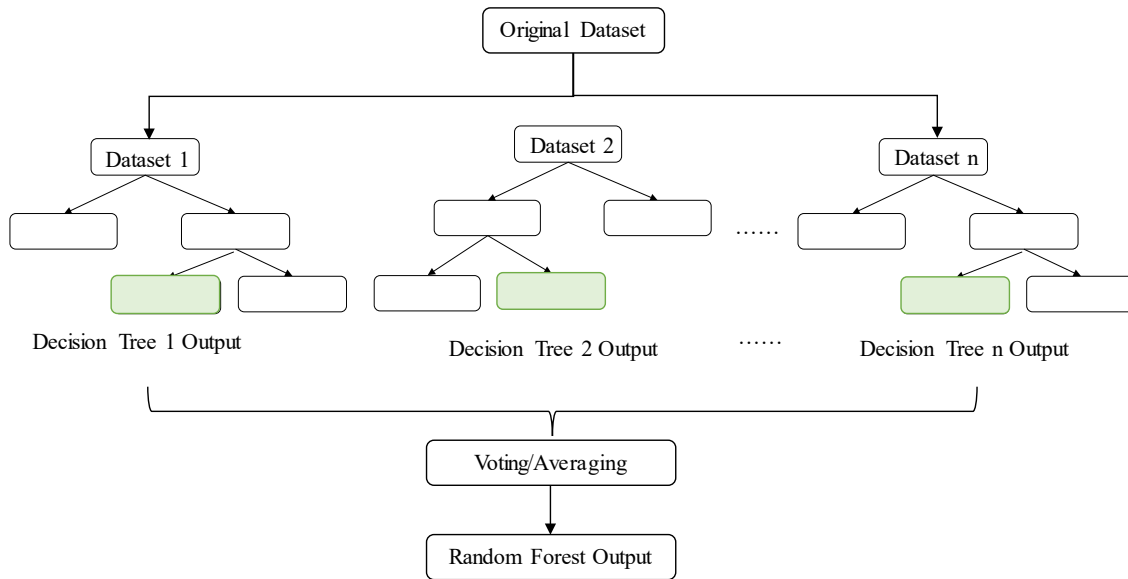


Figure 7. Flow chart. Conceptual design of the random forest approach.

Turning Behavior Modeling

The turning behavior analysis included the development of turning behavior profiles and the identification of factors affecting drivers' turning behaviors. For the latter, the project team used multiple linear regression to identify the correlations with measures related to the turning speeds and acceleration rates. For the former, Gaussian process regression was used to develop continuous vehicle kinematical curves.

The vehicle kinematics measures from different trips were time series data collected at different temporal/spatial locations. Within each trip, the measurements were collected following a predefined frequency (e.g., GPS at 1 Hz and kinematics measurements at 10 Hz) following its unique time axis. In addition, the turning movements from different trips at different intersections followed different paths. Each trip segment for the turning movement therefore had varying durations and distances. Due to these factors, it was necessary to normalize the kinematical measurements onto a single time/space coordinate system for each set of trips in order to profile the vehicle behaviors during the turning movements after accepting a gap at the studied intersections.

For this purpose, the project team first identified the portion of each trip that represented the turning movement after the driver accepted a gap at the studied intersections. Each trip segment was then divided into 10 equal intervals, and the data points of each trip segment within each interval were aggregated into a single value by obtaining the arithmetic mean. The data points were normalized in this manner for all turning movements of trips falling into each analysis scenario and were then used as the basis for profiling the vehicle kinematical behaviors.

The start and end points of the turning portion of each trip were determined primarily based on vehicle speed, acceleration rates, and trip videos. The start point was determined as the point in time when the driver started accelerating longitudinally and laterally from a complete stop after accepting a gap, or when the vehicle started accelerating from a rolling stop after checking and accepting a gap. The end point of the turning movement was determined as the point in time when the vehicle stopped accelerating laterally after merging into the major road traffic flow.

A straightforward way to determine the vehicle kinematic metrics at the 10 turning points would be to simply obtain the average values of the data points from different trips at each of the 10 locations. However, this simplistic approach may not yield reliable results in this case. First, the aggregated data points of different trips for the same interval (i.e., each of the 10 intervals) did not necessarily define a clear normal distribution, particularly when the datasets were relatively small. Therefore, the average values may not necessarily be the center of the data points where the values tend to concentrate. In addition, the aggregated values are discontinuous within each trip, and the goal of the research team was to identify a continuous, smooth curve for each kinematical variable.

To overcome these challenges, the project team used the Gaussian process regression for the construction of vehicle kinematical profiles during turning movements. Gaussian process regression is a nonparametric, Bayesian approach to regression that is considered a robust machine learning method for nonlinear, multivariate interpolation.⁽⁶⁰⁻⁶¹⁾ A Gaussian process is a set of variables such that, for each set of input \mathbf{x} , the joint distribution of the random variables $f(x_i)$ follows the multivariate Gaussian distribution. A Gaussian process regression model can be represented with a mean, which is often assumed to be zero, and a covariance function encoding correlations between random variable pairs $f(x_i)$ and $f(x_j)$:

$$\text{cov}(f(x_i), f(x_j)) = k(x_i, x_j)$$

where $k(x_i, x_j)$ is the covariance function.

The covariance function, or so-called kernel function in this case, captures the relationships between the input observations. A commonly used kernel function is the squared exponential kernel, which constrains nearby input observations to have highly correlated outputs:

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2l^2}\right)$$

where l is the length parameter.

In Gaussian process regression, the Gaussian process is used as a prior probability distribution in the context of Bayesian inference. During regression, the dataset containing all observations is divided into training and testing subsets and the method incorporates the constraints from training data to the testing set. According to the Gaussian process prior, the joint distribution of observed outputs and an unknown test output corresponding to a given input \mathbf{x} is:

$$\begin{bmatrix} Y \\ y \end{bmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} K(X, X) & K(X, x) \\ K(X, x)^T & K(x, x) \end{pmatrix}\right)$$

where

X = the input of the training set,

Y = the output of training set,

y = the output of testing set, and

K = the covariance kernel matrix where its entries correspond to the covariance function evaluated based on the observations.

To obtain a meaningful posterior, the process restricted the joint distribution to contain only those functions that are consistent with the observed output. Because the joint is Gaussian, the predictive distribution obtained by conditioning on the observed output, $p(y|Y)$ is also Gaussian, with mean and variance given by:

$$m(y) = YK(X, X)^{-1}K(X, x)$$

$$\sigma^2(y) = K(x, x) - K(X, x)^TK(X, X)^{-1}K(X, x)$$

The $m(y)$ is the inference for input x and $\sigma^2(y)$ is the variance.

This analysis allowed the researchers to develop continuous profiles for key driver behavior variables after they accept a gap based on scattered time series data. The driver behavioral variables analyzed in this part of the study included lateral acceleration rate and longitudinal acceleration rate.

CHAPTER 3. ANALYSIS RESULTS AND DISCUSSION

DATA OVERVIEW

The research team collected 1,170 gaps, including 466 lags in total from 466 NDS trips. Table 6 summarizes the basic information about the collected accepted and rejected gaps. Note that the project team collected considerably more rejected than accepted gaps/lags. This was primarily because many drivers rejected multiple gaps prior to accepting one. The average accepted and rejected gap sizes were 8.81 seconds and 2.84 seconds, respectively, for all trips.

Table 6. Basic statistics measures of collected gaps.

Statistic	Accepted Gaps/Lags	Rejected Gaps/Lags
Count	270	900
Minimum	2.535	0
Maximum	25.294	15.612
Mean	8.810	2.837
Median	8.289	2.347
Std Deviation	3.069	1.906

Figure 8 shows the frequencies and cumulative distributions of the 1,170 accepted and rejected gaps.

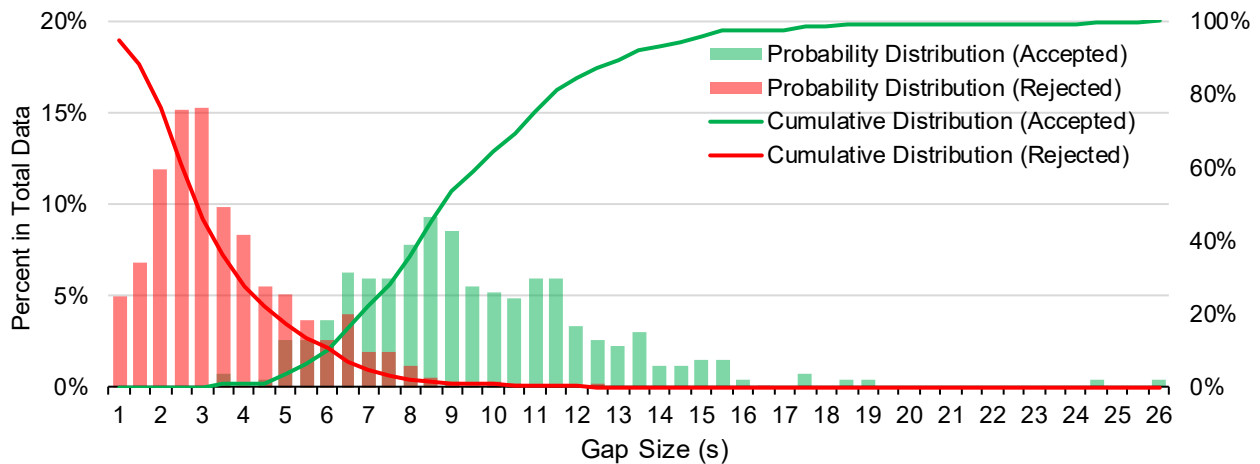


Figure 8. Graph. Probability and cumulative distributions of accepted and rejected gaps.

As illustrated in Figure 8, it was clear that studied drivers rejected all gaps smaller than 2.0 seconds, while the majority of gaps larger than 12.0 seconds were accepted. Gaps with sizes falling outside of the 2.0–12.0 second range would therefore provide limited information with regard to drivers’ gap acceptance behaviors. During this study, the researchers therefore only used gaps/lags between 2 and 12 seconds for analysis. Table 7 further provides some basic statistics of the accepted and rejected gaps with sizes between 2.0 and 12.0 seconds, and Figure 9 shows their frequencies and cumulative distributions.

Table 7. Basic statistics of the collected gaps/lags between 2 and 12 seconds.

Statistic	Accepted Gaps	Rejected Gaps
Count	235	550
Minimum	2.535	2.000
Maximum	11.945	11.878
Mean	7.973	3.829
Median	7.931	3.337
Standard Deviation	2.031	1.688

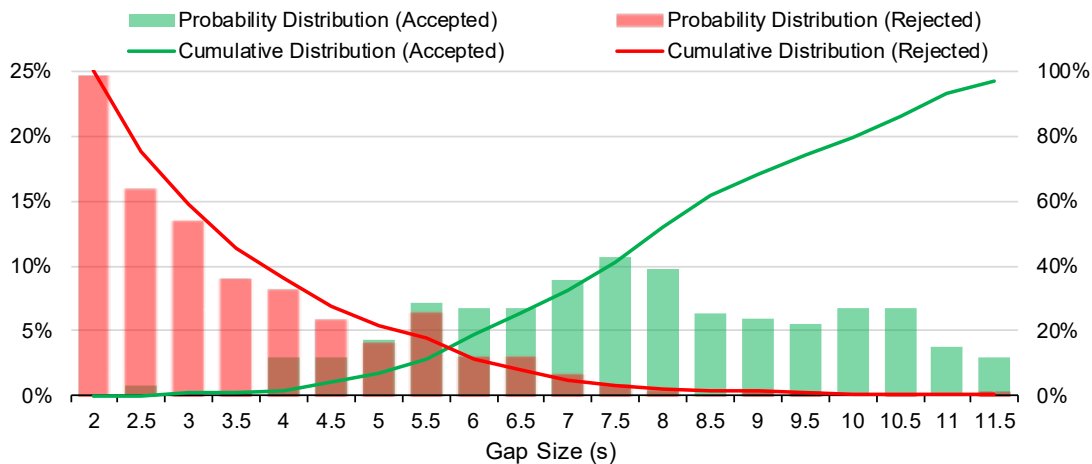


Figure 9. Graph. Probability and cumulative distributions of accepted and rejected gaps between 2.0–12.0 s.

CRITICAL GAP DETERMINATION

The research team used three methods to determine the critical gaps for each scenario based on the SHPR 2 NDS data: included maximum likelihood, binary logistic regression, and probability equilibrium. This section describes the results by method, followed by a comparison of the methods and a discussion of the different critical gap values by scenario.

The maximum likelihood method determined the critical gap with the accepted and maximum rejected gap pairs based on the analyzed trips. For this analysis, we used the 235 trips that involved accepted gaps between 2.0 and 12.0 seconds. Note that this method is based on the assumption that drivers would behave consistently during their gap acceptance decision process. This assumed that drivers would not reject a gap but then accept a subsequent gap of a smaller size than the rejected gap. However, 15 out of the 235 accepted gaps contradicted this assumption. Troutbeck previously suggested reassigning the maximum rejected gap a value slightly less than the accepted gap. However, we did not follow this suggestion, instead excluding these 15 instances and performing the critical gap determination based on the 220 accepted gaps instead. The estimation results using the maximum likelihood method are shown in Table 8 and Figure 10. As shown, the analysis revealed an overall critical gap of 5.4 seconds, with a 6.2-second gap for left-turn movements and 4.9-second gap for right-turn movements.

Table 8. Critical gap sizes by maximum likelihood method.

Scenario	Critical Gap (s)	Standard Deviation
All trips	5.390	1.319
Right-turn trips	4.877	1.096
Left-turn trips	6.164	1.300
Left-turn trips with TWLTL	6.025	1.293
Left-turn trips without TWLTL	6.583	1.253
Right-turn trips with leading vehicle	4.846	1.003
Right-turn trips without leading vehicle	4.899	1.125
Left-turn trips with leading vehicle	5.633	1.488
Left-turn trips without leading vehicle	6.332	1.208

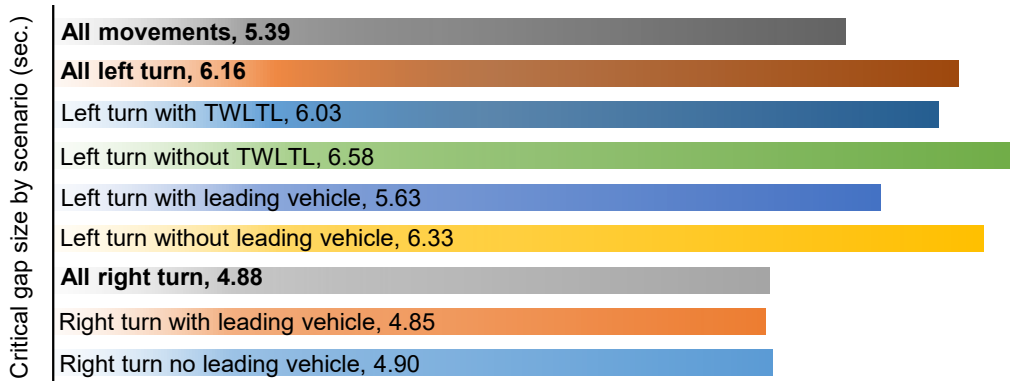


Figure 10. Graph. Critical gap sizes by scenario (maximum likelihood method).

Different from the “trip-based” maximum likelihood method, the binary logistic regression used all accepted and rejected gaps. The modeling process was implemented in SAS and the results are shown in Table 9 and Figure 11. The results showed an overall critical gap of 5.7 second with a 5.4-second gap for right-turn movements and 6.1-second gap for left-turn movements.

Table 9. Critical gap sizes by binary logistic regression.

Scenario	Critical Gap (s)
All 785 gaps	5.715
423 gaps for right-turn trips	5.411
362 gaps for left-turn trips	6.155
243 gaps for left-turn trips with TWLTL	6.176
119 gaps for left-turn trips without TWLTL	6.141
87 gaps for right-turn trips with leading vehicle	5.638
336 gaps for right-turn trips without leading vehicle	5.349
75 gaps left-turn trips with leading vehicle	6.189
287 gaps for left-turn trips without leading vehicle	6.198

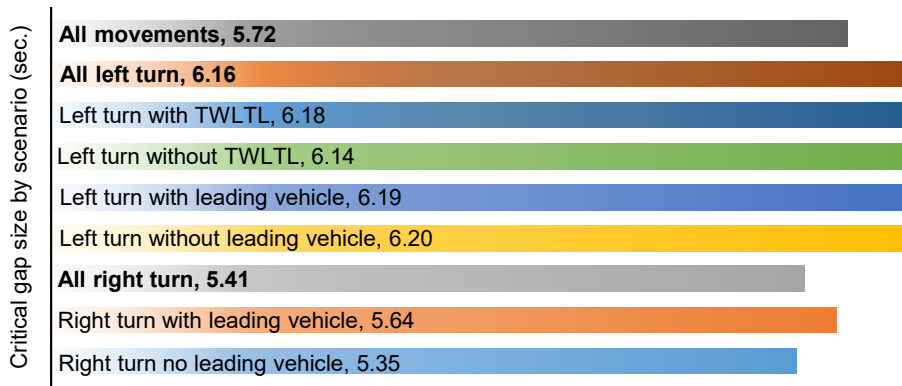


Figure 11. Graph. Critical gap sizes by scenario (binary logistic regression).

The results are presented in Table 10 and Figure 12. The probability equilibrium method also models the drivers' gap acceptance behavior based on all accepted and rejected gaps. Overall, this method showed a critical gap of 5.8 seconds for all trips. In addition, it suggested a critical gap of 5.5 seconds for right-turn movements and 6.2 seconds for left-turn movements.

Table 10. Critical gap sizes by probability equilibrium method.

Scenario	Critical gap (s)	Standard Deviation
All 785 gaps	5.782	1.241
423 gaps for right-turn trips	5.466	1.097
362 gaps for left-turn trips	6.238	1.324
243 gaps for left-turn trips with TWLTL	6.240	1.341
119 gaps for left-turn trips without TWLTL	6.232	1.238
87 gaps for right-turn trips with leading vehicle	5.759	0.883
336 gaps for right-turn trips without leading vehicle	5.365	1.147
75 gaps left-turn trips with leading vehicle	5.981	1.485
287 gaps for left-turn trips without leading vehicle	6.400	1.164

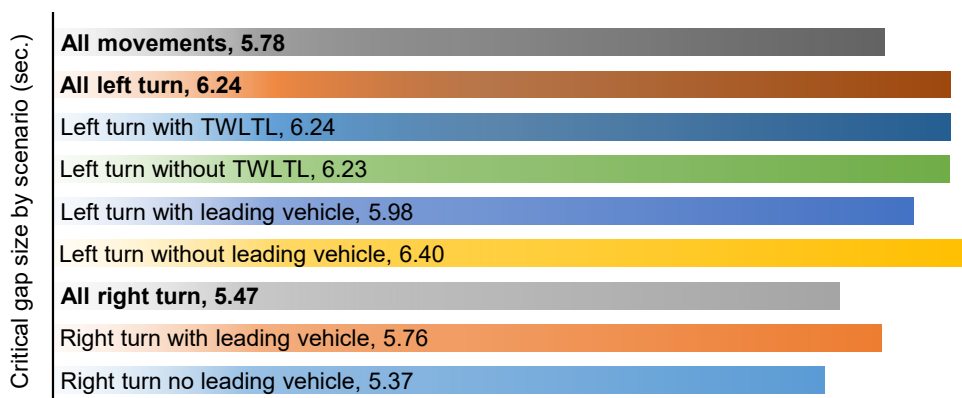


Figure 12. Graph. Critical gap sizes by scenario (probability equilibrium method).

Table 11 and Figure 13 compare the critical gap values for the different scenarios by the three different approaches. Overall, the three methods showed an average critical gap size of 5.63 seconds for all trips and both right-turn and left-turn movements. For right-turn movements, the methods suggested an average critical gap of 5.25 seconds. For left-turn movements, the methods

showed an average critical gap of 6.19 seconds, almost a second longer than that for right-turn movements. In addition, the results showed that:

- There was a shorter critical gap (i.e., 5.93 seconds versus 6.31 seconds) for left-turn movements when a leading vehicle was present. As previously mentioned, the presence of leading vehicles was highly significantly correlated with the presence of following vehicles. This finding seemed to suggest that drivers would take a shorter gap when left turning onto the major roadway when there was more traffic on the minor approach. For right-turning movements, however, this effect was much less evident. In fact, right-turning drivers tended to take a slightly longer gap (i.e., 5.41 seconds versus 5.20 seconds) when leading or following vehicles were present.
- Overall, the presence of TWLTL corresponded to a shorter critical gap size (i.e., 6.15 seconds versus 6.32 seconds) for left-turn traffic. This could suggest that the presence of a TWLTL provides potential space for left-turning drivers to hide upon conflict with major street traffic and therefore increases confidence during drivers' gap acceptance process.
- In most cases, the maximum likelihood method produced lower critical gap sizes. In particular, the critical gap sizes for the right-turn scenarios estimated by this method were shorter than the other two methods. On the other hand, both the binary logistic regression and the probability equilibrium methods showed comparable results for all scenarios. Both the logistic regression and probability equilibrium methods modeled the critical gaps using all collected gaps, while the maximum likelihood method used the studied trips' accepted and maximum rejected gap pairs. In this respect, the former made better use of the collected data than the latter. However, the maximum likelihood was exempted from the biases caused by small rejected gaps. Note that when taking the standard deviation of the critical gaps into consideration, none of these differences was considered significant.

Table 11. Comparison of critical gaps determined by different methods.

Scenario	Critical Gap Size (s)			
	Maximum Likelihood	Logistic Regression	Probability Equilibrium	Mean
All right turn	4.877	5.411	5.466	5.251
Right turn no leading vehicle	4.899	5.349	5.365	5.204
Right turn with leading vehicle	4.846	5.638	5.759	5.414
All left turn	6.164	6.155	6.238	6.186
Left turn without leading vehicle	6.332	6.198	6.400	6.310
Left turn with leading vehicle	5.633	6.189	5.981	5.934
Left turn without TWLTL	6.583	6.141	6.232	6.319
Left turn with TWLTL	6.025	6.176	6.240	6.147
All movements	5.390	5.715	5.782	5.629

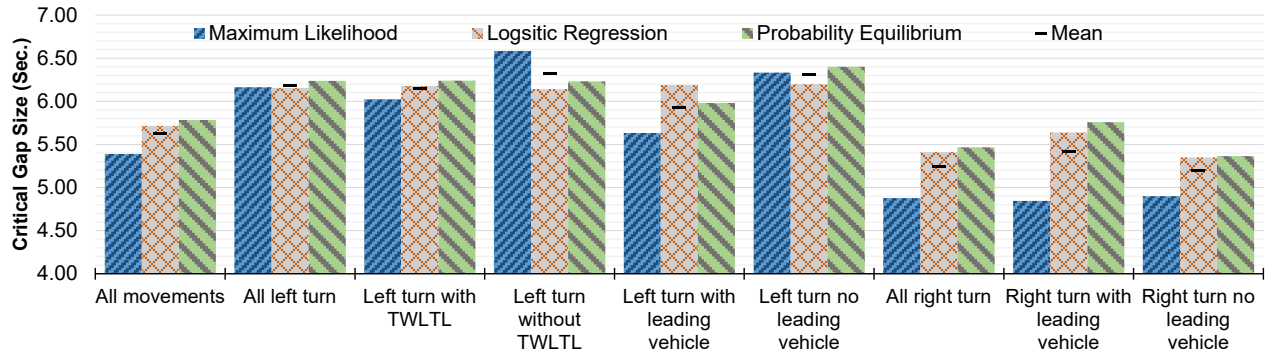


Figure 13. Chart. Comparison of critical gap sizes determined by different methods.

FACTORS AFFECTING GAP ACCEPTANCE DECISION

As previously described, the project team conducted binary logistic regression, decision tree, and random forest analyses to model the gap acceptance decision-making process for a number of scenarios. Similar to the critical gap determination, the project team focused on the 785 gaps collected with a gap length between 2.0 and 12.0 seconds.

Binary Logistic Regression Analysis Results

Table 12 and Table 13 list the summary performance measures and overall model statistics, respectively, for the logistic regression models developed by scenario. Table 14–Table 22 further list the significant parameters included in each model and their associated ORs. With the limited number of samples, to reduce Type II error (failure to reject the null hypothesis when there is an effect), the team used a 10% level of significance for the regression analysis. Readers should note the following regarding the aforementioned tables:

- In the modeling process, a reference level was set for each categorical variable. Parameter estimates and ORs of other values for the same variable were calculated by comparing them against the reference level. The reference level was selected as that which depicted the normal/optimal driving conditions if applicable.
- SAS used maximum likelihood estimations for model estimation and outputted the Akaike information criterion (AIC), Schwarz criterion (SC), and the $-2 \log$ likelihood ($-2 \log L$) as indicators of the overall model goodness of fit (Table 13). These goodness-of-fit measures are generally useful when comparing different models developed using the same dataset; smaller values indicate better models. The measure values themselves may not indicate how well the models described the original data.
- SAS also used likelihood ratio tests, Score tests, and Wald tests to show the joint significance of the independent variables (Table 12). In addition, in order to better understand how well the model described the data and compare the performance of the logistic regression models later with the non-parametric decision tree and random forest models, the researchers applied the regression models to the original datasets to obtain model accuracies.

- In Table 14–Table 22, a positive parameter estimate for a variable or a variable value indicates that the presence or increased level of that variable/variable value correlated to increased probability of a driver accepting a given gap.
- In Table 14–Table 22, an OR for a binary variable refers to the odds of the subject driver accepting a given gap when the condition described by that variable is true as compared to false. An OR for a categorical variable is the odds of a driver accepting a given gap with a variable value against that when the reference level was present instead. An OR calculated for a continuous or an ordinal variable is the odds of gap acceptance compared to the odds of acceptance with a variable value that is one unit or one level lower.

Table 12. Overall model performance summary.

Analysis Scenario	Frequency	Accuracy	Pr > ChiSq		
			Likelihood Ratio	Score	Wald
All movements	785	87.52%	< .0001	< .0001	< .0001
All right turn	423	88.18%	< .0001	< .0001	< .0001
All left turn	362	88.40%	< .0001	< .0001	< .0001
Left turn with TWLTL	243	84.36%	< .0001	< .0001	< .0001
Left turn no TWLTL	119	97.48%	< .0001	< .0001	0.2692
Right turn with leading vehicle	87	90.80%	< .0001	< .0001	0.0052
Right turn no leading vehicle	336	89.58%	< .0001	< .0001	< .0001
Left turn with leading vehicle	75	88.00%	< .0001	< .0001	0.0021
Left turn no leading vehicle	287	90.24%	< .0001	< .0001	< .0001

Table 13. Model fit statistics for all logistic regression models.

Analysis Scenario	AIC		SC		-2 Log L	
	Intercept Only	Intercept and Covariates	Intercept Only	Intercept and Covariates	Intercept Only	Intercept and Covariates
All movements	1,090.2	458.8	1,094.9	519.5	1088.241	432.8
All right turn	588.4	256.1	592.5	292.5	586.4	238.1
All left turn	503.8	206.5	507.7	237.6	501.8	190.5
Left turn with TWLTL	338.9	152.1	342.4	173.1	336.9	140.1
Left turn no TWLTL	167.0	33.8	169.7	53.3	165.0	19.8
Right turn with leading vehicle	122.6	44.8	125.1	59.6	120.6	32.8
Right turn no leading vehicle	467.8	202.3	471.6	229.1	465.8	188.3
Left turn with leading vehicle	106.0	54.9	108.3	68.8	104.0	42.9
Left turn no leading vehicle	399.9	151.3	403.5	169.6	397.9	141.3

Table 14. Significant variables and ORs – all movements.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq	OR (95% CI)
Intercept	-7.8973	0.7846	101.3096	<.0001	-
Is Gap	1.6701	0.3035	30.2766	<.0001	5.313 (2.931 - 9.632)
Gap Size	1.2618	0.0908	192.9732	<.0001	3.532 (2.956 - 4.220)
Wait Time	-0.0254	0.0117	4.7114	0.0300	0.975 (0.953 - 0.998)
Weather	-0.8034	0.4510	3.1731	0.0749	0.448 (0.185 - 1.084)
Turn Type	-1.1271	0.3402	10.9758	0.0009	0.324 (0.166 - 0.631)
Leading Vehicle	1.3679	0.3543	14.9035	0.0001	3.927 (1.961 - 7.864)
Major Road AADT	0.078	0.028	8.0464	0.0046	1.081 (1.023 - 1.142)
Major Approach Left Turn Lane-Right	-1.1817	0.3263	13.1163	0.0003	0.307 (0.162 - 0.581)
Major Road TWLTL	-1.1794	0.3338	12.4857	0.0004	0.307 (0.160 - 0.591)
Minor Road On-Street Parking-Left Side	-1.2803	0.5465	5.4886	0.0191	0.278 (0.095 - 0.811)
Intersection Alignment-Skew to Right	0.1060	0.3216	0.1086	0.7417	0.651 (0.259 - 1.635)
Intersection Alignment-Skew to Left	-0.6411	0.2694	5.6642	0.0173	0.308 (0.153 - 0.622)

Table 15. Significant variables and ORs – right turn.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq	OR (95% CI)
Intercept	-7.5364	0.9061	69.1789	<.0001	-
Is Gap	1.5426	0.3880	15.8108	<.0001	4.677 (2.186 - 10.004)
Gap Size	1.3599	0.1316	106.8534	<.0001	3.896 (3.010 - 5.042)
Leading Vehicle	0.6930	0.4095	2.8637	0.0906	2.000 (0.898 - 4.462)
Major Approach Left Turn Lane-Right	-1.1792	0.3824	9.5107	0.0020	0.308 (0.145 - 0.651)
Major Road TWLTL	-0.8054	0.3971	4.1134	0.0425	0.447 (0.205 - 0.973)
Intersection Alignment-Skew to Right	-0.3548	0.4680	0.5748	0.4484	0.372 (0.091 - 1.523)
Intersection Alignment-Skew to Left	-0.2802	0.3041	0.8493	0.3567	0.400 (0.193 - 0.829)

Table 16. Significant variables and ORs – left turn.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq	OR (95% CI)
Intercept	-9.8343	1.1844	68.9430	<.0001	-
Is Gap	1.5302	0.4526	11.4324	0.0007	4.619 (1.903 - 11.214)
Gap Size	1.2214	0.1329	84.4220	<.0001	3.392 (2.614 - 4.402)
Wait Time	-0.0342	0.0129	7.0253	0.0080	0.966 (0.942 - 0.991)
Weather	-1.1693	0.6269	3.4792	0.0621	0.311 (0.091 - 1.061)
Leading Vehicle	2.0525	0.5329	14.8317	0.0001	7.787 (2.740 - 22.132)
Major Road No. of Through Lanes	1.0860	0.3898	7.7608	0.0053	2.962 (1.380 - 6.360)
Major Road Crosswalk-Left Approach	1.8090	1.0081	3.2199	0.0727	6.105 (0.846 - 44.034)

Table 17. Significant variables and ORs – left turn with TWLTL.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq	OR (95% CI)
Intercept	-8.0327	1.0884	54.4720	<.0001	-
Is Gap	0.8740	0.4822	3.2855	0.0699	2.397 (0.931 - 6.167)
Gap Size	1.1326	0.1449	61.1144	<.0001	3.104 (2.336 - 4.123)
Wait Time	-0.0330	0.0139	5.6348	0.0176	0.967 (0.941 - 0.994)
Leading Vehicle	1.8607	0.5598	11.0498	0.0009	6.428 (2.146 - 19.256)
Major Road Sidewalk	1.7013	0.5892	8.3378	0.0039	5.481 (1.727 - 17.395)

Table 18. Significant variables and ORs – left turn without TWLTL.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq	OR (95% CI)
Intercept	-37.0314	13.8729	7.1253	0.0076	-
Is Gap	16.0587	6.3208	6.4546	0.0111	>999.999 (39.256->999.999)
Gap Size	4.4678	1.6413	7.4099	0.0065	87.162 (3.494 - >999.999)
Major Road Crosswalk Left Approach	11.9387	4.6040	6.7244	0.0095	>999.999 (18.451 - >999.999)
Major Road Sidewalk	5.8406	2.7393	4.5462	0.0330	343.983 (1.603 - >999.999)
Minor Road On-Street Parking-Right Side	-4.8009	2.4938	3.7060	0.0542	0.008 (<0.001 - 1.091)
Lighting	-8.2874	3.6379	5.1898	0.0227	<0.001 (<0.001 - 0.314)

Table 19. Significant variables and ORs – right turn with leading vehicle.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq	OR (95% CI)
Intercept	-15.2763	4.2185	13.1136	0.0003	-
Is Gap	2.5209	1.2289	4.2078	0.0402	12.440 (1.119 - 138.326)
Gap Size	2.2469	0.5614	16.0169	<.0001	9.458 (3.147 - 28.425)
Weather	-7.7954	2.4801	9.8799	0.0017	<0.001 (<0.001 - 0.053)
Annual Miles Traveled	1.4552	0.6296	5.3417	0.0208	4.286 (1.248 - 14.722)
Major Road TWLTL	-3.8049	1.2896	8.7048	0.0032	0.022 (0.002 - 0.279)

Table 20. Significant variables and ORs – right turn without leading vehicle.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq	OR (95% CI)
Intercept	-7.7654	1.0148	58.5538	<.0001	-
Is Gap	2.0367	0.5101	15.9391	<.0001	7.665 (2.820 - 20.833)
Gap Size	1.3299	0.1445	84.6770	<.0001	3.781 (2.848 - 5.019)
Wait Time	-0.0746	0.0357	4.3594	0.0368	0.928 (0.865 - 0.995)
Major Approach Left Turn Lane-Right	-0.6933	0.4070	2.9019	0.0885	0.500 (0.225 - 1.110)
Intersection Alignment-Skew to Right	-0.6068	0.4906	1.5298	0.2161	0.265 (0.111 - 0.632)
Intersection Alignment-Skew to Left	-0.3598	0.3413	1.1109	0.2919	0.265 (0.111 - 0.632)

Table 21. Significant variables and ORs – left turn with leading vehicle.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq	OR (95% CI)
Intercept	-8.9683	2.1769	16.9728	<.0001	-
Is Gap	1.7765	0.8756	4.1162	0.0425	5.909 (1.062 - 32.875)
Size	1.2266	0.2842	18.6251	<.0001	3.410 (1.953 - 5.952)
Lighting	2.6469	1.1353	5.4354	0.0197	14.110 (1.525 - 130.591)
Weather	-2.1911	1.1740	3.4833	0.0620	0.112 (0.011 - 1.116)
Major Road Sidewalk	2.8872	1.3861	4.3384	0.0373	17.942 (1.186 - 271.500)

Table 22. Significant variables and ORs – left turn without leading vehicle.

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr>ChiSq	OR (95% CI)
Intercept	-9.8401	1.3429	53.6924	<.0001	-
Is Gap	1.7452	0.5896	8.7616	0.0031	5.727 (1.803 - 18.189)
Size	1.2524	0.1535	66.5792	<.0001	3.499 (2.590 - 4.727)
Wait Time	-0.0611	0.0301	4.1255	0.0422	0.941 (0.887 - 0.998)
Major Road AADT	0.1010	0.0420	5.8637	0.0155	1.106 (1.019 - 1.203)

The following summarizes the findings relevant to the identified significant factors:

- **Gap size.** Gap size is undoubtedly a significant factor affecting gap acceptance behavior in all analysis scenarios. Overall, the modeling results showed that an increment of 1 second in gap size increased the probability of gap acceptance by about 3.5 times. This effect was also found to be more significant for right-turn trips (i.e., OR = 3.9) than left-turn trips (i.e., OR = 3.4).
- **Gap versus lag.** The variable Is Gap identifies if a gap data point is actually a gap or a lag in the dataset, with 1 indicating a gap. This variable was identified to be significant in all analysis scenarios, suggesting that it has a significant effect on gap acceptance behavior. As shown in the models, for all analysis scenarios, drivers were more likely to accept a gap compared to a lag of the same size. This may be because drivers need additional time to observe the intersection condition after arriving there.
- **Waiting time.** Waiting time was found to be significant in the overall model, and the models for all left-turn trips, left turn with TWLTL trips, left turn without leading vehicle trips, and right turn without leading vehicle trips. The associated negative parameter estimates suggest that drivers were more likely to reject a gap as they wait longer, although the ORs in all aforementioned models were only slightly below 1 (i.e., 0.93–0.98). This finding appears to be contradictory to the common belief that the more drivers wait, the more impatient they become and therefore the shorter gaps they may take. A plausible explanation for this finding may be that more conservative drivers who take longer gaps tend to wait longer.

To further explore this finding, Figure 14 and Figure 15 are scatter plots of accepted and rejected gaps by waiting time, respectively. The plot for accepted gaps seems to show a slight downward trend, indicating longer waiting times correlated with shorter accepted gaps. The plot for rejected gaps suggests that drivers who waited longer tended to reject longer gaps initially, but after waiting longer (for more than 20 seconds), drivers tended to reject shorter gaps.

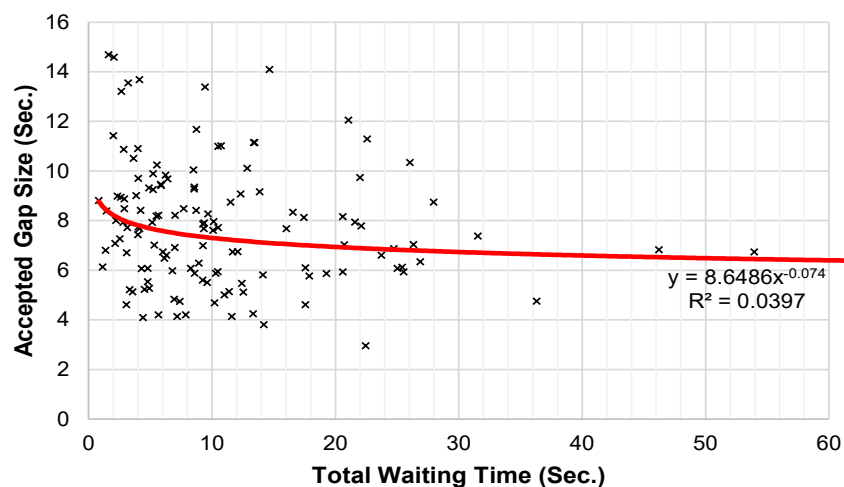


Figure 14. Scatter plot. Accepted gaps by waiting time.

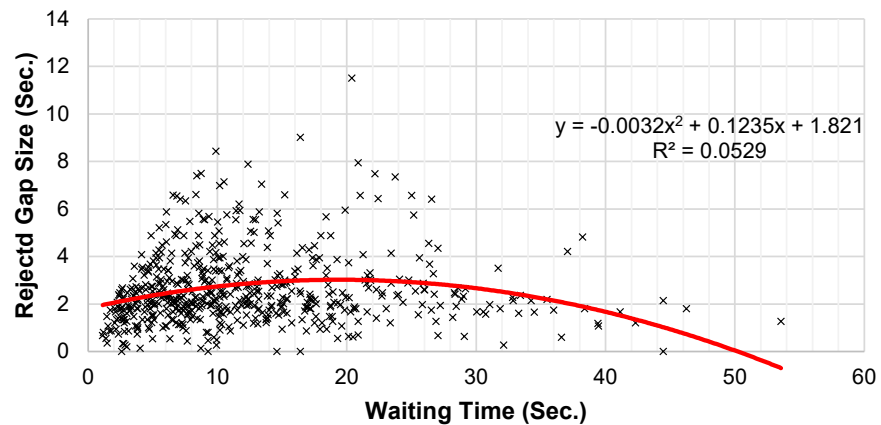


Figure 15. Scatter plot. Rejected gaps by waiting time.

- **Weather.** Weather was found to be significant in the overall model, and models for all left-turn trips, right turn with leading vehicle trips, and left turn with leading vehicle trips. In all these models, drivers tended to accept longer gaps during adverse weather conditions. Note that this analysis only included trips with minor weather conditions (e.g., light rain or fog) since severe weather conditions affected visibility considerably and therefore gap data could not be collected as video images became unclear.
- **Movement type.** The modeling result suggested that under the same exposure conditions, left-turning drivers were less likely to accept a given gap compared to right-turning drivers (i.e., OR = 0.3). Equivalently, drivers accepted larger gaps to complete a left-turn maneuver compared to a right-turn maneuver.
- **Presence of leading vehicle and/or following vehicle.** The analysis result suggests that the existence of leading vehicles and/or following vehicles increased the probability of a driver accepting the given gap (i.e., OR = 3.9 in the overall model). This phenomenon was even more evident for drivers making a left turn (i.e., OR = 7.8) as opposed to drivers making a right turn (i.e., OR = 2.0). This finding is consistent with what was reported by many previous studies. This behavior could be due to one or more of the following factors:
 - Drivers had sufficient time to observe the intersection and traffic pattern while waiting and therefore required shorter time to make a gap acceptance decision.
 - The presence of following vehicles may cause psychological pressure to the driver, prompting them to accept a shorter gap.
- **Major road AADT.** The modeling results showed that increased AADT on major roads correlated with increased probability of accepting a given gap. In another words, increased AADT correlated with shorter gaps overall.

- **Left-turn lane on right major approach.** Based on the overall model and the model for all right-turn trips, the presence of an exclusive left-turn lane on the right major approach increased the probability of a driver rejecting a given gap. Note that there is no conflict between right-turn vehicles and vehicles who attempt to turn left from the right major approach. This result may suggest that drivers need extra observation time when there is a left-turn lane on the major road.
- **Presence of TWLTL on major road.** The presence of TWLTL on the major road was associated with lower probability of gap acceptance in general and for models of right-turning trips. The presence of a TWLTL could be a distraction to drivers and also increase the complexity of the intersection and therefore make drivers more likely to reject a given gap.
- **Intersection alignment.** According to multiple models, minor-road drivers were more likely to accept a given gap if the major and minor roads are intersected perpendicularly. A left-skewed intersection alignment was found to decrease the probability of gap acceptance about three times (e.g., OR = 0.31 based on the overall model). In addition, right-turning trips seemed to be more susceptible to this effect than left-turning trips.
- **Number of through lanes on major road.** The number of through lanes on the major road was found to play a significant role on left-turn drivers' gap acceptance behavior. The modeling results showed that for each additional through lane on the major road, drivers were about three times more likely to accept a given gap (i.e., OR = 3.0 based on the model for all left-turning trips).
- **Major road crosswalk on left approach.** The results showed that the presence of a crosswalk on the left approach of the major road increased the probability of left-turning drivers accepting a given gap.

Decision Tree Modeling Results

During this study, the research team developed decision tree models for different analysis scenarios with the Scikit-learn toolkit in Python.⁽⁶²⁾ The dataset for each analysis scenario was split into 70% training and 30% testing sets. Observations of different classes—i.e., gap acceptance and gap rejection—were balanced by adjusting their associated weights. The team used entropy for node impurity measurement and trimmed the tree by controlling the minimum impurity decreased and minimum sample size on a splitting node. The decision tree exported importance scores of analyzed features, which were calculated as the normalized reduction of the impurity brought by that feature. The models' information, identified important features, and their associated importance scores are presented in Table 23.

Table 23. Model information and important features of decision tree models.

Model Information	Feature Importance
Model scenario: All 785 gaps Minimum decreased impurity: 0.01, minimum samples split: 10 Accuracy in training set: 88.71% Accuracy in testing set: 88.14%	<ul style="list-style-type: none"> • Gap Size: 73.14% • Is Gap: 5.33% • Wait Time: 21.53%
Model scenario: 423 gaps for right-turn trips Minimum decreased impurity: 0.02, minimum samples split: 10 Accuracy in training set: 92.23% Accuracy in testing set: 83.46%	<ul style="list-style-type: none"> • Gap Size: 62.00% • Wait Time: 24.93% • Is Gap: 4.36% • Major AADT: 8.70%
Model scenario: 362 gaps for left-turn trips Minimum decreased impurity: 0.02, minimum samples split: 10 Accuracy in training set: 90.12% Accuracy in testing set: 83.49%	<ul style="list-style-type: none"> • Gap Size: 85.44% • Wait Time: 14.56%
Model scenario: 243 gaps for left-turn trips with TWLTL Minimum decreased impurity: 0.03, minimum samples split: 10 Accuracy in training set: 90.00% Accuracy in testing set: 80.82%	<ul style="list-style-type: none"> • Gap Size: 85.08% • Wait Time: 7.83% • Major Sidewalk: 7.09%
Model scenario: 119 gaps for left-turn trips without TWLTL Minimum decreased impurity: 0.05, minimum samples split: 10 Accuracy in training set: 91.57% Accuracy in testing set: 80.56%	<ul style="list-style-type: none"> • Gap Size: 73.69% • Major AADT: 14.00% • Major Crosswalk Left: 12.31%
Model scenario: 87 gaps for right-turn trips with leading vehicle Minimum decreased impurity: 0.05, minimum samples split: 10 Accuracy in training set: 93.33% Accuracy in testing set: 74.07%	<ul style="list-style-type: none"> • Gap Size: 69.67% • Annual Miles Traveled: 18.08% • Major Sidewalk: 12.26%
Model scenario: 336 gaps for right-turn trips without leading vehicle Minimum decreased impurity: 0.03, minimum samples split: 10 Accuracy in training set: 88.51% Accuracy in testing set: 88.12%	<ul style="list-style-type: none"> • Gap Size: 59.79% • Is Gap: 17.04% • Wait Time: 23.17%
Model scenario: 75 gaps for left-turn trips with leading vehicle Minimum decreased impurity: 0.05, minimum samples split: 10 Accuracy in training set: 90.38% Accuracy in testing set: 73.91%	<ul style="list-style-type: none"> • Gap Size: 75.75% • Major Road Alignment = Curve_out: 9.30% • Is Gap: 7.71% • Annual Miles Traveled: 7.24%
Model scenario: 287 gaps for left-turn trips without leading vehicle Minimum decreased impurity: 0.03, minimum samples split: 10 Accuracy in training set: 92.50% Accuracy in testing set: 91.95%	<ul style="list-style-type: none"> • Gap Size: 78.77% • Is Gap: 9.48% • Wait Time: 11.75%

As shown in Table 23, the developed decision trees for different analysis scenarios had varying prediction performances. All except two of the decision trees had a prediction accuracy higher than 80% for the testing data. The lower accuracy rates in many scenarios for the testing data as compared to the training sets suggested that the models overfitted the training sets. The two models with the severe overfitting problem (i.e., right turn with leading vehicle and left turn with leading vehicle) were developed with very small training sets (61 and 53, respectively). Other models were found to suffer from the overfitting problem (but at a much lesser degree) although a combination of trimming methods was used in the modeling process. The variables Is Gap, Gap Size, and Wait Time were identified as important by most of the decision trees. In addition, the variable Gap Size was identified to be the most important feature that reduced the tree impurity by about 60%–85% among constructed models.

The developed decision trees are illustrated in Figure 16–Figure 24. Tree nodes in blue represent gap acceptance while nodes in orange represent gap rejection. A node in a lighter color indicates that the purity of that node is low. Readers should note that the decision tree models were highly dependent on the training sets. Shuffling the training and testing datasets or a slight change in the predictor may result in great differences in the obtained tree structures. The structures presented in this section are only solutions that are related to the results shown in Table 23. Figure 16–Figure 24 may provide a better understanding of the roles that different features played on drivers’ intersection gap acceptance behavior:

- **Gap Size.** Gap Size was selected as the splitting feature for the root node of all developed models, suggesting its overwhelming importance among all analyzed features. Drivers tended to accept longer gaps and reject smaller gaps. The gap size value documented in the root node was the size that best split accepted and rejected gap observations. To a certain degree, it could be considered as the critical gap for that analysis scenario. However, the researchers found that this value could vary considerably if shuffling the training and testing dataset or if making a slight change in the dataset.
- **Wait Time.** Previously, the logistic regression results showed that longer waiting time was associated with higher probability of gap rejection. The decision tree models provided more information on how wait time influenced drivers’ gap acceptance behavior. In the decision tree developed for all trips (Figure 16), wait time played a determinant role when the exposed gap size was larger than 4.591 seconds. Within this range, observations with wait time shorter than 4.625 seconds were all accepted, while longer wait times seemed to be more often associated with gap rejection. However, further looking into observations with a wait time longer than 4.625 seconds showed that wait times longer than 9.03 seconds were more likely to be associated with gap acceptance. A similar trend could also be found in other decision tree models with the feature Wait Time (Figure 17 and Figure 18). This result seemed to suggest that the correlation between wait time and the probability of accepting a gap is not linear (also see Figure 15) and very long wait times (e.g., > 9 seconds) tended to prompt drivers to accept a gap.
- **Is Gap.** The Is Gap feature was selected to be a splitting feature by multiple decision tree models. The decision tree structures illustrated that drivers were more likely to accept a gap than a lag (Figure 16, Figure 17, Figure 22, Figure 23 and Figure 24). This feature is often selected after Gap Size and Wait Time, suggesting that it is of less importance than those two features.
- **Major Road AADT.** The major road AADT was found to be important in decision tree models developed for all right-turning trips (Figure 17) and left-turning without TWLTL trips (Figure 20). Based on the decision tree for right-turning trips, drivers tended to reject a given gap at conditions where the major-road AADT was smaller and the gap size was smaller than 4.591 seconds. However, when drivers experienced a long gap (> 4.591 seconds), they were more likely to accept the gap if the major-road AADT was smaller. For left-turning trips with smaller major road AADT, the results showed that

drivers were more likely to accept a given gap when the gap size was larger than 5.815 seconds.

- **Major Road Sidewalk.** The decision tree model constructed for right-turn trips with leading vehicle (Figure 21) showed that when exposed to large gaps (> 5.488 seconds), drivers tended to reject a given gap at intersections where the major road had sidewalks. However, according to the decision tree constructed for left-turning trips with TWLTL (Figure 19), if drivers were exposed to a gap between 4.243 and 6.716 seconds, they were more likely to accept the gap if the major road had sidewalks.
- **Major Road Crosswalk – Left Approach.** According to the model constructed for left-turning trips without TWLTL on the major road (Figure 20), when given a small gap (i.e., < 5.815 seconds), drivers tended to accept the gap if there was a crosswalk on the left approach of the major road.
- **Annual Miles.** According to the model developed for right-turning trips with a leading vehicle (Figure 21), when the size of a given gap was smaller than 5.488 seconds, drivers who travelled 20,000 miles each year were more likely to accept the gap, while drivers who travelled less each year were more likely to reject it.

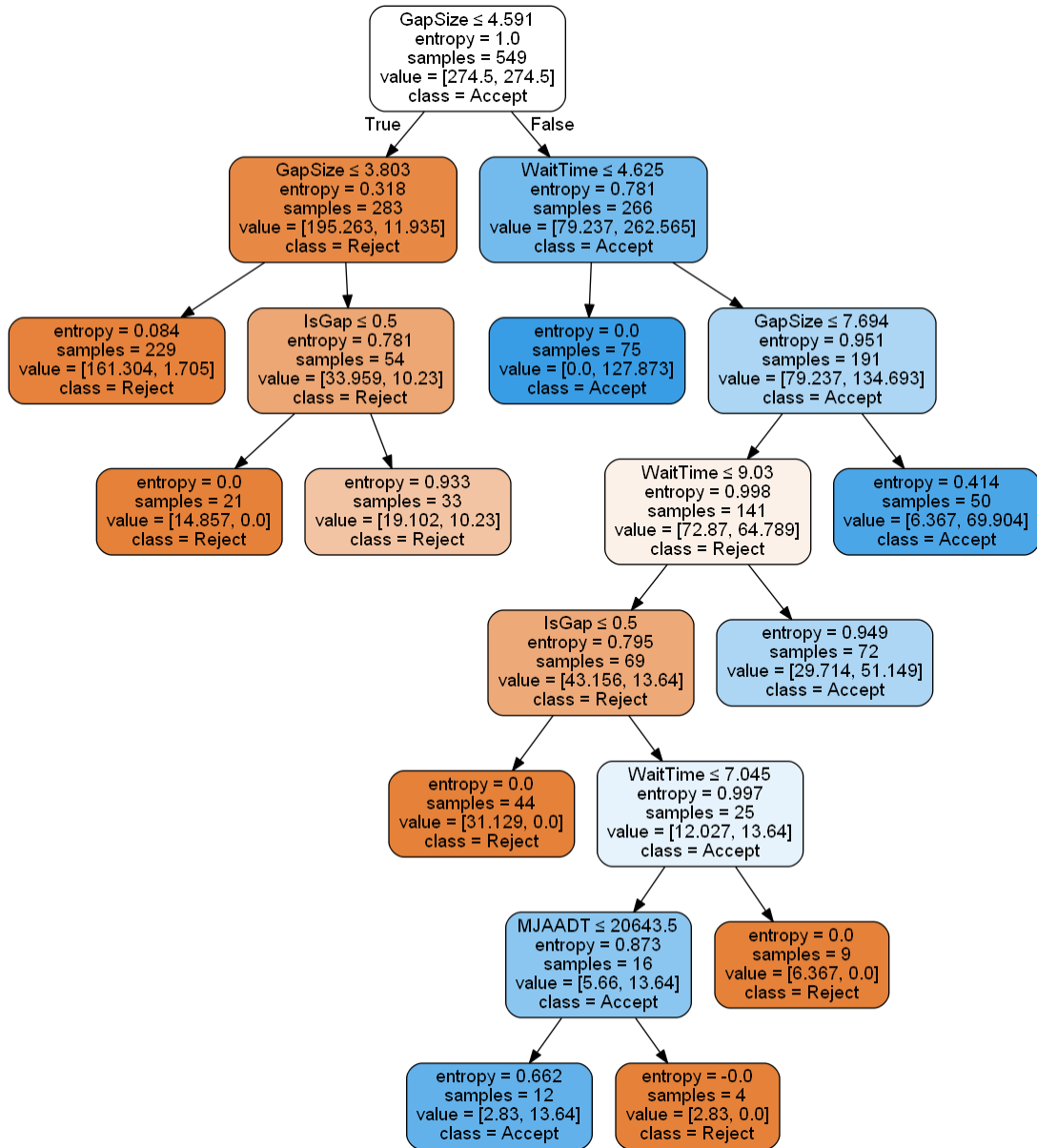


Figure 16. Flow chart. Decision tree for all 785 gaps.

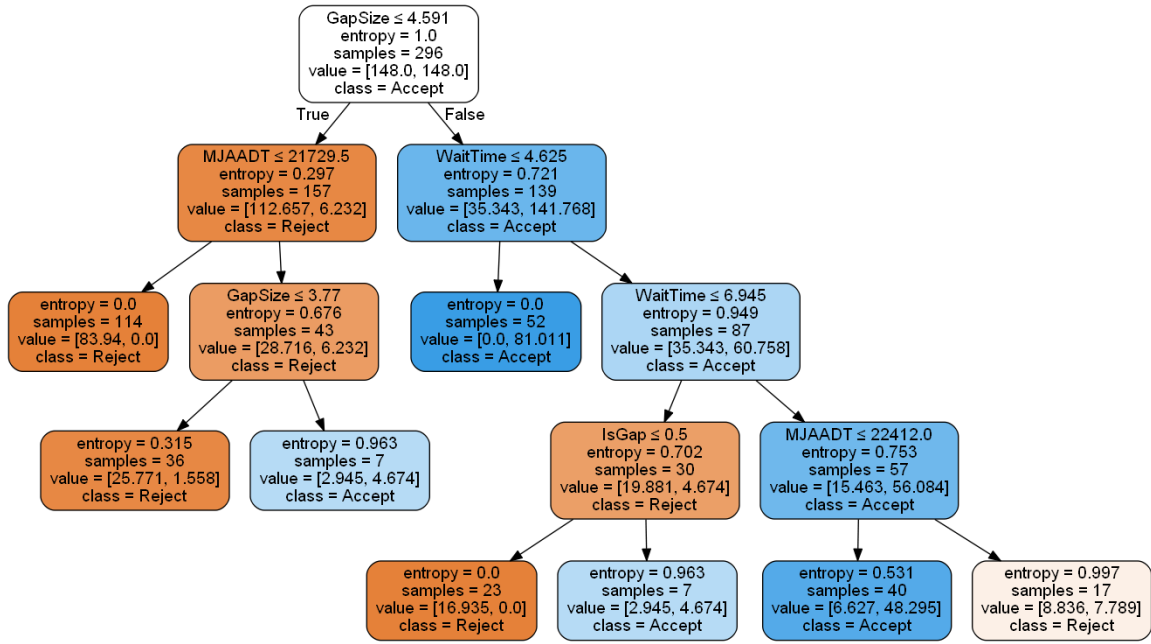


Figure 17. Flow chart. Decision tree for 423 gaps for right-turn trips.

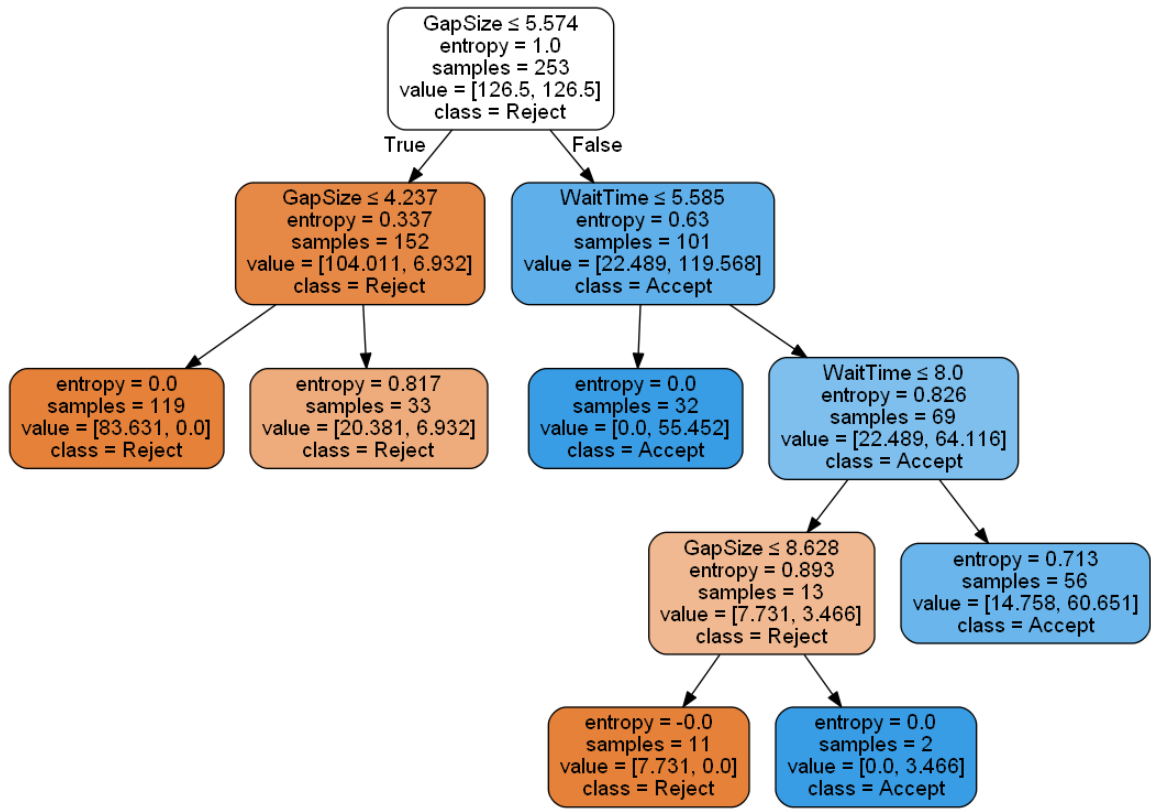


Figure 18. Flow chart. Decision tree for 362 gaps for left-turn trips.

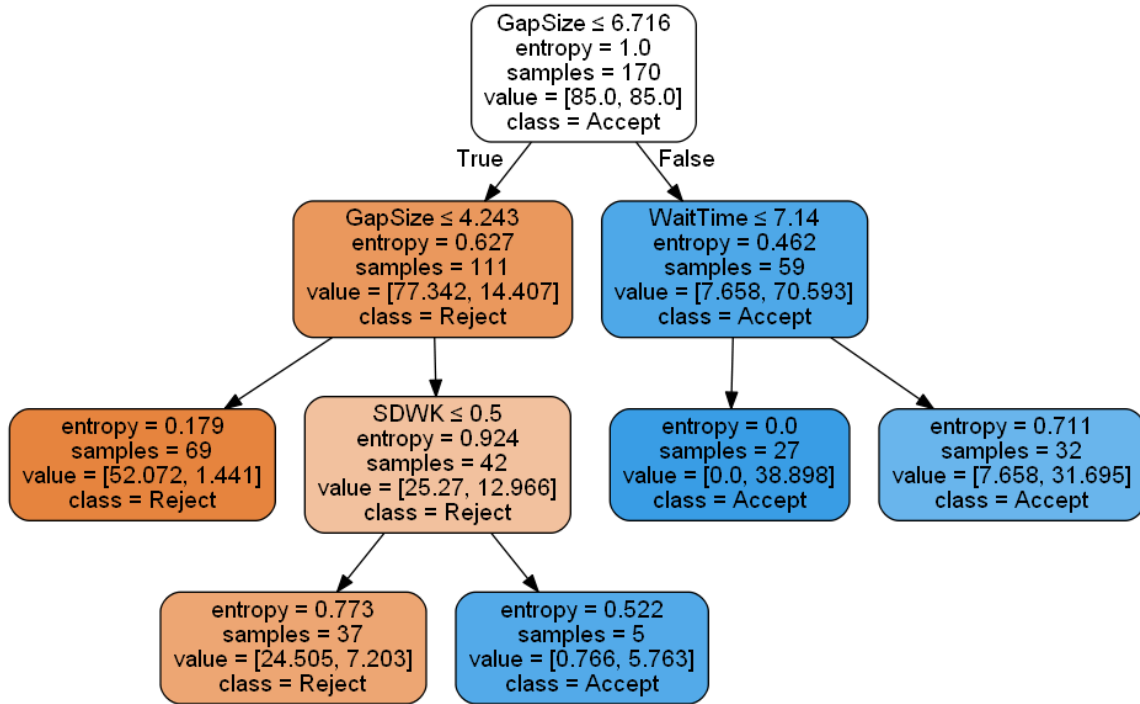


Figure 19. Flow chart. Decision tree for 243 gaps for left-turn trips with TWLTL.

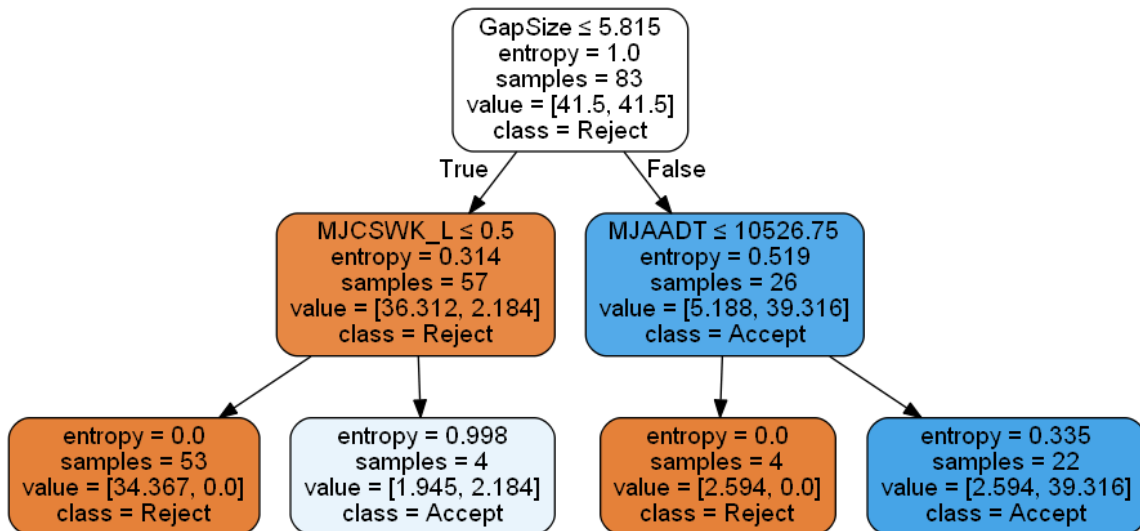


Figure 20. Flow chart. Decision tree for 119 gaps for left-turn trips without TWLTL.

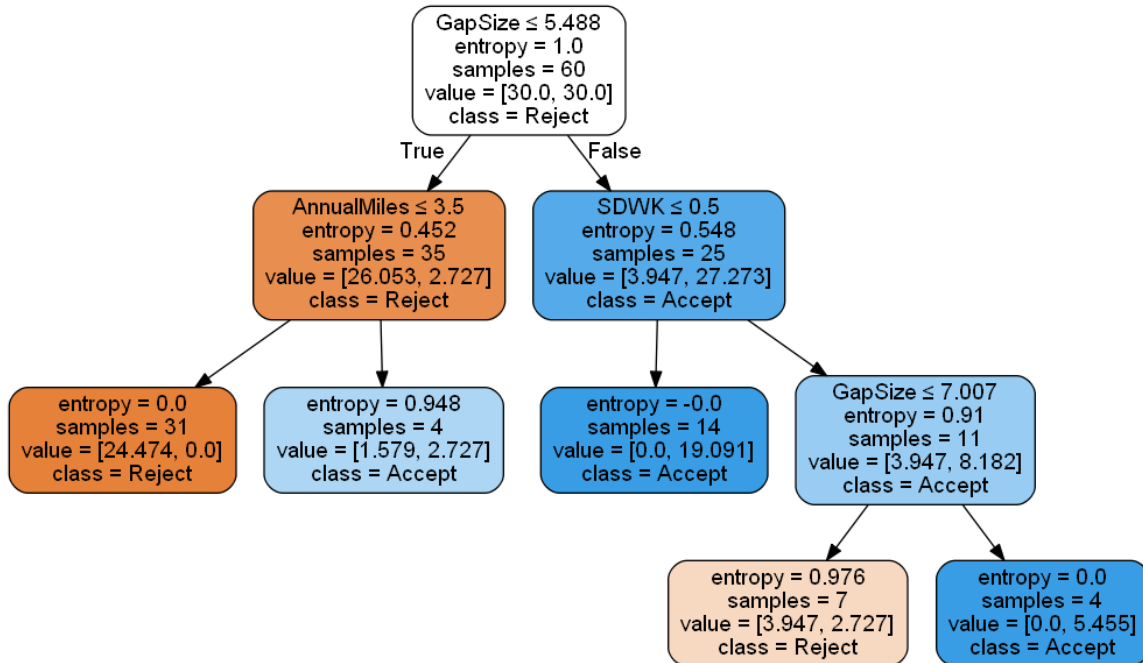


Figure 21. Flow chart. Decision tree for 87 gaps for right-turn trips with leading vehicle.

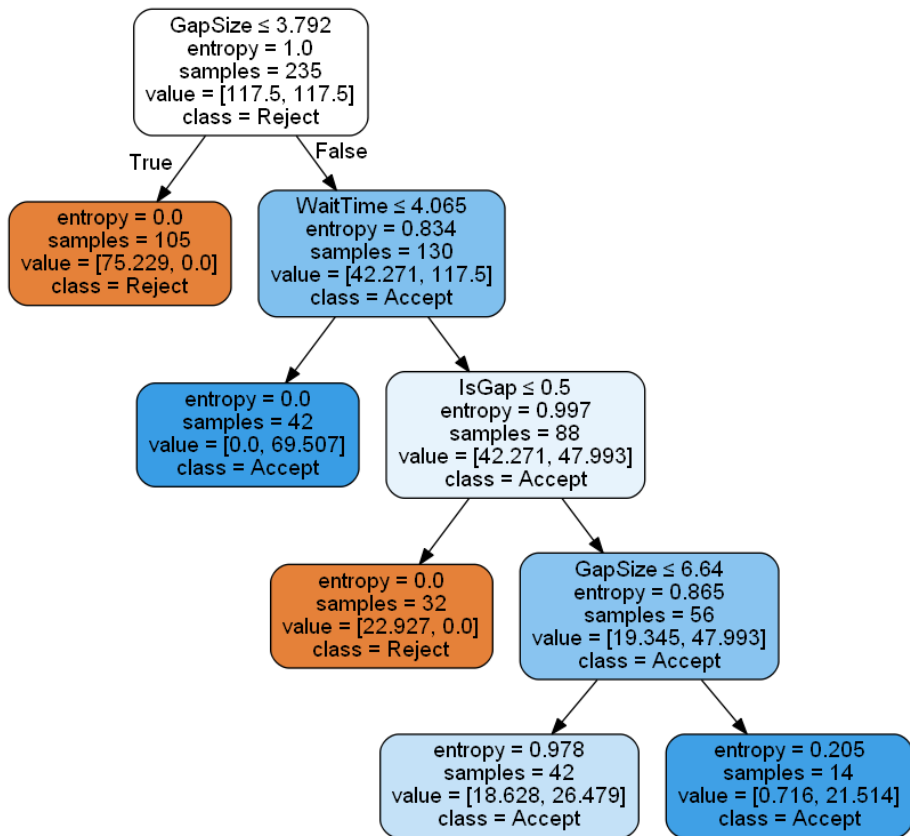


Figure 22. Flow chart. Decision tree for 336 gaps for right-turn trips without leading vehicle.

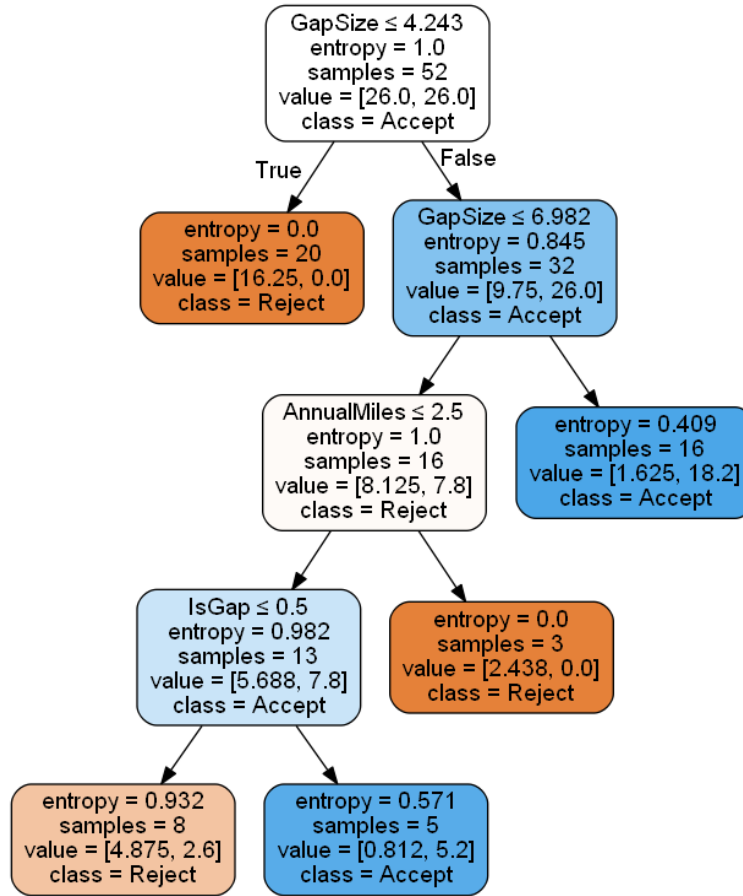


Figure 23. Flow chart. Decision tree for 75 gaps for left-turn trips with leading vehicle.

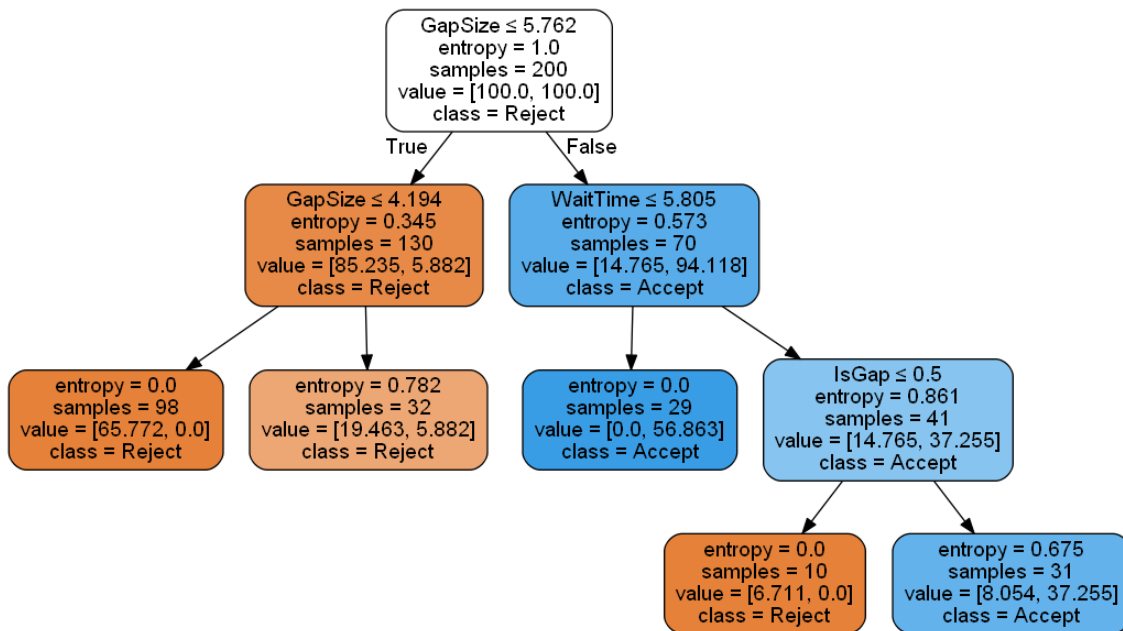


Figure 24. Flow chart. Decision tree for 287 gaps for left-turn trips without leading vehicle.

Random Forest Modeling Results

While decision trees are robust in identifying significant variables and critical variable values, they are prone to overfitting and results are sensitive to changes in data. To develop a thorough understanding of the variables affecting gap acceptance decisions, the project team also conducted a random forest machine learning analysis. During this analysis, the random forest models were developed for the same analysis scenarios using the Scikit-learn toolkit in Python.⁽⁶²⁾ Similarly, for each scenario, the dataset was split into 70% training and 30% testing. Observations of different groups were balanced by adjusting their associated weights. Entropy was also used as the criterion of selecting splitting features. The modeling information, prediction performance, important features (with a feature importance greater than 2%), and their associated feature importance scores are summarized in Table 24 (only features with importance scores greater than 2% are listed).

Table 24. Model information and important features of random forest models.

Model Information	Feature Importance
Model scenario: All 785 gaps Number of trees: 1000, maximum tree depth: 3 Accuracy in training set: 86.70% Accuracy in testing set: 86.86%	<ul style="list-style-type: none"> • Gap Size: 44.36% • Wait Time: 21.91% • Major Road AADT: 4.30% • Leading Vehicle: 3.31%
Model scenario: 423 gaps for right-turn trips Number of trees: 1000, maximum tree depth: 3 Accuracy in training set: 90.88% Accuracy in testing set: 82.68%	<ul style="list-style-type: none"> • Gap Size: 40.38% • Wait Time: 22.18% • Major Road AADT: 4.91% • Annual Miles: 2.36% • Intersection Alignment: 2.34% • Major Approach Left turn Lane – Right: 2.08%
Model scenario: 362 gaps for left-turn trips Number of trees: 1000, maximum tree depth: 4 Accuracy in training set: 90.91% Accuracy in testing set: 83.49%	<ul style="list-style-type: none"> • Gap Size: 48.15% • Wait Time: 16.38% • Major Road AADT: 4.59% • Annual Miles Traveled: 3.77% • Leading Vehicle: 2.43% • Is Gap: 2.01%
Model scenario: 243 gaps for left-turn trips with TWLTL Number of trees: 1000, maximum tree depth: 3. Accuracy in training set: 90.59% Accuracy in testing set: 84.93%	<ul style="list-style-type: none"> • Gap Size: 39.23% • Wait Time: 17.94% • Leading Vehicle: 5.17% • Annual Miles Traveled: 5.12% • Major Road AADT: 3.71% • Age Group: 2.78% • Intersection Alignment: 2.48% • Is Gap: 2.15%
Model scenario: 119 gaps for left-turn trips without TWLTL Number of trees: 100, maximum tree depth: 2 Accuracy in training set: 92.77% Accuracy in testing set: 80.56%	<ul style="list-style-type: none"> • Gap Size: 30.79% • Wait Time: 16.52% • Major Road AADT: 11.8% • Major Road Alignment: 10.79% • Intersection Alignment: 6.51% • Minor Approach Number of Lanes: 3.1% • Annual Miles Traveled: 2.95%

Model Information	Feature Importance
Model scenario: 87 gaps for right-turn trips with leading vehicle Number of trees: 100, maximum tree depth: 2 Accuracy in training set: 97.59% Accuracy in testing set: 83.33%	<ul style="list-style-type: none"> • Gap Size: 32.69% • Wait Time: 10.94% • Major Road AADT: 9.94% • Annual Miles Traveled: 7.2% • Major Road Alignment: 6.58% • Age Group: 5.50% • Intersection Alignment: 3.05% • Number of Violations: 2.28% • Major Road Sidewalk: 2.75%
Model scenario: 336 gaps for right-turn trips without leading vehicle Number of trees: 1000, maximum tree depth: 5 Accuracy in training set: 95.74% Accuracy in testing set: 91.09%	<ul style="list-style-type: none"> • Gap Size: 43.61% • Wait Time: 23.43% • Annual Miles Traveled: 3.46% • Major Road AADT: 3.4% • Is Gap: 3.1%
Model scenario: 75 gaps for left-turn trips with leading vehicle Number of trees: 1000, maximum tree depth: 4 Accuracy in training set: 95.32% Accuracy in testing set: 86.14%	<ul style="list-style-type: none"> • Gap Size: 42.38% • Wait Time: 23.73% • Major Road AADT: 3.84% • Is Gap: 2.65% • Major Approach Left Turn Lane-Right: 2.55%
Model scenario: 287 gaps for left-turn trips without leading vehicle Number of trees: 100, maximum tree depth: 5 Accuracy in training set: 93.00% Accuracy in testing set: 87.36%	<ul style="list-style-type: none"> • Gap Size: 51.55% • Wait Time: 14.62% • Major AADT: 4.65% • Intersection Alignment: 3.72% • Is Gap: 2.30%

During the model development process, the research team tried different combinations of tree numbers and tree depth for the random forest models. The model information included in Table 24 is for the tables that resulted in the highest prediction accuracy levels on the testing sets. In general, the random forest models had better prediction performance and were less susceptible to overfitting than the decision tree models even with datasets of smaller sample sizes. For each tree in the forest, the method only considered a random subset of the analyzed features, which therefore allowed the features with lower impact on the modeling results to be involved in the models. Due to this, the random forest analysis found a much larger number of features to be more important.

Again, the results suggested that Gap Size played a dominant role in each random forest model with a feature importance score between 30% and 50% depending on the analysis scenario (note that this importance is lower than the decision tree analysis suggested). Wait Time and Major Road AADT were also identified to be important by all analysis scenarios with feature importance scores between 10% and 25% and 3.5% and 12%, respectively. Annual Miles Traveled and Intersection Alignment were identified by multiple random forest models to be important as well. Other features such as Leading Vehicle, Major Approach Left Turn Lane – Right, and Major Road Alignment were found to be important for some scenarios as well. These findings in general are consistent with the results obtained during the logistic regression and decision tree analysis, although the feature Is Gap was not considered important by the random forest models. Note that unlike the logistic regression and decision tree models, random forest was unable to reveal the detailed influence of these features.

Comparison of Model Performance

Each of the three methods showed advantages as well as disadvantages during the modeling of drivers' gap acceptance behavior. A comparison of the three methods showed that the logistic regression method was straightforward in identifying significant variables and quantifying their impacts. However, the method was unable to explain complex correlations (e.g., the impact of wait time). The decision tree method, on the other hand, was able to reveal some complex correlations in the data and identified critical data points that served as thresholds in decision-making. The dichotomous structure associated with decision trees was able to show which decision was more likely to be taken under different conditions, but it could not quantify the influence as well as logistic regression did. Random forest models worked like a black box and did not provide information about the influence of the analyzed features' presence on responses, but they were able to identify a more complete set of variables affecting the output compared to the decision tree models. Although the logistic regression models quantified the impact of individual variables on the model outcome, they could not measure the importance of the variable for the overall performance of the model. The machine learning methods, therefore, were superior in this regard.

DRIVER BEHAVIOR AFTER ACCEPTING GAPS

Understanding the different behaviors of drivers merging into the major road traffic stream after accepting a gap can be valuable for various purposes. For example, such behaviors can be used in microscopic simulation to develop more realistic algorithms for simulating turning vehicles at intersections. In addition, such information may also help highly automated vehicle technology developers design automated vehicle behaviors at intersections that better mimic human driver behaviors. The information can also be used to create baselines for safety analysis relevant to turning behavior at intersections.

Using the collected naturalistic driving data, the project team conducted a comprehensive analysis of driver behaviors merging into the major road traffic stream after they accepted a gap. The analysis looked at two aspects of the driver behaviors: vehicle kinematics characteristics during turning and influential factors affecting the driver turning behavior in the form of vehicle kinematics.

Vehicle Kinematical Characteristics during Turning

During this analysis, the research team used Gaussian process regression to develop continuous vehicle kinematics profiles for the analysis scenarios to depict drivers' typical turning behavior after gap acceptance. Based on the data availability, the project team conducted this analysis for both longitudinal and lateral acceleration rates. The team initially attempted to develop speed profiles as well. However, the speed data of the analyzed trips exhibited large variances and Gaussian process regression did not show reliable results.

The vehicle kinematics profiles were developed for the turning movements only for each trip after accepting a gap into the major road. The start and end points of the turning portion of each trip were determined primarily based on vehicle speed, acceleration rates, and trip videos. The start point was determined as the point in time when the driver started accelerating longitudinally

and laterally from a complete stop after accepting a gap, or the vehicle started accelerating from a rolling stop after checking and accepting a gap. The end point of the turning movement was determined as the point in time when the vehicle stopped accelerating laterally after merging into the major road traffic flow. To normalize the vehicle kinematical behaviors, the team aggregated the time series data of each trip portion into 10 data points, which were then used as the basis for this analysis.

Overall Profiles for Left-turning and Right-turning Trips

Table 25 summarizes the basic statistics for the vehicle kinematical profiles of all analysis scenarios used for the Gaussian process regression analysis, followed by Figure 25 comparing the acceleration profiles between left-turning trips and the right-turning trips. Figure 26 and Figure 27 further show the detailed acceleration profiles for the overall left- and right-turning scenarios after drivers accepted a gap. In the detailed profile figures, the red dots represent the aggregated data points for acceleration rates of different trips. The blue curves are the vehicle kinematics profiles based on the Gaussian process regression results. The grey areas are the associated 95% confidence intervals for each data point.

Table 25. Basic statistics measures of gaussian process modeling profiles.

Scenario	No. of Trips	Longitudinal Acceleration (m/s ²)			Lateral Acceleration (m/s ²)		
		Max	Average	Std	Max	Average	Std
All left turn	202	2.024	1.248	0.364	2.479	1.661	0.722
Left turn with TWLTL	147	2.027	1.212	0.379	2.535	1.704	0.749
Left turn without TWLTL	55	2.029	1.345	0.343	2.334	1.545	0.651
Left turn with leading vehicle	54	1.851	1.178	0.297	2.480	1.676	0.743
Left turn without leading vehicle	148	2.087	1.275	0.389	2.479	1.655	0.715
Left turn for younger drivers	124	2.035	1.305	0.326	2.607	1.742	0.764
Left turn for mid-age drivers	27	2.068	1.339	0.353	2.267	1.520	0.618
Left turn for older drivers	51	1.990	1.062	0.475	2.389	1.540	0.683
All right turn	234	1.780	1.532	0.170	2.395	1.474	0.698
Right turn with leading vehicle	55	1.863	1.624	0.170	2.583	1.573	0.747
Right turn without leading vehicle	179	1.756	1.503	0.171	2.343	1.444	0.684
Right turn for younger drivers	99	1.838	1.562	0.191	2.578	1.564	0.755
Right turn for mid-age drivers	52	1.702	1.517	0.168	2.047	1.293	0.606
Right turn for older drivers	83	1.775	1.504	0.159	2.408	1.479	0.690

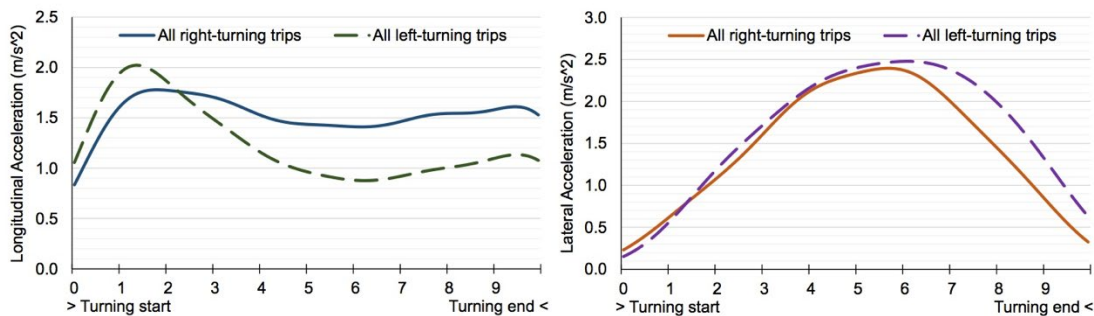
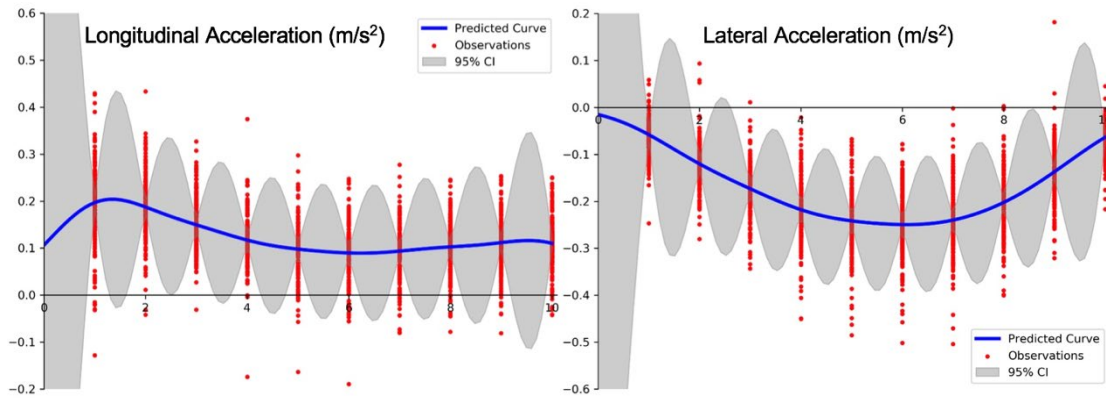


Figure 25. Graphs. Comparison of acceleration profiles between left- and right-turning trips.



Note: X-axis is the continuous timeline from the beginning of the turn to the end of the turning movement.

Figure 26. Graphs. Gaussian process regression profiles for all left-turn trips.

The illustrations suggest that:

- The maximum longitudinal acceleration for left-turning vehicles was approximately 2.0 m/s^2 (with an average acceleration rate of 1.25 m/s^2), higher than the average maximum longitudinal acceleration rate (i.e., 1.8 m/s^2 with an average rate of 1.53 m/s^2). Overall, both left- and right-turning vehicles accelerated quickly after they accepted a gap, and then reduced to a lower but prolonged acceleration rate while turning to reach a desired speed. The longitudinal acceleration curves suggest that left-turning vehicles accelerated more abruptly compared to right-turning vehicles to reach a higher speed, and then accelerated continuously at a lower rate compared to right-turning vehicles. These behavioral differences also seemed to suggest that right-turning drivers accelerated at a higher level of comfort/ease compared to left-turning drivers when merging into the major road traffic stream.
- The maximum lateral acceleration rate for left-turning vehicles was approximately 2.5 m/s^2 (with an average rate of 1.66 m/s^2), slightly higher than that for right-turning vehicles (i.e., 2.4 m/s^2 with an average rate of 1.47 m/s^2). Although both lateral acceleration profiles resembled a parabolic curve, the peak value for the left-turning profile was reached later in the turning process than for the right-turning vehicles. This difference suggests that left-turning drivers tend to reach the opposite side of the major road first and then turn quickly into the travel direction/lane.

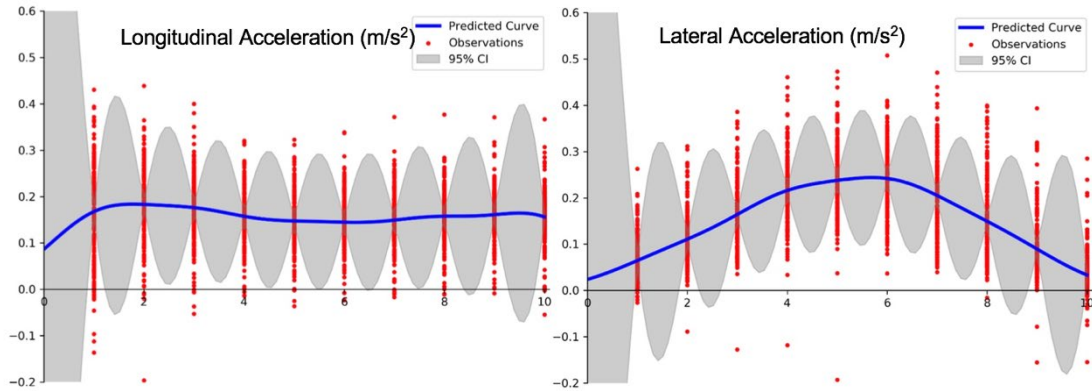


Figure 27. Graphs. Gaussian process regression profiles for all right-turning trips.

Acceleration Profiles for Left-Turning Scenarios

Figure 28 compares the longitudinal and lateral acceleration profiles for left-turning vehicles with and without a TWLTL, followed by the detailed Gaussian process regression results and original data in Figure 29 and Figure 30. The comparison suggests that the presence of a TWLTL particularly affected the longitudinal accelerating behavior of left-turning drivers. Compared to the acceleration profile for the left turn without TWLTL scenario, drivers clearly accelerated less smoothly after they entered the intersection box.

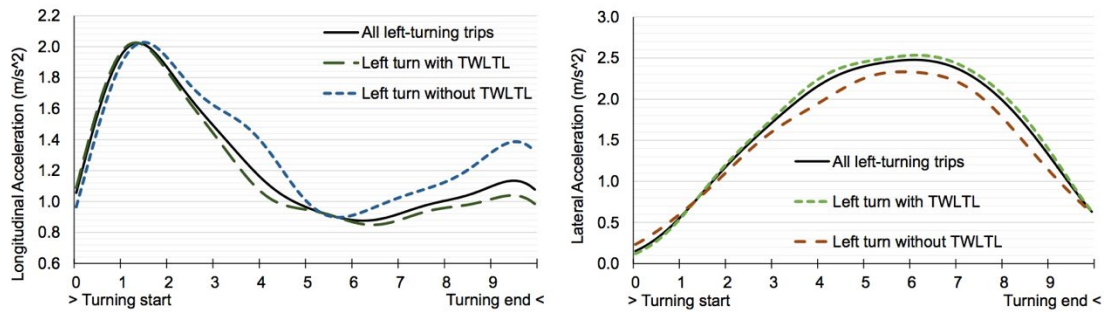


Figure 28. Graphs. Acceleration profiles for left-turning trips (with versus without TWLTL).

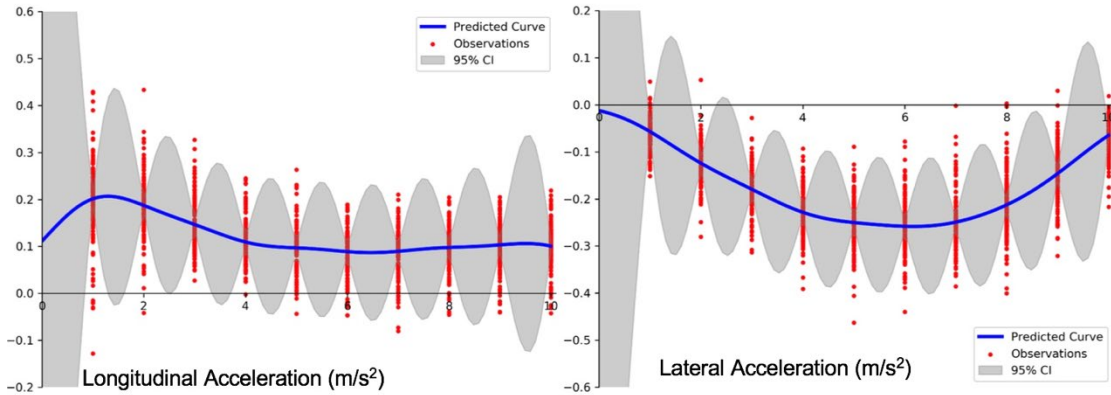


Figure 29. Graphs. Gaussian process regression profiles for left turn with TWLTL.

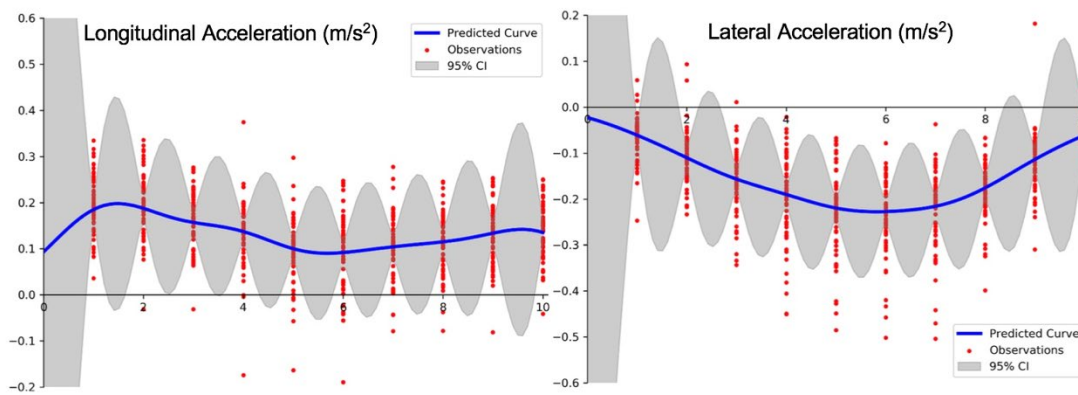


Figure 30. Graphs. Gaussian process regression profiles for left turn without TWLTL.

Figure 31 compares the acceleration profiles for left-turning trips with and without leading/following vehicles, followed by Figure 32 and Figure 33 showing the original data points and Gaussian process regression confidence intervals for the curves individually. As previously noted, the presence of leading vehicles frequently occurred with the presence of following vehicles. Based on the illustrations, the presence of leading and/or following vehicles did not appear to have any effect on left-turning drivers' lateral acceleration behavior, but did affect their longitudinal acceleration behavior. In general, left-turning vehicles with leading and/or following vehicles accelerated more slowly at the beginning (indicated by the lower peak) and maintained a flatter acceleration profile later during turning (indicated by the higher minimum).

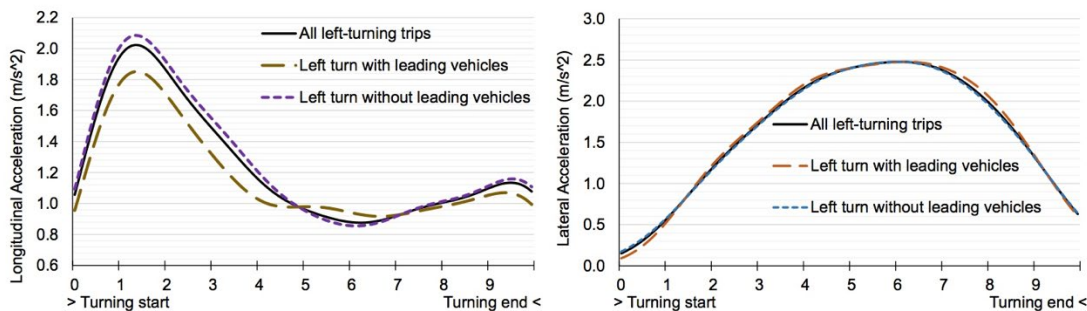


Figure 31. Graphs. Acceleration profiles for left-turning trips (with versus without leading vehicles).

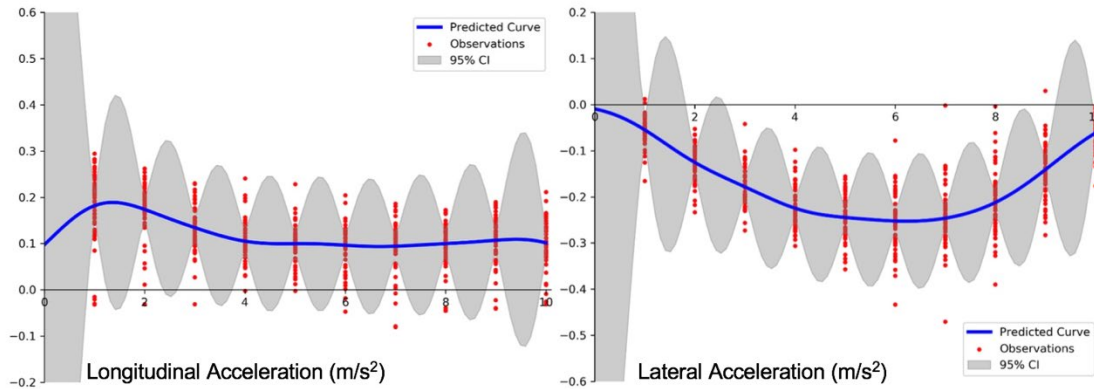


Figure 32. Graphs. Gaussian process regression profiles for left turn with leading vehicles.

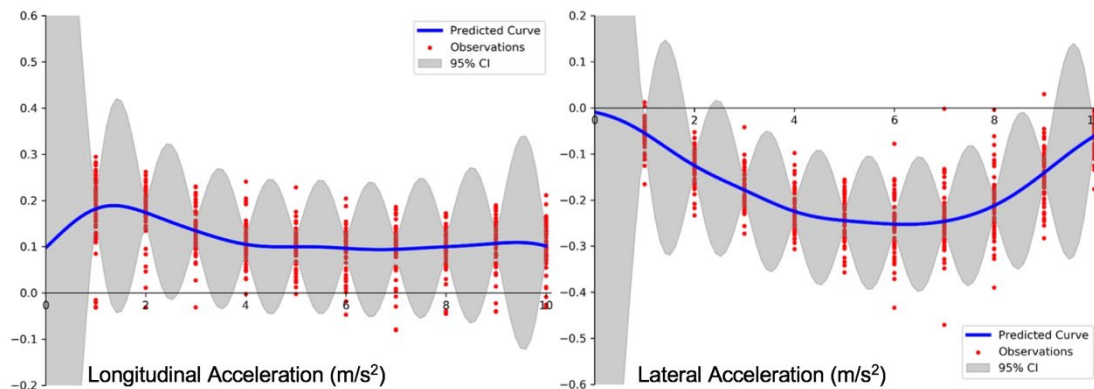


Figure 33. Graphs. Gaussian process regression profiles for left turn without leading vehicles.

Figure 34 compares the acceleration profiles among different driver groups, followed by individual profiles along with the associated data and confidence intervals in Figure 35–Figure 37. The comparison seems to suggest that drivers of different age groups accelerated differently when turning left after accepting a gap. In terms of longitudinal acceleration, older drivers (i.e., drivers ≥ 65 years) exhibited a pause (a drop in longitudinal acceleration rate) at the sixth data point during the turn movement, and then gradually increased acceleration rate. Overall, older drivers had an average longitudinal acceleration rate of 1.06 m/s^2 compared with the 1.3 m/s^2 average rate for other age groups (see Table 25). Younger drivers (i.e., drivers < 25 years) initially followed the same acceleration pattern as the other age groups but maintained a higher acceleration rate later, particularly compared to older drivers.

In terms of lateral acceleration, younger drivers turned faster than the other driver groups, following a smoother parabola. Older drivers, however, reached the peak when turning left until the seventh data point, and then quickly reduced the lateral acceleration rate after that. The shape of the lateral acceleration peak for older drivers was therefore also sharper than that for the other age groups. The mid-age (25–64 years) group turned more slowly compared to younger drivers and followed an equally smooth but lower parabola of lateral acceleration rates during the turn (these drivers may have taken a larger turning curve by turning into a further lane).

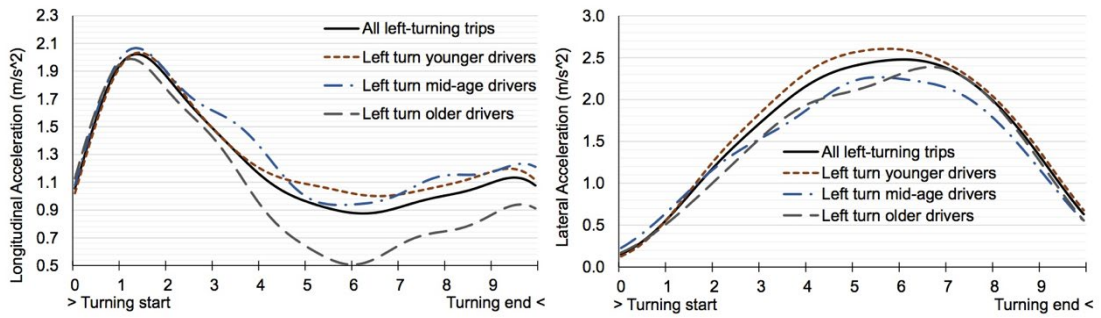


Figure 34. Graphs. Acceleration profiles for left-turning trips (by age group).

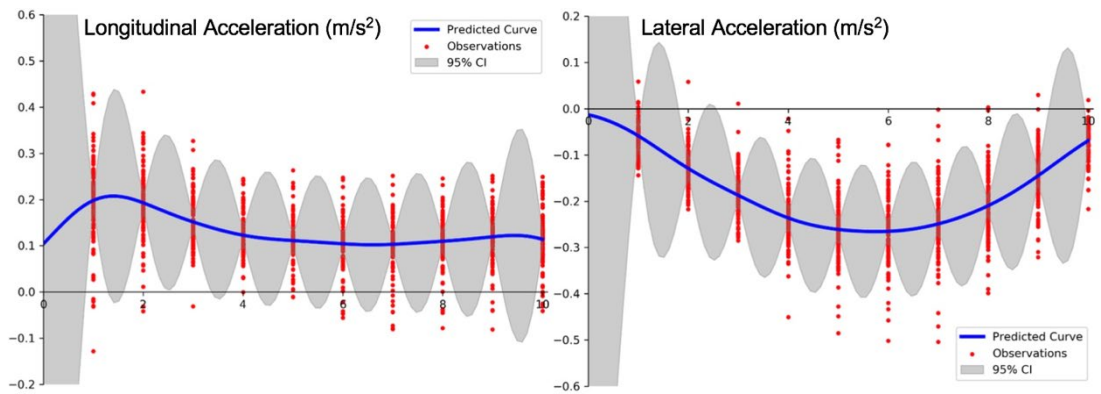


Figure 35. Graphs. Gaussian process regression profiles for left-turn trips for young drivers.

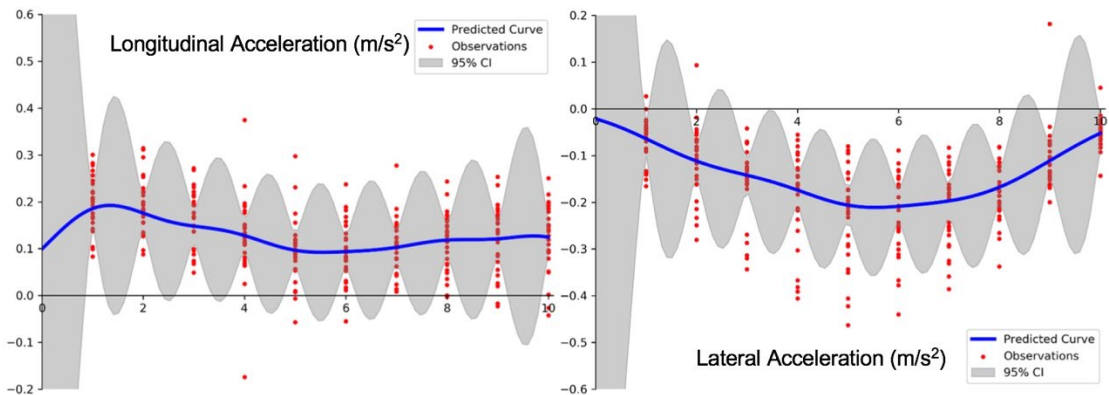


Figure 36. Graphs. Gaussian process regression profiles for left-turn trips for mid-age drivers.

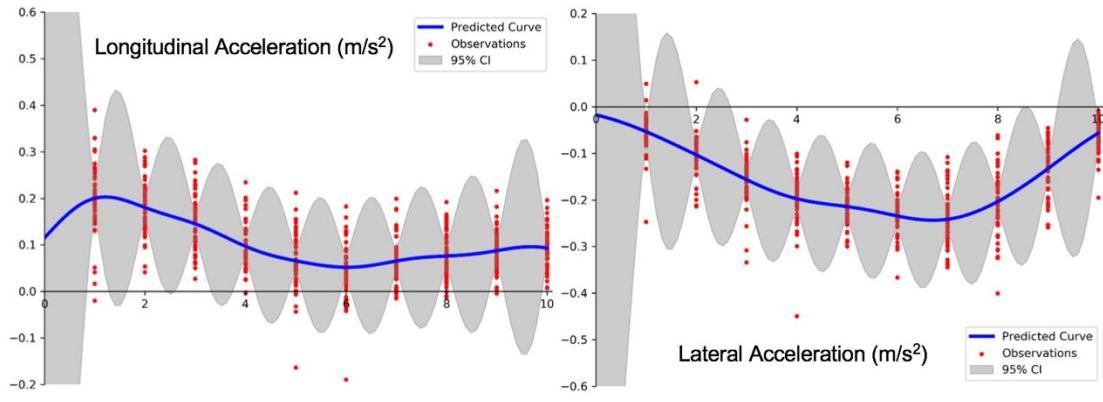


Figure 37. Graphs. Gaussian process regression profiles for left-turn trips for older drivers.

Acceleration Profiles for Right-Turning Scenarios

Figure 38 compares the longitudinal and lateral acceleration rates for two right-turn scenarios: with leading/following vehicles and without leading/following vehicles. Figure 39 and Figure 40 provide more details of the curves individually by displaying the original data points and the confidence intervals generated by the Gaussian process regression analysis. The comparison suggests that drivers accelerated faster when turning right after an accepted gap if there were leading and/or following vehicles. In addition, the lateral acceleration curve seems to suggest that they also turned faster in the with-leading-vehicle scenario. These observations suggest that right-turning drivers tended to turn more decisively and/or swiftly either due to a higher level of confidence/ease following another vehicle and/or the psychological pressure of the following vehicles.

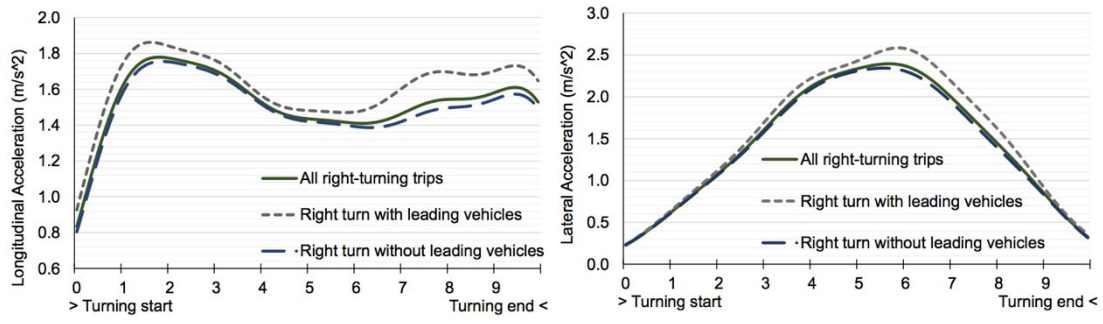


Figure 38. Graphs. Acceleration profiles for right turn (with versus without leading vehicles).

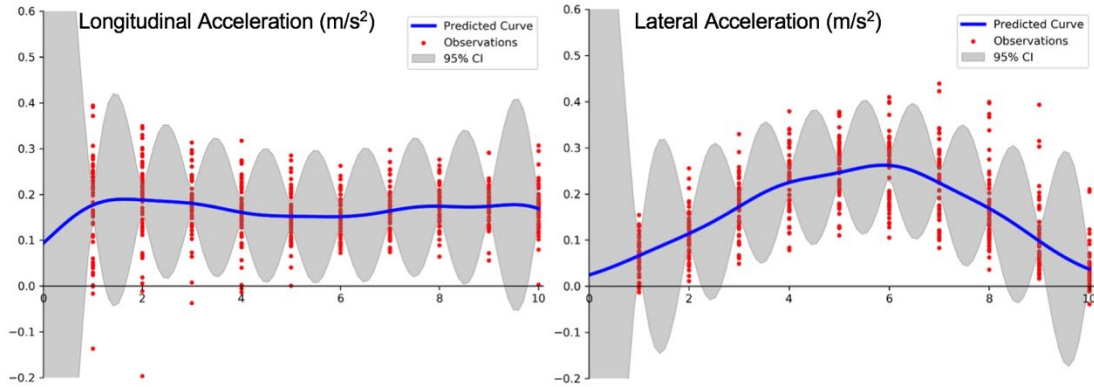


Figure 39. Graphs. Gaussian process regression profiles for right-turn trips with leading vehicle.

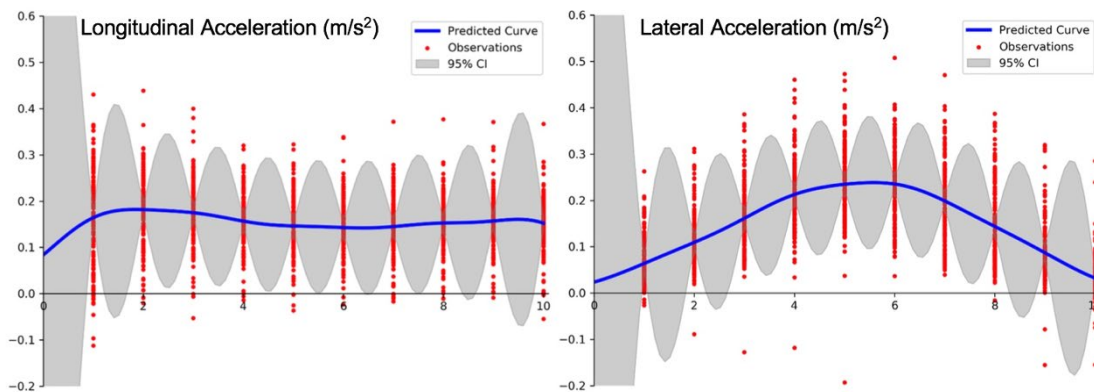


Figure 40. Graphs. Gaussian process regression profiles for right-turn trips without leading vehicle.

Figure 41 compares the right-turn acceleration profiles by age groups, followed by Figure 42–Figure 44, which provide more details on the curves individually. By comparison, drivers younger than 25 both accelerated and turned faster when turning right after accepting a gap. Drivers from 25 to 64, on the other hand, tended to turn and accelerate more slowly and smoothly.

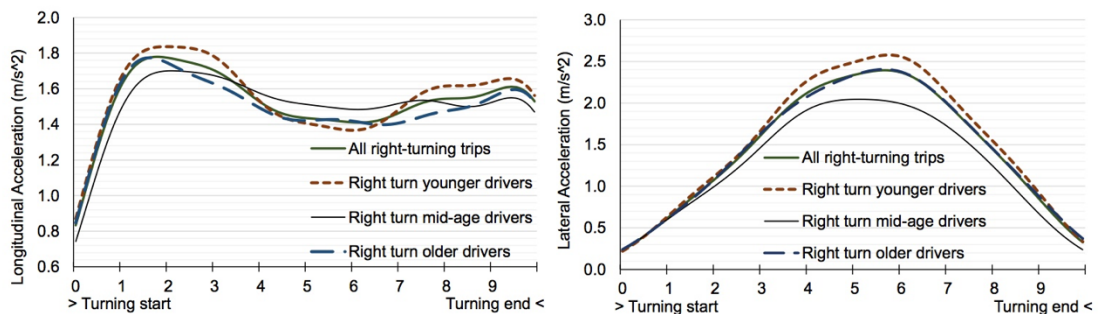


Figure 41. Graphs. Acceleration profiles for right turn (with versus without leading vehicles).

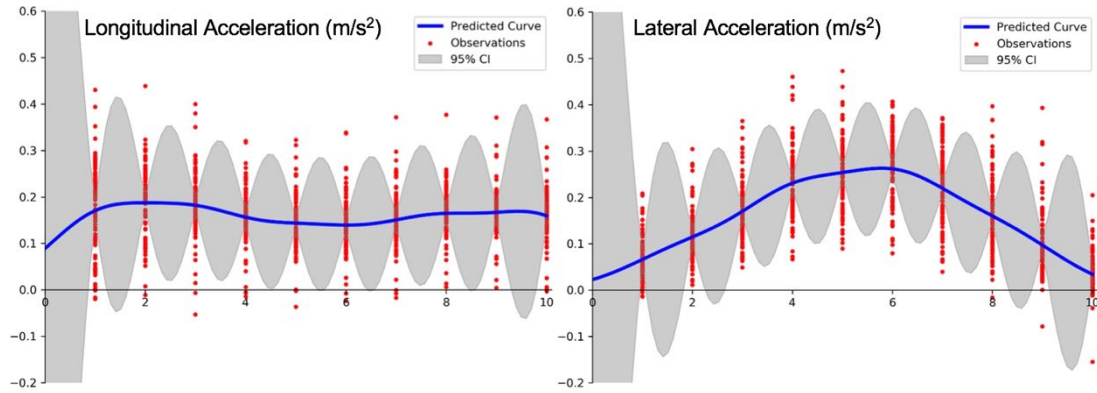


Figure 42. Graphs. Gaussian process regression profiles for right-turn trips for younger drivers.

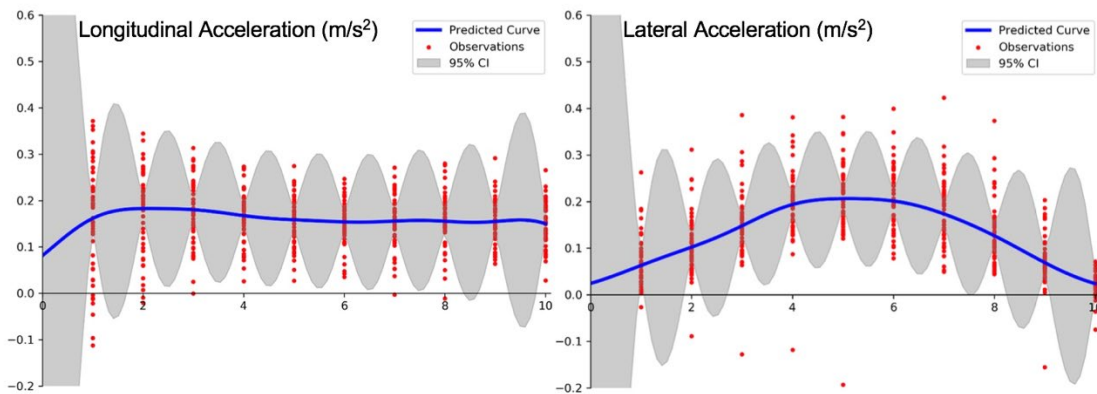


Figure 43. Graphs. Gaussian process regression profiles for right-turn trips for mid-age drivers.

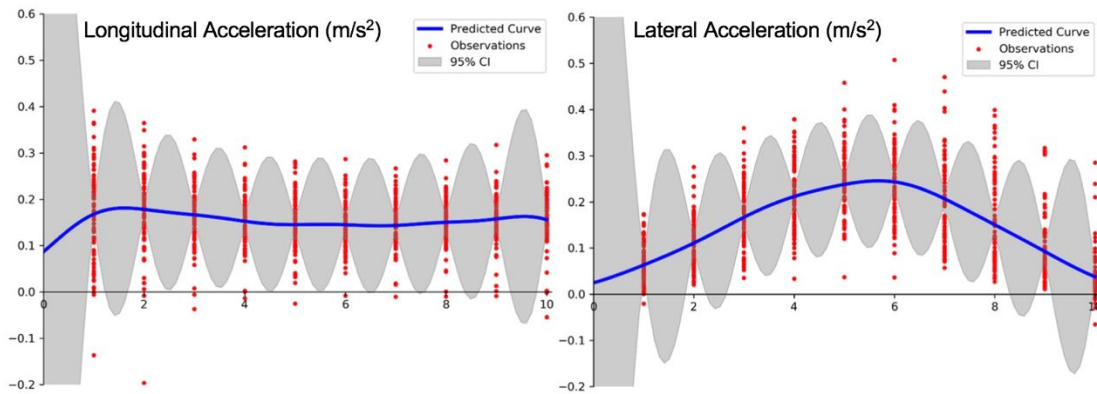


Figure 44. Graphs. Gaussian process regression profiles for right-turn trips for older drivers.

Factors Affecting Vehicle Kinematics After Accepting Gaps

To further understand how different factors affect drivers' turning behavior after accepting a gap, the research team conducted a multiple linear regression analysis of the key vehicle kinematical

measures, including average speed, maximum longitudinal acceleration rate, standard deviation of longitudinal acceleration rate, maximum lateral acceleration rate, and standard deviation of lateral acceleration rate. During this analysis, the project team used the 252 trips that involved accepted gaps. All trips involved left- or right-turning maneuvers. The significant variables were selected at a 0.05 level of significance.

Table 26 summarizes the significant correlations affecting drivers' turning behaviors after they accepted a gap, followed by Table 27–Table 31, which list detailed linear models for the aforementioned dependent variables. The analysis showed that the following factors affected driver behavior when turning into the major road traffic stream after accepting a gap:

- Trip-related factors affecting driver speed and acceleration when turning:
 - **Accepted gap size.** Accepted gap sizes affected both longitudinal and lateral accelerating behaviors. When drivers accepted a shorter gap size, they tended to accelerate and turn faster. Meanwhile, shorter accepted gap sizes also correlated with lower acceleration variances.
 - **Gap versus lag.** For the same gap size, the results showed that a driver would travel more slowly and turn more slowly after accepting a gap instead of a lag.
 - **Turn type.** Compared to right-turn maneuvers, left-turning trips were associated with higher variance of longitudinal acceleration rate and higher lateral acceleration rate.

- Driver-related factors affecting turning behaviors:
 - **Age group.** When analyzing the data by age group, the results suggested that younger drivers (< 25 years) in general had a slower speed while turning after they accepted a gap but had both a higher maximum lateral acceleration rate and lateral acceleration variance. Older drivers (≥ 65 years) also had a higher lateral acceleration variance when turning after they accepted a gap. Mid-age drivers (between 25 and 64 years), however, were found to have a lower maximum lateral acceleration rate when turning.
 - **Annual miles traveled during the previous 3 years.** The results showed that those who drove more often tended to turn at a higher speed and had higher lateral acceleration (maximum and variance) after they accepted a gap.
 - **Maximum crash severity during the previous 3 years.** Similarly, the results showed that drivers who previously had more severe crashes tended to turn at a higher speed and accelerate faster after they accepted a gap.

- Roadway-related factors affecting driver speed and acceleration when turning:
 - **Major road AADT.** The results showed that drivers tended to travel more slowly when turning onto roadways with a higher AADT. However, the results also showed that such drivers had a higher acceleration variance when turning, suggesting effects from major street traffic.
 - **Major road number of through lanes.** The results showed that when turning into roadways with more through lanes, drivers tended to travel faster and turn faster (i.e., higher maximum lateral acceleration rate). However, the results also

showed that drivers turn more smoothly, as shown by the lower lateral acceleration variance.

- **Minor road crosswalk.** Perhaps contrary to expectations, the results showed that the presence of a crosswalk on the minor road approach correlated with higher turning speeds and higher maximum lateral acceleration rates after gap acceptance.
- **On-street parking.** The results showed that on-street parking on both the major and minor streets correlated with lower lateral acceleration (both maximum rate and variance), suggesting that drivers turned more carefully when there was on-street parking.
- **Roadway lighting.** Based on the modeling results, drivers turned more smoothly when the major street was lighted, as evidenced by lower lateral and longitudinal acceleration variance.
- **Other roadway factors affecting drivers' turning behavior.** In addition to the aforementioned factors, major road turn lanes, major road sidewalks, and intersection skew also had effects on drivers' turning behavior.

Table 26. Summary of correlation directions for factors affecting turning behavior.

Variable	Average Speed	Longitudinal Acceleration		Lateral Acceleration	
		Maximum	Variance	Maximum	Variance
Intercept	15.51	0.267	0.092	0.259	0.101
Trip-related factors					
Accepted gap size	-	-0.005	-0.002	-0.004	-0.002
Is gap (compared to a lag)	-2.49	0.022	-	-	-0.006
Is left turn (compared to right turn)	-	-	0.012	0.034	-
Driver-related factors					
Age group-younger	-3.34	-	-	0.030	0.023
Age group-mid age	-	-	-	-0.054	-
Age group-older	-	-	-	-	0.013
Annual miles traveled (previous three years)	2.35	-	-	0.021	0.008
Maximum crash severity (previous three years)	2.11	0.020	-	-	-
Roadway-related factors					
Major road AADT	-0.22	-	0.001	-	-
Major road number of through lanes	3.17	-	-	0.034	-0.011
Minor road crosswalk	3.49	-	-	0.036	-
Major road on street parking-right side	-	-	-	-0.078	-0.027
Minor road on street parking-left side	-	-	-	-0.100	-0.021
Roadway lighting	-	-	-0.043	-	-0.019
Minor road on street parking-right side	-4.04	-	-	-	-
Major road right turn lane-left approach	-	-	-	-0.051	-
Major road sidewalk	-	-	-	-	0.009
Major road TWLTL	3.64	-	-	-	-
Intersection skew-no	-1.95	-	-	-	-
Intersection skew-towards right	-	0.034	-	-	-
Intersection skew-towards left	-	-	-0.010	-	-

Table 27. Linear regression model for average speed.

Variable	Parameter Estimate	Standard Error	F Value	Pr > F
Intercept	15.505	1.448	114.7	<.001
Is Gap	-2.492	0.588	18.0	<.001
Major Road TWLTL	3.640	0.633	33.1	<.001
Annual Miles	2.351	0.348	45.6	<.001
Max Crash Severity	2.110	0.575	13.5	<.001
Major Road AADT	-0.215	0.070	9.4	0.002
Major Road No. of Through Lanes	3.172	0.749	17.9	<.001
Minor Approach Crosswalk	3.494	1.235	8.0	0.005
Minor Approach Parking-Right	-4.043	1.130	12.8	0.000
Intersection Alignment-Skew to Right	-1.954	0.792	6.1	0.014
Age Group-Younger	-3.339	0.606	30.4	<.001

Model $Pr > F \leq .0001$; R-Square = 0.408; and Adjusted R-Square = 0.384.

Table 28. Linear regression model for maximum longitudinal acceleration.

Variable	Parameter Estimate	Standard Error	F Value	Pr > F
Intercept	0.267	0.015	303.8	<.001
Accepted Gap Size	-0.005	0.001	10.3	0.002
Is Gap	0.022	0.009	6.3	0.013
Max Crash Severity	0.020	0.008	7.3	0.007
Intersection Alignment-Skew to Right	0.034	0.014	5.6	0.019

Model $Pr > F \leq .0001$; R-Square = 0.145; and Adjusted R-Square = 0.131.

Table 29. Linear regression model for standard error of longitudinal acceleration.

Variable	Parameter Estimate	Standard Error	F Value	Pr > F
Intercept	0.092	0.011	72.9	<.001
Accepted Gap Size	-0.002	0.000	11.6	0.001
Left Turn	0.012	0.003	14.4	<.001
Major Road AADT	0.001	0.000	11.7	0.001
Intersection Alignment-Skew to Left	-0.010	0.005	4.3	0.039
Lighting	-0.043	0.009	20.6	<.001

Model $Pr > F \leq .0001$; R-Square = 0.177; and Adjusted R-Square = 0.161.

Table 30. Linear regression model for maximum lateral acceleration.

Variable	Parameter Estimate	Standard Error	F Value	Pr > F
Intercept	0.259	0.023	129.3	<.001
Accepted Gap Size	-0.004	0.001	8.2	0.005
Left Turn	0.034	0.011	10.3	0.002
Annual Miles	0.021	0.005	16.3	<.001
Major Road No. of Through Lanes	0.034	0.010	10.7	0.001
Major Approach Right Turn Lane-Left	-0.051	0.011	21.3	<.001
Minor Approach Crosswalk	0.036	0.015	5.7	0.017
Major Road On-Street Parking-Right Side	-0.078	0.037	4.4	0.038
Minor Road On-Street Parking-Left Side	-0.100	0.022	20.7	<.001
Age Group-Younger	0.030	0.010	9.3	0.003
Age Group-Mid Age	-0.054	0.013	16.8	<.001

Model Pr > F ≤ .0001; R-Square = 0.276; and Adjusted R-Square = 0.246.

Table 31. Linear regression model for standard error of lateral acceleration.

Variable	Parameter Estimate	Standard Error	F Value	Pr > F
Intercept	0.101	0.009	115.1	<.001
Accepted Gap Size	-0.002	0.000	24.7	<.001
Is Gap	-0.006	0.003	4.7	0.031
Annual Miles	0.008	0.002	21.3	<.001
Major Approach Left Turn Lane-Left	-0.011	0.003	13.4	<.001
Major Road Side Walk	0.009	0.003	8.0	0.005
Major Approach On-Street Parking-Right Side	-0.027	0.011	5.5	0.020
Minor Road On-Street Parking-Left	-0.021	0.005	18.2	<.001
Lighting	-0.019	0.008	5.3	0.022
Age Group-Younger	0.023	0.004	32.9	<.001
Age Group-Older	0.013	0.004	10.3	0.002

Model Pr > F ≤ .0001; R-Square = 0.310; and Adjusted R-Square = 0.281.

CHAPTER 4. SUMMARY AND CONCLUSION

Safety at unsignalized intersections continues to be a major concern for transportation agencies and roadway users. To improve intersection safety, this project conducted a comprehensive study of gap acceptance behaviors at unsignalized intersections using the SHRP 2 NDS data. The project team conducted a number of analyses to achieve the research objectives. First, the team identified the critical gaps for a number of common scenarios using three widely accepted methods: binary logistic regression, maximum likelihood, and probability equilibrium. The team then went on to develop a complete understanding of the factors affecting gap acceptance decisions using logistic regression and machine learning techniques. Finally, researchers further investigated drivers' longitudinal and lateral acceleration behaviors during turning after accepting a gap and factors affecting their turning behaviors.

For the purpose of this project, the team collected 1,170 accepted and rejected gaps/lags based on 466 NDS trips at 60 unsignalized T-intersections in the states of Washington and North Carolina. To collect complete information about the involved trips, drivers, and intersections, the project team utilized a number of data sources, including time series data measuring vehicle kinematics for the analyzed trips, forward-facing and rear-view videos for the analyzed trips, driver demographic and driving history data, the RID database, and satellite images.

The following summarizes the major findings of this project:

- **Critical gap determination.** The study showed an overall critical gap of 5.3 seconds for right-turning trips and 6.2 seconds for left-turning trips (Table 11). For right-turning vehicles, the presence of leading and/or following vehicles resulted in a 0.2-second longer critical gap (i.e., 5.4 versus 5.2 seconds). For left-turning vehicles, the presence of leading and/or following vehicles corresponded with a 0.4-second shorter critical gap (i.e., 5.9 versus 6.3 seconds). In addition, the presence of a TWLTL also resulted in a shorter critical gap for left-turning vehicles by 0.2 seconds (i.e., 6.1 versus 6.3 seconds). Both the left-turn and right-turn critical gap sizes determined in this study were lower than those included in the HCM (Table 1) and AASHTO's Green Book (Table 2) by approximately 1 second.

When comparing the three different methods used for critical gap determination, in most cases, the maximum likelihood method produced lower critical gap sizes. In particular, the critical gap sizes for the right-turn scenarios estimated by this method were shorter than the other two methods. On the other hand, both the binary logistic regression and the probability equilibrium methods showed comparable results for all scenarios. Both the logistic regression and probability equilibrium methods modeled the critical gaps using all collected gaps, while the maximum likelihood method used the accepted and maximum rejected gap pairs of the studied trips. In this respect, the former made better use of the collected data than the latter. However, the maximum likelihood is exempted from the biases caused by small rejected gaps. Note that when taking the standard deviation of the critical gaps into consideration, none of these differences was considered significant.

- **Factors affecting gap acceptance decisions.** This analysis used three methods (logistic regression, decision tree, and random forest) to identify a comprehensive set of factors that affect gap acceptance decisions at intersections. The results showed a large number of factors that could have affected drivers' gap acceptance decisions. For example, being a gap instead of a lag, presence of leading and/or following vehicles, higher AADT, intersection being unskewed, and increased number of through lanes all increased the probability of a driver accepting a given gap for multiple scenarios. On the other hand, adverse weather conditions reduced the probability of acceptance for a given gap. In addition, the modeling results suggested that a waiting time of longer than 9 seconds tended to increase the probability of gap acceptance.

The decision tree models further suggested that gap sizes were the determinant factor for gap acceptance. Gap sizes were responsible for 60%–85% of the model weights depending on the analysis scenario. In addition to gap size, the analysis found that wait time, gap type (i.e., gap versus lag), and major road AADT had relatively heavy impacts on the model performance. The random forest analysis once again showed that gap size had the dominant effect on gap acceptance decision, albeit at a lower level of importance (30%–51% of model weight) compared to the decision tree models.

Each of the three methods showed advantages as well as disadvantages. The logistic regression method was straightforward in identifying significant variables and quantifying their impacts. However, the method was unable to explain nonlinear correlations. The decision tree method, on the other hand, was able to reveal some complex correlations in the data and identified critical data points that served as thresholds in decision-making. The dichotomous structure associated with decision trees was able to show which decision was more likely to be taken under different conditions, but could not quantify the influence as logistic regression did. Random forest models worked like a black box and did not provide information about the influence of the analyzed features' presence on responses. However, they were able to identify a more complete set of variables that affect the output compared to the decision tree models. Although the logistic regression models quantified the impact of individual variables on the model outcome, they could not measure the importance of the variables for the overall performance of the model. Machine learning methods, therefore, are superior in this regard.

- **Driver turning behaviors and impact factors.** During this project, the research team used Gaussian process regression to develop continuous vehicle kinematics profiles for different turning behaviors, and multiple regression to identify factors affecting how drivers turn after accepting a gap. The maximum longitudinal acceleration for left-turning vehicles was approximately 2.0 m/s^2 (with an average acceleration rate of 1.25 m/s^2), higher than the average maximum longitudinal acceleration rate (i.e., 1.8 m/s^2 with an average rate of 1.53 m/s^2). Overall, both left- and right-turning vehicles initially accelerated quickly after they accepted a gap, and then reduced to a lower but prolonged acceleration rate while turning to reach a desired speed.

The maximum lateral acceleration rate for left-turning vehicles was approximately 2.5 m/s^2 (with an average rate of 1.66 m/s^2), also slightly higher than that for right-turning vehicles (i.e., 2.4 m/s^2 with an average rate of 1.47 m/s^2). Although both lateral acceleration profiles resembled a parabolic curve, the peak value for the left-turning profile was reached later in the turning process than that for the right-turning vehicles.

The results suggest that drivers accelerated faster when turning right but slower when turning left if leading and/or following vehicles were present. In addition, drivers younger than 25 generally turned faster in terms of both longitudinal and lateral acceleration rates.

In addition to the driver age and presence of a leading/following vehicle, a number of other factors affected how drivers turned after they accepted a gap. For example, gap size, gap type, annual miles traveled, and past crash experience all correlated with how drivers turned. Several roadway-related factors, such as major road AADT, number of through lanes, on-street parking, and roadway lighting affected how aggressively and smoothly drivers turned to merge into the major road traffic stream after they accepted a gap.

This information can be valuable for microscopic simulation algorithms and tools. The constructed vehicle kinematical profiles can also serve as the typical/baseline turning scenario for safety analyses of driver behaviors at intersections. In addition, such information may be used for automated vehicle technology development to allow more consistent vehicle behavior at intersections. Vehicle behaviors that are more consistent with human driver expectations will undoubtedly improve safety in a traffic environment with vehicles of mixed levels of automation and non-motorized vehicles.

Traditionally, gap acceptance has largely been studied using observational data collected from outside of vehicles based on video or manual timing at intersections. By using observational data, existing research has studied driver gap judgements from an external point of view, without taking into account the role of driver behavior. In addition, most previous studies have been based on a limited number of intersections. Depending on local traffic and driver characteristics, studies frequently found significantly different critical gap values at different intersections. The availability of the SHRP 2 NDS data has provided a unique opportunity to study driver gap acceptance behavior from a different angle. Naturalistic driving data can help provide a better understanding of this problem by allowing the study of driver behavior in an unobtrusive way. The detailed driver, trip, and vehicle kinematical data made possible by use of the NDS database was fully utilized in this study to provide new insights to the gap acceptance behaviors at unsignalized intersections.

REFERENCES

1. National Highway Traffic Safety Administration. (2014). *Fatality Analysis Reporting System (FARS)*. <https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars>
2. Federal Highway Administration. (2020). *Safety - Unsignalized intersections*. <https://safety.fhwa.dot.gov/intersection/conventional/unsignalized/>
3. Transportation Research Board. (2000). *Highway capacity manual*. National Research Council.
4. Harwood, D. W., Mason, J. M., Brydia, R. E., Pietrucha, M. T., & Gittings, G. L. (1996). *Intersection sight distance* (No. Project 15-14 [1]).
5. Troutbeck, R. J., & Brilon, W. (2017). *Revised monograph on traffic flow theory*. Federal Highway Administration. <https://www.fhwa.dot.gov/publications/research/operations/tft/>
6. Hamed, M. M., Easa, S. M., & Batayneh, R. R. (1997). Disaggregate gap-acceptance model for unsignalized T-intersections. *Journal of Transportation Engineering*, 123(1), 36-42.
7. *Access management manual*. (2003). Transportation Research Board.
8. Pant, P. D., & Balakrishnan, P. (1994). Neural network for gap acceptance at stop-controlled intersections. *Journal of Transportation Engineering*, 120(3), 432-446.
9. Polus, A. (1983). Gap acceptance characteristics at unsignalised urban intersections. *Traffic Engineering and Control*, 24(5), 255-258.
10. Yan, X., Radwan, E., & Guo, D. (2007). Effects of major-road vehicle speed and driver age and gender on left-turn gap acceptance. *Accident Analysis & Prevention*, 39(4), 843-852.
11. Ragland, D. R., Arroyo, S., Shladover, S. E., Misener, J. A., & Chan, C. Y. (2006). Gap acceptance for vehicles turning left across on-coming traffic: Implications for intersection decision support design (Paper 06-2696). *Transportation Research Board 85th Annual Meeting Compendium of Papers*.
12. Al-Mhairat, E. W. (2016). Gap acceptance regression model at un-signalized intersections (major-minor roads) in Amman-Jordan. *International Journal of Civil Engineering Research*, 7(1), 1-13.
13. Nabae, S. (2011). *An evaluation of gap acceptance behavior at unsignalized intersections*. [M.S. Thesis, Oregon State University].

14. Beanland, V., Lenné, M. G., Candappa, N., & Corben, B. (2013). Gap acceptance at stop-controlled T-intersections in a simulated rural environment. *Transportation Research Part F: Traffic Psychology and Behaviour*, 20, 80-89.
15. Kyte, M., Tian, Z., Mir, Z., Hameedmansoor, Z., Kittelson, W., Vandehey, M., & Troutbeck, R. (1996). *Capacity and level of service at unsignalized intersections* (NCHRP Project 3-46 final report. Transportation Research Board.
16. Tupper, S. M. (2011). *Safety and operational assessment of gap acceptance through large-scale field evaluation*. [M.S. Thesis, University of Massachusetts Amherst].
17. Karthika, P. T., & Koshi, B. I. (2014). Gap acceptance behavior of drivers at t-intersections. *International Journal of Engineering Research & Technology*, 3(11), 935-938.
18. Dissanayake, S., Lu, J. J., & Ping, Y. I. (2002). Driver age differences in day and night gap acceptance capabilities. *IATSS Research*, 26(1), 71-79.
19. Lerner, N. D., Huey, R. W., McGee, H. W., & Sullivan, A. (1995). *Older driver perception-reaction time for intersection sight distance and object detection. Volume I* (No. FHWA-RD-93-168). Chicago.
20. Akçelik, R. (2007, December). A review of gap-acceptance capacity models. In *The 29th Conference of Australian Institutes of Transport Research (CAITR 2007)*, University of South Australia, Adelaide (pp. 5-7).
21. Ragland, D. R., Arroyo, S., Shladover, S. E., Misener, J. A., & Chan, C. (2005, December). *Gap acceptance for vehicles turning left across on-coming traffic: implications for intersection decision support design*. Research Report, Safe Transportation Research & Education Center, UC Berkeley.
22. Teply, S., Abou-Henaidy, M. I., & Hunt, J. D. (1997). Gap acceptance behaviour - Aggregate and logit perspectives: Part 1. *Traffic Engineering & Control*, 38(09), 474-482.
23. Shaaban, K., & Hamad, H. (2018). Group gap acceptance: A new method to analyze driver behavior and estimate the critical gap at multilane roundabouts. *Journal of Advanced Transportation*. <https://doi.org/10.1155/2018/1350679>
24. Maurya, A. K., Amin, H. J., & Kumar, A. (2016). Estimation of critical gap for through movement at four leg uncontrolled intersection. *Transportation Research Procedia*, 17, 203-212.
25. Paschalidis, E., Choudhury, C. F., & Hess, S. (2018). Modelling the effects of stress on gap-acceptance decisions combining data from driving simulator and physiological

- sensors. *Transportation Research Part F: Traffic Psychology and Behaviour*, 59, 418-435.
26. Mitsopoulos-Rubens, E., Triggs, T., & Regan, M. (2009). Comparing the gap acceptance and turn time patterns of novice with experienced drivers for turns across traffic. In *Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*.
 27. Gattis, J., & Low, S. (1999). Gap acceptance at atypical stop-controlled intersections. *Journal of Transportation Engineering*, 125(3), 201-207.
 28. Hewitt, R. H. (1983). Measuring critical gap. *Transportation Science*, 17(1), 87-109.
 29. Zohdy, I. H. (2009). *Modeling permissive left-turn gap acceptance behavior at signalized intersections*. [Doctoral dissertation, Virginia Tech].
 30. Petzoldt, T., Hoang, Q., & Bogda, K. (2017). Time to arrival estimates, (pedestrian) gap acceptance and the size arrival effect. In *Proceedings of the 9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, Manchester Village, Vermont.
 31. Fitzpatrick, K. (1991). Gaps accepted at stop-controlled intersections. *Transportation Research Record*, 1303(11), 103-112.
 32. Abhigna, D., Brahmanekar, D. P., & Ravishankar, K. V. R. (2020). Multi vehicle-type right turning gap-acceptance and capacity analysis at uncontrolled urban intersections. *Periodica Polytechnica Transportation Engineering*, 48(2), 99-108.
 33. Pollatschek, M. A., Polus, A., & Livneh, M. (2002). A decision model for gap acceptance and capacity at intersections. *Transportation Research Part B: Methodological*, 36(7), 649-663.
 34. Lord-Attivor, R., & Jha, M. K. (2012, January). Modeling gap acceptance and driver behavior at stop controlled (priority) intersections in developing countries. In *Proceedings of the 6th WSEAS International Conference on Computer Engineering and Applications*, World Scientific and Engineering Academy and Society (pp. 29-38).
 35. Nagalla, R., Rothuganti, P., & Pawar, D.S. (2017, May 16-19). Analyzing gap acceptance behavior at unsignalized intersections using support vector machines, decision tree and random forests. In *Proceedings of the 8th International Conference on Ambient Systems, Networks and Technologies*, Madeira, Portugal.
 36. Pawar, D. S., Patil, G. R., Chandrasekharan, A., & Upadhyaya, S. (2015). Classification of gaps at uncontrolled intersections and midblock crossings using support vector machines. *Transportation Research Record*, 2515(1), 26-33.

37. Pant, P. D., & Balakrishnan, P. (1994). Neural network for gap acceptance at stop-controlled intersections. *Journal of Transportation Engineering*, 120(3), 432-446.
38. Catchpole, E. A., & Plank, A.W. (1986), The capacity of a priority intersection. *Transportation Research Part B: Methodological*, 20(6), 441-456.
39. Brilon, W., Koenig, R., & Troutbeck, R. J. (1999). Useful estimation procedures for critical gaps. *Transportation Research Part A: Policy and Practice*, 33(3-4), 161-186, 1999.
40. Mohan, M., & Chandra, S. (2016). Review and assessment of techniques for estimating critical gap at two-way stop-controlled intersections. *European Transport-trasporti Europei*, 61(61).
41. Maze, T. H. (1981). A probabilistic model of gap acceptance behavior. *Transportation Research Record*, 795, 8-13.
42. Ashworth, R. (1969). The analysis and interpretation of gap acceptance data. *The First Annual Conference of the British Universities Transport Study Group*, University of Leeds.
43. Guo, R. (2010). Estimating critical gap of roundabouts by different methods. In *Proceedings of the 6th Advanced Forum on Transportation of China* (pp. 84-89).
44. Troutbeck, R. J. (1992). *Estimating the critical acceptance gap from traffic movements* (Research report 92-5), Queensland University of Technology.
45. McGowen, P., & Stanley, L. (2012). Alternative methodology for determining gap acceptance for two-way stop-controlled intersections. *Journal of Transportation Engineering*, 138(5), 2012.
46. Wu, N. (2006, July). A new model for estimating critical gap and its distribution at unsignalized intersections based on the equilibrium of probabilities. In *Proceeding of the 5th International Symposium on Highway Capacity and Quality of Service* (pp. 1-10).
47. Zhong, X., Zhu, X., Zhang, Y., & Liu, X. (2007). Left-turn gap acceptance behavior of tee type of unsignalized intersection. In *International Conference on Transportation Engineering 2007* (pp. 2975-2980).
48. Miller, A. J. (1971, June). Nine estimators for gap-acceptance parameters. In *Proceedings of the International Symposium on the Theory of Traffic Flow and Transportation*. Berkeley, California.
49. American Association of State Highway and Transportation Officials. (2010). *A policy on geometric design of highways and streets* (The Green Book).

50. Sahraei, M. A., & Puan, O. C. (2014, March.). Determination of gap acceptance at priority intersections. In *Proceedings of the 8th SEATUC Symposium*, Vol. 4.
51. Dutta, M., & Ahmed, M. A. (2018). Gap acceptance behavior of drivers at uncontrolled T-intersections under mixed traffic conditions. *Journal of Modern Transportation*, 26(2), 119-132.
52. Hankey, J. M, Perez, M. A., & McClafferty, J. A. (2016). *Description of the SHRP 2 naturalistic database and the crash, near-crash, and baseline data sets*. Task report for the Second Strategic Highway Research Program. Virginia Tech Transportation Institute.
53. Campbell, K. (2012). The SHRP 2 Naturalistic Driving Study: Addressing driver performance and behavior in traffic safety. *TR News*, 282, 30-35.
54. Transportation Research Board of the National Academy of Sciences. (2013). The 2nd Strategic Highway Research Program Naturalistic Driving Study Dataset. Available from the SHRP 2 NDS InSight Data Dissemination web site: <https://insight.shrp2nds.us>
55. McLaughlin, S. B., & Hankey, J. M. (2015). *Naturalistic driving study: Linking the study data to the roadway information database* (No. SHRP 2 Report S2-S31-RW-3). Virginia Tech Transportation Institute.
56. Troutbeck, R. J. (2014). Estimating the mean critical gap. *Transportation Research Record*, 2461(1), 76-84.
57. *SAS/STAT 9.2 User's Guide*, Second Edition. SAS Institute Inc. https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_reg_sect038.htm
58. *Decision Tree Learning*. Wikipedia.org. https://en.wikipedia.org/wiki/Decision_tree_learning. Accessed 06/12/2020
59. *Random Forest*. Wikipedia.org. http://en.wikipedia.org/wiki/Random_forest. Accessed 06/12/2020.
60. *Gaussian Process*. (n.d.) In Wikipedia. https://en.wikipedia.org/wiki/Gaussian_process. Accessed 06/12/2020.
61. *Gaussian Process Regression Models*. MathWorks, Inc. <https://www.mathworks.com/help/stats/gaussian-process-regression-models.html>. Accessed 06/12/2020.
62. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, 12, 2825-2830.