Gap Acceptance Behavior in the Lane-Changing Model in Congested Traffic

Nick D. Sangwa

Southeast University

On-Campus Supervisor: Professor Wusheng Hu

Off-Campus Supervisor: Professor Lu Sun

Table of Contents

Abstract

Including heterogeneity in the driver populace can significantly improve the accuracy of microscopic traffic simulation models. Yet, compared to car-following models, heterogeneity in lane-changing models has not been ventured into enough. To integrate the overlooked heterogeneity into lane-changing models, this research will construct a Multinomial logit model and use AIC & BIC to regulate them. One dataset was mined from the NGSIM vehicle trajectory dataset. It is expected that there will be at least one recognized ideal model and they will be split evenly between multinomial logit models. Additional influencing factors may be statistically noteworthy in the multinomial logit models; this will indicate if the suggested model will be effective in data mining for extracting unseen relations. While various variables will be measured across the board, the class will be named matching to its common aspects. The multinomial models will be tested for results in precision & accuracy. The results of this study are crucial for possibly bettering traffic simulation, traffic safety and procedure, increasing our knowledge on microscopic traffic flow, and traffic operations & management. The key work of this study is to construct a multinomial logit models based on the extracted NGSIM trajectory data, examine the results of any produced models, and determine the characterization of the lane-changing behavior with integrated driver heterogeneity. This is done through the Alteryx Designer and SPSS software. The main contribution of this study show a defined class observed from the results of the multinomial logit model. The conclusion determines that the results can be used for different aspects of traffic solutions.

Keywords: *Lane-changing, Gap Acceptance, NGSIM, Heterogeneity, Multinomial Logit Model*

Acceptance Behavior in the Lane-Changing Model in Congested Traffic

1.0 Introduction

The original lane-changing model covered numerous urban driving states in which traffic signs, transit lanes, obstructions, and the existence of heavy operating vehicles affect drivers' lane selection. It was primarily proposed for microscopic traffic simulation tools (Gipps, 1986). Since its inception, the topic of lane changing has been majorly studied through microsimulation. Microscopic traffic simulation packages deliver a virtual environment to assess innovative traffic management strategies and analyze their results. Microscopic traffic simulation packages are proficient in analyzing traffic behavior in various lane configurations, traffic alignments and states of traffic flow. Microsimulation as a category of computerized analytical tools can perform highly intricate analyses of microscopic activities. They are used in the application of a variety of transportation and traffic studies. They have been used to mimic the behavior of individual vehicles within a predefined road system and are used to forecast the probable impact of fluctuations in traffic patterns resulting from alterations to the physical environment or traffic flow. It is essential to increase their accuracy in modelling drivers' decision making due to the growing reliance on them. One of the most important parts of a microscopic traffic simulation software is the lane-changing model and consequently, it is crucial to ensure that the lane-changing behavior of motorists are precisely displayed in these simulations. Creating a precise lane-changing model for motorists is now considered an essential component of traffic model development. Fluctuations on roadways appear as an effect of lane changing instead of car following, for the periods of heavy traffic conditions (Mauch and Cassidy 2002; Laval and Daganzo 2006). Therefore, understanding the influences which effect drivers' lane-changing behaviors, in addition to developing the capabilities to model these decisions, have a significant part to play in the advancement of traffic management strategies. The advancement of these strategies are imperative to the improvement of traffic safety, traffic flow conditions, and automation in the driving industry. These are the larger concerns. The aspect of these concerns to be answered is how we improve lane-changing behavior models to more accurately represent the heterogeneity of the drivers on the road today. We have to take a look into the lane changing processes first, then figure what exactly is a major challenge that has halted further understanding of these processes.

Lane changes are separated into two categories: mandatory lane changes and discretionary lane changes. Drivers, to stay on their route, perform mandatory lane changes. Merging and diverging are two types of mandatory lane changes describing when motorists have to move from an on-ramp to the roadway or from the roadway to an off-ramp. One of the most central effects of lane changing on traffic flow conditions could be the traffic flow fluctuations caused by them. Through heavy traffic conditions, the fluctuations emerge as an effect of lane changing instead of car following (Mauch and Cassidy 2002; Laval and Daganzo 2006). The numerous lane changing actions of merging and diverging can create bottleneck areas on the roadway resulting in a flow breakdown under heavy traffic density (Cassidy and Bertini 1999; Daganzo et al. 1999; Hoogendoorn and Bovy 2001; Daganzo 2002; Wall and Hounsell 2005). Therefore, it is imperative to construct accurate merging and diverging behavior models in microscopic traffic simulation to replicate real-life traffic conditions.

As of lately, the capability of microscopic traffic simulation models to embody reality has brought worries because of the lack of knowledge on driver heterogeneity (Ossen et al. 2006; Hoogendoorn et al. 2006; Ossen and Hoogendoorn 2005; Ossen and Hoogendoorn 2011; Kim and Mahmassani 2011). These studies have shown that drivers have individual behaviors under the same traffic conditions. This concept is called inter-driver heterogeneity. The same driver may behave in a different way in changing traffic conditions. This is called intra-driver heterogeneity. Heterogeneity amongst motorists has been studied in macroscopic traffic models in many different papers (Sun et al. 2014; Sun 2014, 2015; Pan and Sun 2012; Xiong et al, 2010). Accommodating the heterogeneity in the driver populace is unavoidable for creating more precise car-following models in the future (Kim et al. 2013). Nonetheless, heterogeneity in lane-changing models has not been acknowledged with much importance in most literature. The merging procedure is a systematic decision process that mirrors the active optimization of drivers' merging tactics by being aware of changes in surrounding traffic conditions. A driver is projected to have different decision guidelines and/or risk taking behavior for these two categories of lane changes, on different highway infrastructures (i.e. freeways, highways, and arterial streets), and under various traffic congestion stages, which normally correlate with various times during a day. Numerous previous studies use a systematic two-step model to describe the merging process: First, acceptable gap searching; second, merging execution (Ahmed 1999; Choudhury et al. 2009). Adjacent gaps are constantly likened with the "critical gap". This is defined as the minimum length of the acceptable gap, to decide whether the merging driver would accept the existing adjacent gap. Also, Sasoh (2002) proposed a model in which lane-changing needs 2 seconds, in order to model realistic driving behavior in lane-changing. The continuous-comparison process was condemned because it halts the continuation of merging actions and disregards the truth of real merging driving behavior (Sun and Elefteriadou 2014; Wang 2005; Wan et al. 2017). Hence, it is more accurate to describe and model the merging process in four steps as follows: (1) Acceptable Gap Searching, (2) Speed and Position Adjustment, (3) Gap Acceptance, and (4) Merging Maneuver or Lane Changing.

Compared to the merging process, key literature lacked information on the diverging process. The diverging process is unlike the merging process since diverging vehicles may not be in a lane adjacent to the auxiliary lane. Therefore, a five-step diverging process is as follows: (1) Move to the adjacent lane, (2) Acceptable Gap Searching, (3) Speed and Position Adjustment, (4) Gap Acceptance, (5) Diverging Maneuver or Lane Changing. This study emphasizes the gap selection step of the merging and diverging process: Acceptable Gap Searching. It uses a sequential method to form a gap selection model by integrating driver heterogeneity.

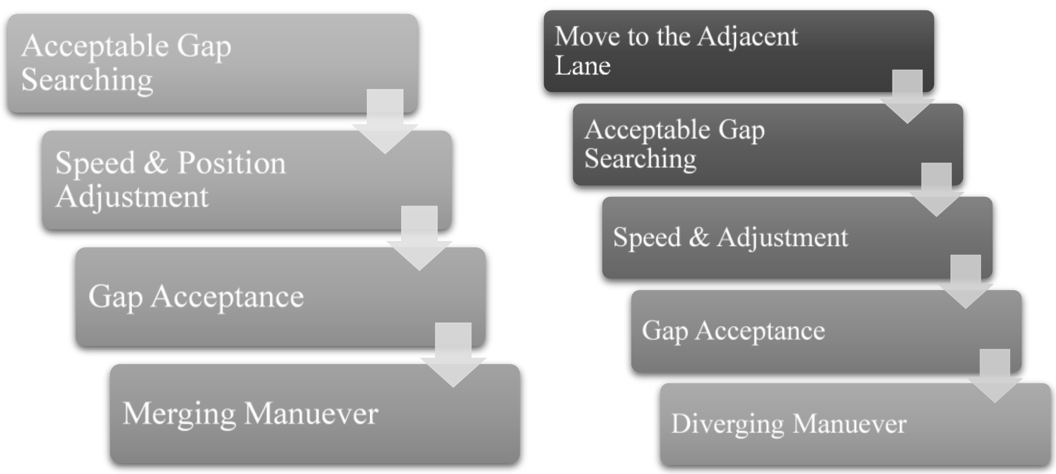


Figure 1(a): Driver Merging Process **Figure 1(b):** Driver Diverging Process

2.0 Literature Review

2.1 History

2.1.1 Modeling Gap Acceptance

The proposed topic of this thesis has been addressed previously by numerous researchers in the field. Their investigations have shown that amid the existing models, most studies have been aimed at gap acceptance representation through defining the critical gaps (Yang and Koutsopoulos 1996; Ahmed

1999; Lee 2006; Toledo et al. 2007, 2009). Gap acceptance models were primarily constructed to approximate the capabilities of non-signalized intersections. It is presumed that a driver executes a lane change when both his/her lead and lag gaps in the objective lane are longer than the critical gaps in the changing lane. The definition of the critical gap is the difference between these models. An exponential distribution was supposed by Herman and Weiss (1961), lognormal distribution was supposed by Drew et al. (1967), and normal distribution for critical gaps was supposed by Miller (1972). Gipps (1986) was donned to be the earliest to utilize the gap acceptance concept to develop a comprehensive structure for lane-changing models. While Gipps' model concentrated on urban driving conditions, the identical standard has been used in freeway mandatory lane changing models (Hidas 2002; Wang 2005) and microscopic traffic simulation software (Bloomberg et al. 2000; PTV 2004; SIAS 2005). Different definitions of critical gap were used in these simulations and software. For instance, in the VISSIM application, the critical gap hinges on the acceptable maximum deceleration for the lane-changing driver and his presumed follower for a mandatory lane change. Ten various driver categories can be defined through variable gap acceptance values in the CORSIM application (Bloomberg et al. 2000).

2.1.2 Traffic Simulation Model Theory

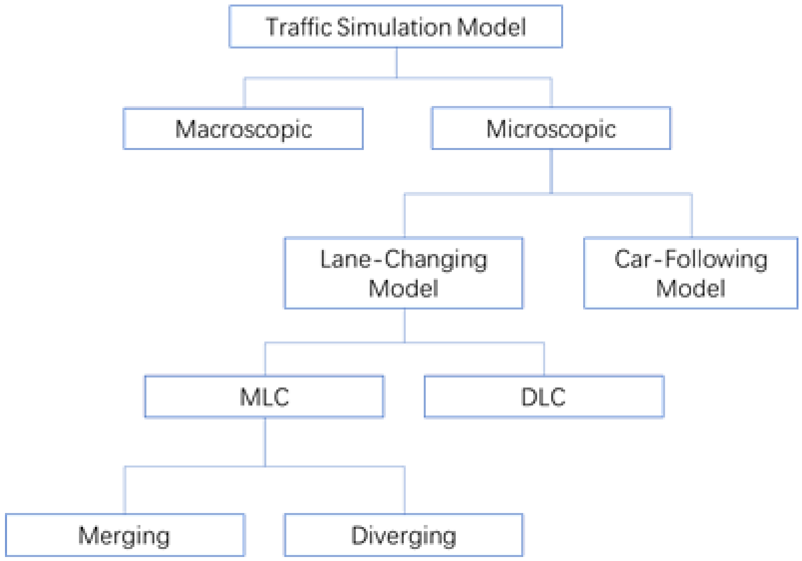


Figure 2. *Traffic Simulation Model Tree Chart (Munigety, 2016)*

The most predominant commercial microscopic traffic simulation tools could not replicate accurate traffic behaviors near merge areas under congestion as reported by Sarvi and Kuwahara (2007). "Forced" and "cooperative" lane-changing models were suggested to define distinctive behaviors of vehicles under congestion to surmount the deficiency (Ahmed et al. 1996; Hidas 2002; Hidas 2005). Built upon a series of studies (Ahmed 1999; Lee 2006; Rao 2006), an outline for merging process with latent plans was presented by Choudhury et al. (2007). Normal merge, merge with courtesy, and forced merge were measured in this outline. Marczak et al. (2013) saw that in this outline, rejected gaps were ignored and only accepted gaps were measured; and that some estimated coefficients in the model were not substantial.

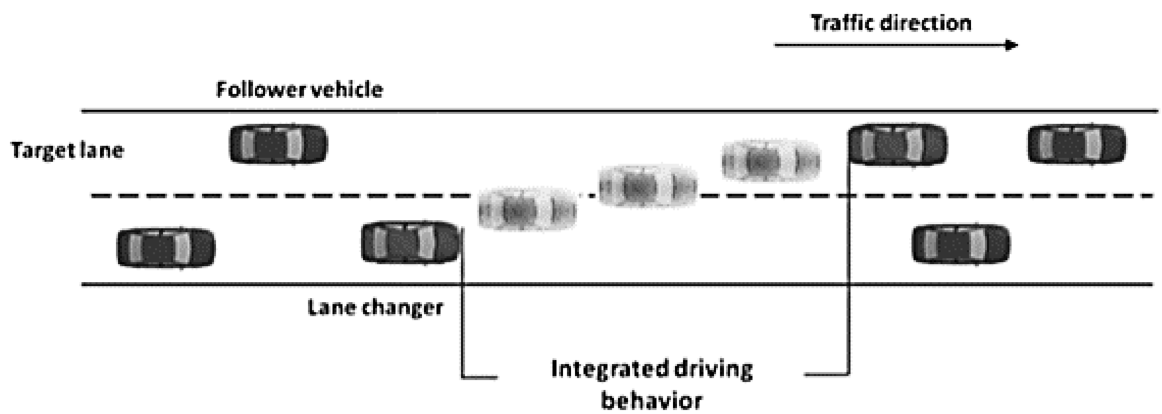


Figure 3. *Integrated driving behavior within the lane changer*

Gap acceptance using critical gaps is the most commonly utilized method in lane changing models. It is presumed that a driver will accept the adjacent gap only if both lead and lag gaps are longer than the critical gap (Punzo et al. 2005; Ma and Andreasson 2007; Hastie et al. 2001; Martin et al. 2012). Yet, this is frequently varying with the actual observations that vehicles remain taking lane changes only when the lead or lag gap is longer than the critical gap (Daamen et al. 2010; Marczak et al. 2013; Chu, 2014). Kita (1993) to surmount this deficiency created a binary logit model. Weng and Meng (2011) and Marczak et al. (2013) to forecast merging decisions in short-term work zone merging areas and to compare the gap acceptance of merging decisions at two sites were using the same kind of model. Marczak et al. (2013) established that the traffic conditions affect the parameters of the merging process model in his conclusion, but did not analyze the results of traffic conditions on the merging process. Traffic density is frequently used as a gauge on traffic conditions in the macroscopic approaches (Sun and Zhou 2005; Sun et al. 2010; Sun 2015). Still, it has not been effectively used in demonstrating lane-changing behaviors because of its inability to define vehicles' microscopic traffic behavior (Park et al. 2015). Because of these confinements, a hybrid model has been presented to replicate realistic traffic conditions, where the sections of the roadway are in homogeneous cells (Laval and Daganzo 2005, Park et al. 2015). Adding on to the fact that none of the studies has analyzed the effect of traffic conditions on the diverging process.

2.1.3 Exchanges between MLC & DLC

Toledo *et al.* created a probabilistic lane-changing decision model to outline the exchanges between Mandatory Lane-Changes and Discretionary Lane-Changes. The exchanges amongst MLC and DLC are obtained by keeping in mind mutual types of lane changes in a single effective operation. Note that even in a mandatory lane-changing situation, a motorist doesn't need to switch lanes straightaway. A discrete choice's center is used to model drivers' strategic and useful lane-changing choices. The model is controlled using maximum-likelihood estimation. The lane-changing decision model includes the choice of the destination lane and the choice for accepting a gap. Four clusters of explanatory variables were observed in the model underlying lane-changing decisions: regional variables (gaps and speeds), path plan variables (distance from the intended off-ramp), network knowledge and experience, and driving style & abilities. In the target lane model, the group of target lane selections include: remaining in the current lane, maneuvering to the right, and maneuvering to the left adjacent lane. The target lane decision model, the probability of choosing a specific lane, and the critical gap model are equivalent to those in Ahmed's model. In this model, the decision of accepting a target gap is based on a target lane choice. The model presumes that the driver will change to the target lane based on the acceptance of the lead and lag gaps in the target lane and does not count any other gaps. Toledo *et al.* identified the critical lead and lag gaps as the minimum acceptable gaps. When the existing target lead and lag gaps are greater than their corresponding critical values, they will be accepted. A log-normal distribution is supposed for the critical gaps to ensure that they are always positive. According to this model, after selecting a target lane and searching for gaps of acceptable sizes, the focus vehicle driver executes a sequence of accelerations and decelerations in order to move into the target lane. Toledo used a conditional probability to select whether a lead/lag gap is suitable or not. In Toledo's model, the subject vehicle employs a three-stage acceleration behavior model to select the target gap. Initially, if the subject vehicle wishes to remain in the current lane, a stay-in-the-lane selection model applies. Secondly, if the driver selects the available target gap and shifts into an adjacent lane, an acceleration model applies for the changing lane. Thirdly, if the subject vehicle accelerates or decelerates for changing lane but later rejects the target gap, a target gap acceleration model is used. This work has reformed the central outlook of gap acceptance models and how the focus affects traffic flow theory.

2.1.4 Modeling the Diverging Process

The safety of exit ramps and deceleration lanes have brought some concern and observation in the past on modeling the diverging process (Harwood & Graham 1983; Lundy 1967; McCartt et al. 2004; Oppenlander & Dawson 1970). These previous studies were all dedicated to the relationships between crash rates, and the geometrical components of deceleration lanes and exit ramps. None of these studies have attempted to investigate the diverging process. It was Bham (2009) who conducted the only study about diverging process in which the critical gaps of both merging and diverging behavior were projected through a method of the median rejected, mean rejected, and the largest rejected gaps less than the accepted gaps (LRLA).

2.1.5 Discretionary Lane-Changes

A lane change that is meant to better the observed driving conditions less immediately (e.g. overtaking a slow proceeding vehicle) is called a Discretionary Lane Change (DLC) (also known as free lane changes or desired lane changes). Discretionary lane changes are performed by motorists who are looking to better driving conditions for themselves or other motorists (Daamen et al. 2010). It usually happens when a driver desires a faster speed, more spacing, a further line of sight, and better ride quality inside the target lane (Balal et al., 2016, Balal et al., 2014, Pan et al., 2016, Zheng, 2014). Discretionary lane changing occurs in situations like if the driver wishes to pass a heavy vehicle or if a driver wishes to yield its way to another merging vehicle. A discretionary lane change transpires at the driver's own discretion and vehicle velocity is normally an aspect of it (i.e. the following vehicle is travelling too fast). DLC comprises of two decisions: whether the driving conditions are satisfactory, and if the motorist is not pleased, whether there is another lane that would be better choice than the current lane. The term satisfactory driving conditions suggests that the motorist is satisfied with the driving conditions in their current lane. Obviously, the motivations and resulting driving behavior for the two types of lane change (MLC or DLC) are different. Consequently, a motorist is predicted to have different decision rules or parameters for these two types of lane change.

History of DLC in Models

Authors have observed that, in some models, it was unfeasible to distinguish discretionary lane change from mandatory lane change. PARAMICS did not differentiate well between mandatory lane change and discretionary lane change. The lane changing model in PARAMICS was built on gap acceptance theory. AIMSUN split an off-ramp freeway segment upstream of into three zones, where discretionary lane changes took place in the most upstream zone to differentiate between discretionary and mandatory lane changes. TransModeler used the discrete choice approach to model drivers' lane changing decisions. It reflected on three types of lane change: discretionary, mandatory and forced lane changes. A discretionary lane change is usually observed when a driver is displeased with the current speed. There were two discretionary lane change models: target lane model and neighboring lane model. The neighboring lane model had the target lanes adjacent to the original lane. In comparison, the target lane models moved the subject vehicle by more than one lane. A logit model calculated the probabilities of drivers selecting the left or right adjacent lane in the neighboring lane models. The parameters of the utility function were average velocity gain and slow lead vehicle. Once a target lane has been selected, the subject vehicle looks for an appropriate gap in the target lane to maneuver into. Gipps (1986) is possibly one of the first to document a lane change study in a signalized street. His driver's decision making structure entailed of the possibility, necessity and interest to change lanes. He then proposed a lane changing model including mandatory and discretionary lane changes. The decision parameters for discretionary lane changes included the subject vehicle's “safe speed”, relative speed between the original lane and the target lane, speed of vehicles in the target lane, and headway between preceding and subject vehicle. He figured that the purpose of a discretionary lane changes are for the subject vehicles to increase speed or to better their position in the traffic stream. Koutsopoulos (1996) depicted his own method of using discrete choices for modeling lane-changing behavior. This model was based on the gap acceptance model. A scarce amount of existing lane-changing models were based on real traffic data and were typically tested through simulations because they did not generate incidents or interrupt traffic flow.

It is not entirely revealed how differences in traffic conditions relate to discretionary lane-changing. There is lack of firsthand evidence about discretionary lane-changing, since it is difficult to detect motivations for discretionary lane changing. The relationship between discretionary lane changes and traffic environments were investigated by Knoop et al. (2010), but that study focused on the effect of one factor on a number of free flow lane changes at the same time even if many of the influencing aspects were looked into independently, such as difference in velocity, traffic conditions, and traffic density values in original and target lanes. These aspects could differ in what leads to a number of discretionary lane changes in free flow condition. This study was inadequate because of that. Park et al.

(2014) suggested the logistic regression behavior model of discretionary lane changing particularly under congested traffic conditions with attention to a variety of explanatory variables. More recent literature covers the updated geometric lane changing models for simulations (e.g. Li et al., 2014).

2.1.6 Lane-changes and Traffic Flow, Operations, & Management

Lane changing movements could have a significant effect on traffic flow characteristics as a consequence of their interfering effect on neighboring vehicles. The interference effect of lane changing is more prominent when heavy vehicles change lanes compared to when passenger cars perform the same maneuver (Munigety, 2016). This is because of the physical effects that heavy vehicles inflict on surrounding traffic. Heavy vehicle drivers generally change into the slower lanes to avoid obstructing the high-speed moving vehicles that advance from the rear. Meanwhile, passenger car drivers raise their speed according to the speeds of the lead and lag vehicles in the target lane. They more frequently move into the high-speed lanes to gain speed advantages.

Major roadways usually have multiple lanes and vehicles can make both longitudinal and lateral movements. A lane-changing area is a section, where one or more streams of vehicles systematically can change their lanes. These areas could be nearby a merging junction and lane-drops, upstream to a diverging junction, inside a weaving section, or around a cloverleaf interchange (Milam and Choa, 1998; Cassidy and Rudjanakanoknad, 2005). Since bottlenecks (Hall and Agyemang-Duah, 1991) and accidents (e.g. Golob et al., 2004) have a habit of occurring in these sections, it is vital that we comprehend the occurrences associated with lane-changing traffic (Jin, 2018).

2.2 The Multinomial Logit Model

Logistic regression is an expansion of simple linear regression. Logistic regression is a commonly-used technique as it permits binary variables, the summation of binary variables, or polytomous variables (variables with more than two classes) to be modeled. Logistic and linear regression fit in the same group of models called Generalized Linear Models (GLM): in each case, an occurrence is related to a linear arrangement of explanatory variables. For logistic regression, the dependent variable (response variable) abides by a Bernoulli distribution for the parameter p (p is the mean probability that a result will arise) when the trial is repeated once, or a Binomial (n, p) distribution if the trial is repeated n times. The probability parameter p is here a linear combination of explanatory variables. The most usual operations used to link probability p to the explanatory variables are the Logit models and the Probit models.

Where the dependent variable is dichotomous (having two outcomes) or binary in nature, we can't use simple linear regression. Binary logistic regression is the statistical procedure used to predict the correlation between predictors (independent variables) and the predicted variable (dependent variable) where the predicted variable is binary. There must be two or more predictors for a logistic regression. The predictors can be continuous variables, nominal variables, or categorical variables. All predictor variables are confirmed in one block to calculate their predictive ability whilst controlling for the influences of other predictors in the model.

The rules for a Binary logistic regression model are for a sufficient sample size, because too few cases for numerous predictors is flawed, and an absence of multicollinearities (which are high intercorrelations among the predictors), and no outliers. This can make it difficult to run tests on large sample sizes, especially when large sample sizes are needed for accurate representation of data.

While a model used in the binary case with only two outcomes is based on a binomial distribution, when there are more than two outcomes, the model we use is based on multinomial distribution. The multinomial, also known as polytomous, logistic regression model is a straightforward extension of the binomial logistic regression model. They are used when the dependent variable has more than two nominal outcomes. Dummy coding of independent variables is fairly ordinary. The multinomial logistic regression estimates a separate binary logistic regression model for each of these dummy variables. The result is M-1 binary logistic regression models. Each one tells the result of the predictors on the probability of success in that category in comparison to the reference category. Each model has its own intercept and regression coefficients. The chosen predictors can affect each category differently. One may ask themselves, “Why not just run a series of binary regression models?” That could be done, and people used to, before multinomial regression models started being commonly available in software. You would probably get comparable results. However, running them together means they are estimated simultaneously, which means the parameter estimates are more proficient. There are generally less unexplained errors. The multinomial Logit model is also an extension of multiple regression modelling, where the dependent variable is discrete instead of continuous, permitting the modeling of discrete outcomes. Multinomial logit is designed for outcomes that are not complexly interrelated. In particular, for this study, we are interested in characterizing the probability of individual choices conditioned to the values of the attributes and characteristics.

When the response categories are ordered, a multinomial regression model could be run. The disadvantage is that information about the ordering is thrown away. An ordinal logistic regression model keeps that information, but it is somewhat more intricate. In the Ordinal Logistic Regression (also known as Ordinal logit model and OLS), the occurrence being modeled is not having a result in a single category as in the binary and multinomial models; the occurrence being modeled has an outcome in a specific category or any previous category before it. In the OLS, each outcome has its own intercept but the same regression coefficients. This means the complete odds of any occurrence can fluctuate, but the effect of the predictors on the odds of an occurrence happening in every subsequent category is the same for every category. This is an assumption of the model that needs to be checked. It is usually violated. It also makes the very significant assumption that the correlation of predictors to the odds of a response occurring in the next higher order category is the same regardless of which categories are being compared. This is called the proportional odds assumptions or the parallel regression assumption. Unfortunately, this assumption is difficult to encounter in real data.

The goal of the multinomial logistic regression is to construct a model that explains the relationship between the independent variables and the dependent variable. When using this regression, one category of the dependent variable is selected as the reference category. Separate odds ratios, the odds of an event occurring given some factor compared to the odds of an event occurring in the absence of that factor, are determined for all independent variables for each category of the dependent variable with the exception of the reference category, which is omitted.

Multinomial Logistic Regression has been used to analyze many traffic situations already. For example, it has been used to analyze some Missouri work zone crash data to identify significant factors which affect the severity of crashes in the area. This specific type of regression analysis was used due to the mixed nature of data. Multinomial regression was used to compare crashes by severity; Property Damage Only against crashes with Minor Injuries and Disabling Injuries/ Fatal.

2.3 Driver Heterogeneity

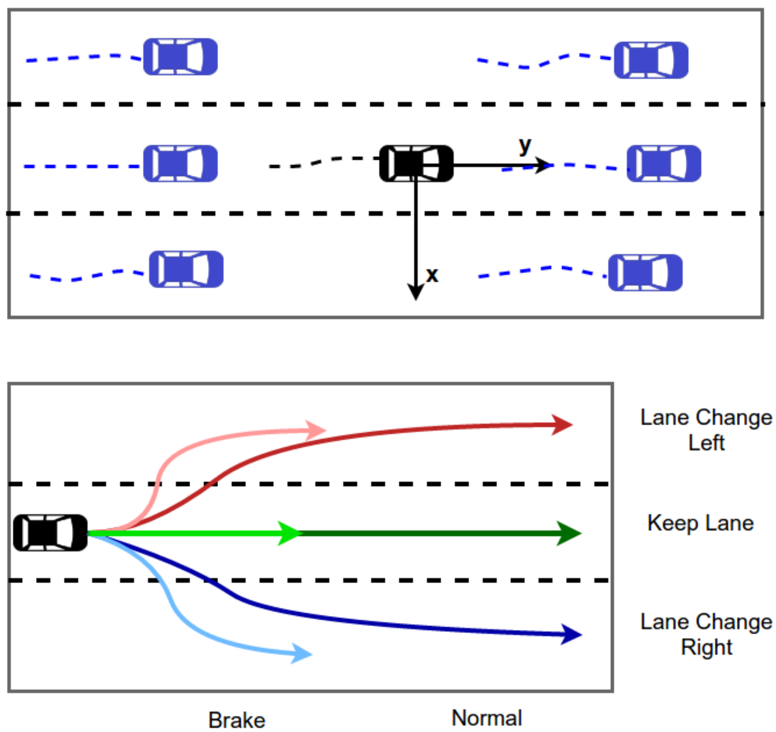


Figure 4. *Driver lane choice*

Recently, driver heterogeneity has been deemed as a vital element in driver behavior (Pan and Sun 2012; Sun et al. 2011; Sun 2014) and its effects are crucial for the furthering of traffic studies. Numerous studies have been taken to investigate heterogeneity in car following models by developing the joint distribution of model coefficients dependent on an empirical basis (Ossen et al. 2006; Hoogendoorn et al. 2006; Ossen and Hoogendoorn 2005; Ossen and Hoogendoorn 2011; Kim and

Mahmassani 2011). Also, heterogeneity has been effectively studied in numerous macroscopic traffic models (Sun et al. 2001; Sun et al. 2010; Sun et al. 2011). Driver heterogeneity in lane-changing models have not gained similar attention. Traffic simulation software packages, like VISSIM and PARAMICS, are using an aggressiveness parameter to differentiate between aggressive and timid drivers (PTV 2004; SIAS 2005). The research that integrates driver heterogeneity is mainly a series of studies from Ben-Akiva's group: Ahmed (1999), Lee (2006), Rao (2006), Choudhury et al. (2007), Toledo et al. (2009). A lane-changing decision configuration involves latent (also known as unobservable) levels of decision structure suggested by previous research. Driver heterogeneity was measured by utilizing an individual-specific latent variable. Conversely, they did not deliberate the differences between the parameters amongst different drivers.

3.0 Problem Statement

Driver heterogeneity is now deemed as a crucial component of driver behavior. Heterogeneity in lane-changing models has not been acknowledged in equal emphasis as in car-following models. This study aims to create a logit models to define the driver gap selection behavior during the merging, diverging, and discretionary processes, while bearing in mind the heterogeneity between drivers' characteristics, and maneuvering strategies. One trajectory dataset will be utilized to calibrate the suggested multinomial logit model. The multinomial logit models are proposed to define the lane-change maneuvering processes while including driver heterogeneity.

3.1 Motivation: Why is this research timely?

There are many reasons that this research is timely starting with the need to go into a deeper understanding microscopic traffic flow. Vehicle angles were calculated and added to these dataset to include additional factors and variables in driver lane decision such as driver angle which contains the driver's visual angle. We are redefining the lane-changing maneuver process to include the recording of multiple rejected gaps and analyzing some differences in behaviors in multiple situations (mandatory lane change, discretionary lane change, congested flow, and different structural environments). We can learn different methodologies and understanding how they work to achieve various goals on a wider scale. Adding to the existing knowledge in this area of research is imperative to attain that goal. The fact is that every day we are moving closer to having automated cars on the common road. That affects future traffic safety in total. These models could be used for the algorithms made. Traffic density and conditions has to be included in these models too. The study will present an overall methodological structure for characterizing lane-changing behaviors while integrating driver heterogeneity. The findings of this study should improve the understanding of driving behaviors in microscopic traffic flow. This will aid the improvement of traffic safety, operation, and simulation.

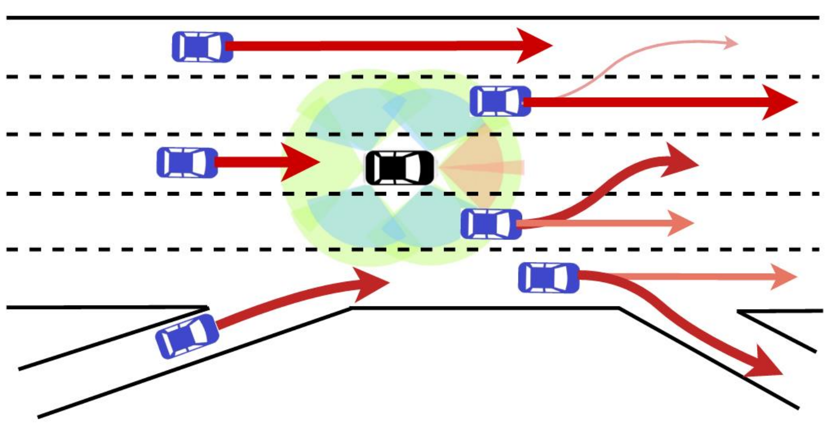


Figure 5. *Driver Angular Vision & Lane Targeting*

3.2 Aims

My aims are to:

1. Identify how all variables do, or do not, play a factor in driver behavior.
2. Expose driver bias towards any continuous or categorical variables.

3. Verify any behaviors in congested traffic behavior.
4. Define any possible class of driver behavior.
5. Analyze how traffic density influences driver lane-changing decisions.
6. Recognize how the same drivers may behave differently in varying situations.

3.3 Research Objectives

The main objectives of this research are to develop an overall methodological structure for characterizing lane-changing behaviors while integrating driver heterogeneity and to prove that traffic density and conditions are actual influential factors in drivers' lane-changing behaviors. To accomplish these objectives, the following tasks are performed:

- Construct Multinomial Logit Models based on the extracted trajectory data.
- Examine results of all the models and determine the characterization of the lane-changing behavior with the integrated driver heterogeneity.

4.0 Research Methodologies

4.1 Research Environment

The NGSIM vehicle trajectory data gathered on a section of southbound U.S. Highway 101 (Hollywood Freeway) in Los Angeles, CA will be chosen in this study in order to examine the merging & diverging process in weaving sections (Alexiadis et al. 2004). Figure 6 displays the sections for the U.S. Highway 101 dataset.

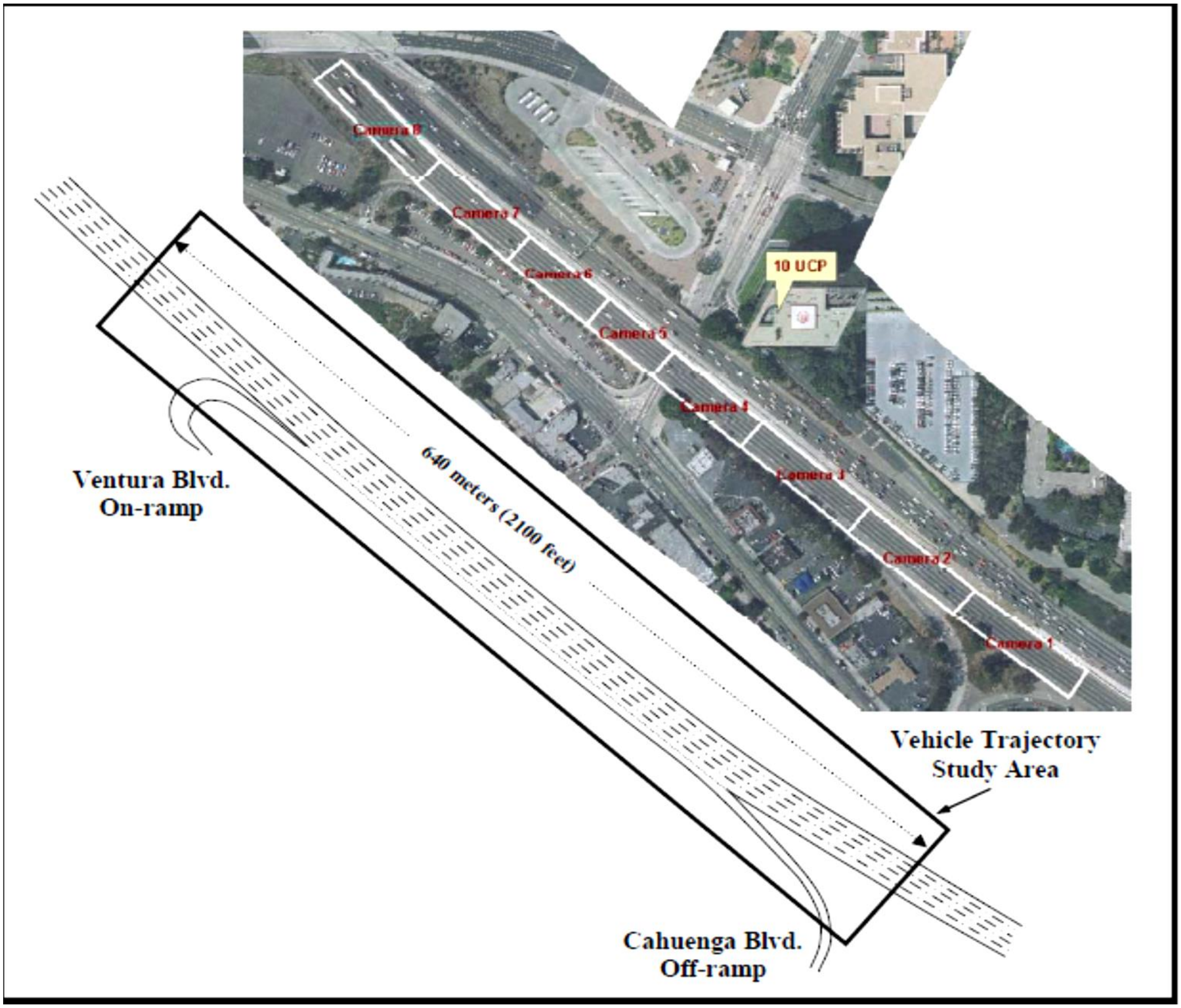


Figure 6. U.S. Highway 101 study corridor from NGSIM (Cambridge Systematics, Inc. 2005)

The segment of US-101 is 640 meters (2100 ft.) long and has five main lanes and one auxiliary lane. The vehicle trajectory datasets were accumulated from 7:50 A.M. to 8:35 A.M, Pacific Time on June 15, 2005. The road segment has eight cameras covering the whole area. All vehicle trajectory subsets were updated at a resolution of 10 fps (frames per second) (Cambridge Systematics, Inc. 2005). The location's dataset has three subsets.

Table 1. *Constants in the US-101 & I-80 Datasets*

Site	**US-101**
Merge Dataset(s)	1
Diverge Dataset(s)	1
Total Sample Size	6,101
Length	2100 ft / 640 m
Upstream Length	578 ft / 176 m
Middle Length	698 ft / 213 m
Downstream Length	824 ft / 251 m
Lanes	5
Merge Lane	1
Auxillary Lane(s)	1 (On & Off Ramp)
Adjacent Lane(s)	0
On-Ramp	Ventura Blvd
Off Ramp	Cahuenga Blvd
Merge Decision Point	Global X: 6451525.29 Global Y: 1872909.55
Diverge Decision Point	Global X: 6452049.22 Global Y: 1872451.53
Ending Point	Global X: 6452734.58 Global Y: 1871874.94
Cameras	8
Camera Cell Length	262 ft / 80 m
Frames per Second	10
Total Frames	23,963
Detector Stations	5
Detector Cell Length	525 ft / 160 m
Special Length Details	
Subsets	3 (7:50 AM - 8:05 AM; 8:05 AM - 8:20 AM; 8:20 AM - 8:35 AM)
Subset Sizes	2,169; 2,017; 1,915
Subset Time Frame	15
Total Minutes	45
Subset Frames	7,192; 9,944; 4,652
Speed Units	Detector: MPH | Trajectory: FPS

4.2 Trajectory Datasets

4.2.1 U.S.101 Trajectory Dataset

The US-101 dataset was collected on June, 2005 by FHWA on a segment of U.S. Highway 101, called Hollywood Freeway, in Los Angeles, California. The data represents travel southbound of US-101. Eight video cameras are installed on a 36-story building, on 10 Universal City Plaza, right next to US-101. Similar to I-80, the eight cameras recorded eight sub-sections of the study area, respectively. The US-101 trajectory dataset has also three subsets, each 15 minutes (7:50 a.m. 8:05 a.m., 8:05 a.m. ~ 8:20 a.m., and 8:20 a.m. ~ 8:35 a.m.). This study uses the 8:05 a.m. ~ 8:20 a.m. subset. The study site is approximately 2100 feet in length, with five mainline lanes and one auxiliary lane connecting to the Ventura Blvd on-ramp and the Cahuenga Blvd off-ramp. This US-101 vehicle dataset is also useful for the modeling of the lane-changing behavior in congested conditions because its structural characteristics, including the on-ramp and off-ramp, and the data collected during morning rush hour**.**

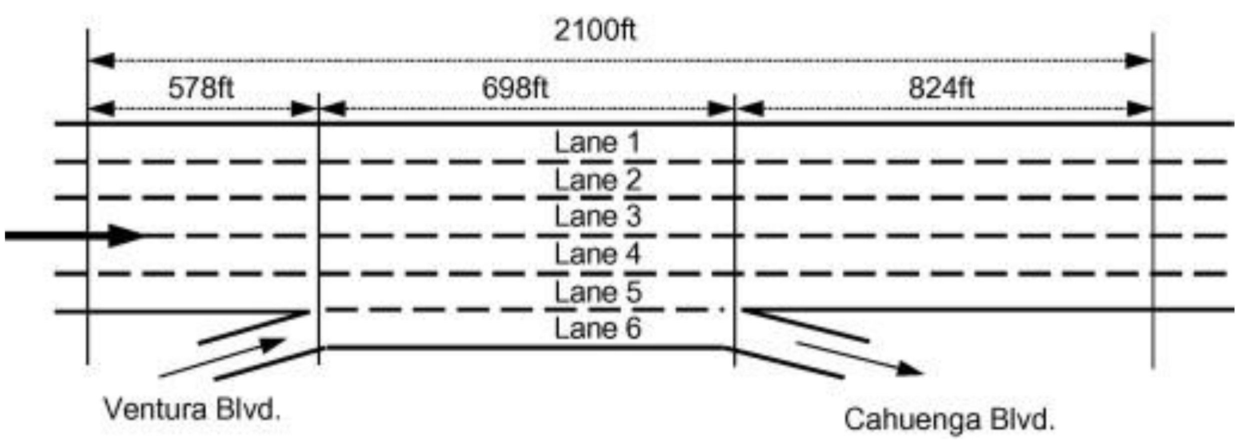


Figure 8. *US-101 Structural Layout*

The velocity of mainline traffic varied from 27.3 km/h to about 51.50 km/h. The average velocity was about 41.03 km/h. The average initial velocity of the on-ramp merging vehicles in the auxiliary lane is about 49.48 km/h. That is a higher average velocity than the mainline traffic.

Table 2. *Time mean speed & Traffic Flow*

US-101

Traffic Condition	Time Period (PDT)	Flow (VPH)	Time Mean Speed	
			(m/s)	(km/h)
Congested	8:05 AM - 8:20 AM	8016	11.1	39.96

4.3 Research Respondents

The respondents in the datasets come in the form of vehicle ID numbers. Matched with every vehicle ID number is a frame ID number, so one may look up vehicle information at different time frames. The frames were taken at 10 fps (frames per second) and the trajectory data was recorded in that interval. The variable “global time” represents the specific times the datasets were recorded in epoch time. In the US-101 dataset there were a total of 6,101 drivers recorded & more than 20,000 frames, resulting in 4.1 million lines of data. We will be analyzing data from the 8:05 AM – 8:20 AM subset which has about 3,109 subject vehicles.

Due to the large size of the data files, the lines of data were reduced to 3 million lines of data (There were 4.5 million lines in the total population). Additional subsets were made from the following categories: discretionary lane-changes, mandatory lane-changes, merging maneuvers, and diverging maneuvers.

4.4 Research Instruments

Due to the size of the dataset and lack of personal funds, this research was done mostly on Alteryx Designer, Minitab, and SPSS. Alteryx Designer was used for extracting data from the data files, adding columns, and verifying data. Minitab was used for calculating basic frequencies and statistics due to its inability to handle larger sample sizes. SPSS was used for creating dummy variables, creating the models, and analyzing distributions, due to its ability to handle large datasets.

4.5 Research Procedures

4.5.1 Gathering of Data

To improve the accuracy of lane-changing models, large vehicle trajectory data sets are mandatory at the individual (microscopic) level. Data such as velocity, acceleration, lane changes, and the variables defining the relationship between the subject vehicle and other vehicles including relative speed, relative position, time, and space headways are required with great time resolution to estimate lane-changing model parameters. This data was acquired from the American Federal Highway Association (FWHA). The datasets are free to download and use from their website. Traffic data provided by the Federal Highway Administration’s NGSIM project was used to build these lane-changing models. The NGSIM data set is an open-source dataset that has been used in previous research for simulation model development and testing. NGSIM datasets include detector data and vehicle trajectories on a segment of U.S. Highway 101 (Hollywood Freeway) in Los Angeles, CA, (going southbound). The datasets represent two traffic conditions: states when congestion was building up (i.e. the period of the first 15 min), which are indicated as the transition periods, and congested states (i.e. the period of the remaining 30 min). The data used from our subset represents the transformation from transition to congested traffic flow.

4.5.2 Selecting Independent & Dependent Variables

A variable is any parameter in the study that can change. To comprehend the idea of independent and dependent variables, one ought to comprehend the meaning of variables in general. Variables are described as the properties or types of characteristics of specific events or objects.

Independent variables are variables that are controlled or are changed by researchers and whose effects are measured and evaluated. This is the variable that is stable and unaltered by the other variables you are trying to measure. It denotes the condition of an experiment that is systematically manipulated by the researcher. It is the apparent cause. The other name for the independent variable is the predictor. The independent variables are called this per se because independent variables predict or estimate the values of the dependent variables in the model.

The other variables are considered the dependent variables. These are the variables that depend on other factors that are measured. These variables are anticipated to change as a result of an investigational operation of the independent variable or variables. It is the supposed effect. The dependent variable denotes the kind of variable that processes the effect of the independent variable on the experimental component. We can also say that the dependent variable is the type of variables that is completely dependent on an independent variable. The other name for dependent variables is predicted variables. The dependent variables are called this per se because they are the values that are predicted or presumed by the predictor (independent) variables.

Variable selection is a means to an end and not an end itself. The goal is to build a model that predicts appropriately or clarifies the relationships in the data. A variable in research purely denotes a person, place, thing, or occurrence that you are attempting to measure in any particular way. The greatest way to comprehend the difference between dependent and independent variables is to understand that the significance of each is understood by what the words tell you about the variable you are using. "The [independent variable] causes a change in [dependent variable] and it is not possible that [dependent variable] could cause a change in [independent variable]." Insert the names of variables you are using in the sentence. This will help you recognize each type of variable.

4.6 Definitions

4.6.1 Definition of Terms

Lane-Changing Model – First intended for microscopic traffic simulation tools (Gipps, 1986), the original lane-changing model covers various urban driving situations in which traffic signals, transit lanes, obstructions, and the presence of heavy operating vehicles affect drivers' lane selection.

Lane Change – The transfer of a vehicle from one lane to an adjacent lane.

Freeway – An express highway, especially one with controlled access (Toll-Free Highway).

Passing Lanes – A lane designated for passing vehicles.

Urban Way – A city highway for drivers who want to pass slower traffic.

On-Ramp – A lane for traffic entering a turnpike or freeway.

Off-Ramp – A one-way road leading off a main highway.

Exit Ramp – A short section of road which allows vehicles to enter or exit a highway

Car-Following Models – (Microscopic Traffic Flow Model) – Time-continuous models defined by ordinary differential equations describing the complete dynamics of vehicles' positions and velocities.

VISSIM – A visual block diagram language for simulation of dynamic systems and model-based design of embedded systems.

Driverless Car – (Automated Car) – A car which does not have a human driver in control, but may be controlled by AI or any intelligence system.

Traffic Safety – Refers to methods and measures for reducing the risk of a person using the road network from being killed or seriously injured. The users of a road include pedestrians, cyclists, motorists, their passengers, and passengers of an on-road public transport, mainly buses and trams.

Traffic Flow – The study of interactions between vehicles, drivers, pedestrians, cyclists, other travelers, and infrastructure, with the aim of understanding and developing an optimal road network with efficient movement to traffic and minimal traffic congestion problems.

Macroscopic – Of or relating to large-scale or general analysis.

Visual Angle – The angle a viewed object subtends at the eye, usually stated in degrees of arc.

Traffic Simulation – The mathematical modeling of transportation system through the application of computer software to better help plan, design, and operate transportation systems.

Microsimulation – A category of computerized analytical tools that perform highly detailed analysis of activities such as highway traffic flowing through an intersection, financial transactions, or spreading of disease. It is also a term used in traffic modelling and typified by software packages simulating the behavior of individual vehicles within a predefined road network and used to predict the likely impact of changes in traffic patterns resulting from changes to traffic flow or from changes to the physical environment.

B – This is the unstandardized regression weight. It is measured just a multiple linear regression weight and can be simplified in its interpretation.

S.E. – Like the multiple linear regression, this is how much the unstandardized regression weight can vary by. It is similar to a standard deviation to a mean.

Wald χ^2– This is the test statistic for the individual predictor variable. A multiple linear regression will have a *t* test, while a logistic regression will have a χ^2 test. This is used to determine the *p* value.

p – This is used to determine which variables are significant. Typically, any variable that has a *p* value below .050 would be significant. When the p-value is small, you can reject the null hypothesis and conclude that the parameter is not equal to 0 and it does contribute to the model.

EXP(B) – this is the odds ratio. This is the measurement of likelihood.

95% *CI OR* – this is the 95% confidence interval for the odds ratio. With these values, we are 95% certain that the true value of the odds ratio is between those units. If the confidence interval does not contain a 1 in it, the *p*-value will end up being less than .050.

Parameter estimates (also called coefficients) are the change in the response associated with a one-unit change of the predictor, all other predictors being held constant. The unknown model parameters are estimated using least-squares estimation.

A **standardized parameter estimate** (commonly known as standardized beta coefficient) removes the unit of measurement of predictor and response variables. They represent the change in standard deviations of the response for 1 standard deviation change of the predictor. You can use them to compare the relative effects of predictors measured on different scales.

Classification is the problem of predicting a discrete class label output for an example.

Regression is the problem of predicting a continuous quantity output for an example.

Upstream refers to the direction the traffic is flowing from.

Downstream refers to the direction the traffic is flowing towards

4.6.2 Definition of Models and Techniques

Multinomial Logit Model

One of the explanations on why Multinomial Logistic Regression is a good choice for this data is that it does not presume normality, linearity, or homoscedasticity (Starkweather, 2011). The driver chooses among more than two choices, once again, making the choice that provides the greatest utility.

In the multinomial situation, the observed response is simply a label for the nominated choice; it could be anything: a brand, the name of a place, or the type of travel mode. Numerical assignments are not meaningful.

The goal of the MNL regression is to create a model that describes the relationship between the independent variables and the dependent variable. When utilizing this regression, one category of the dependent variable is chosen as the reference category. Separate odds ratios, the odds of an event occurring given some factor compared to the odds of an event occurring in the absence of that factor, are determined for all independent variables for each category of the dependent variable with the exclusion of the reference category, which is absent.

The Multinomial Logit Model can be effectively utilized to model the lane-changing process. Picture a dependent variable with M categories; the probability of drivers "i" choosing "j" must add up to 1.0:

$$\sum_{j=1}^{J} p_{ij} = p_{i1(Initial\_Gap)} + p_{i2(Nested\_Gap)} + p_{i3(Accepted\_Gap)} = 1 \tag{1}$$

We can model the probability of each outcome as:

$$p_{ij} = \frac{e^{\sum_{j=1}^{K} \alpha + \beta_{kj} X_{kji}}}{\sum_{j=1}^{J} e^{\sum_{j=1}^{K} \alpha + \beta_{kj} X_{kji}}} \tag{2}$$

where p_{ij} is the probability that driver i accepts the jth offered gap; $x_j(j=1\cdots J)$ represents the descriptive variables; α is the constant and $\beta_j(j=1\cdots J)$ is the parameter to be assessed.

Solved by adding constraint; Coefficient sum to zero.

$$\sum_{j=1}^{J} \beta_{jk} = 0 \tag{3}$$

The log-odds of each response can be defined using the equation:

$$\eta_{ij} = \log \pi_{ij}\pi_{iJ} = \alpha j + \boldsymbol{x'}_I \boldsymbol{\beta}_j \tag{4}$$

The log-odds of choosing the initial gap as the accepted should not change if "nested gap" is added or removed from the decision process. If initial gap is rejected, those drivers should choose the nested gap as the accepted gap in similar pattern to the rest of the sample.

In Multinomial logistic regression one category is chosen as a "reference"…

- Probability of initial gap rejection vs. gap acceptance
- Probability of nested gap rejection vs. gap acceptance
- Probability of initial gap rejection vs. nested gap rejection

$$\Pr(Y_i = 1) = \frac{e^{\beta_1 \cdot \mathbf{X}_i}}{\sum_{k=1}^{K} e^{\beta_k \cdot \mathbf{X}_i}} \quad \Pr(Y_i = 2) = \frac{e^{\beta_2 \cdot \mathbf{X}_i}}{\sum_{k=1}^{K} e^{\beta_k \cdot \mathbf{X}_i}} \quad \ldots\ldots \quad \Pr(Y_i = K) = \frac{e^{\beta_K \cdot \mathbf{X}_i}}{\sum_{k=1}^{K} e^{\beta_k \cdot \mathbf{X}_i}} \tag{5}$$

In this model the dependent variable can be any categorical variable. It doesn't need to be positive or sequential. The output will include two tables: factors affecting the probability of choosing to reject the initial gap vs. accepting the gap & factors affecting probability of rejecting the nested gap vs. accepting the gap.

4.7 Model Parameter Estimation

As soon as the vehicle concerning a new adjacent gap i appears, the merging or diverging vehicle driver n may either choose "$y_n^i = 1$" accepting the gap, or "$y_n^i = 0$" rejecting the gap. A driver may only accept one gap after they reject 0 or several gaps. Therefore, the gap choice order of a driver is,

$$(\underbrace{0\ldots0}_{I_{n-1}}, 1) \tag{6}$$

where I_n is the number of offered gaps for the n^{th} merging or diverging driver.

For given class assignment, the contribution of merging or diverging vehicle n to the likelihood would be the joint probability of the gap choice sequence $(\underbrace{0\ldots0}_{I_{n-1}}, 1)$ (Greene and Hensher 2003; Greene 2007):

$$P_{n|c} = \prod_{i=1}^{I_n} P_{n|c}^{i} \tag{7}$$

The class assignment remains unidentified. Let H_{nc} specify the prior probability that the merging or diverging vehicle n fits to the class c. The unconditional likelihood of observing the gap choice sequence of merging or diverging vehicle n is displayed in Equation (10):

$$P_n = \sum_{c=1}^{C} H_{nc} P_{n|c} \tag{8}$$

Thus, the log-likelihood (*LL*) of all merging or diverging vehicle drivers in each dataset is stated in Equation (11) (Greene and Hensher 2003; Greene 2007):

$$LL = \sum_{n=1}^{N} \ln P_n = \sum_{n=1}^{N} \ln[\sum_{c=1}^{C} H_{nc} (\prod_{i=1}^{I_n} P_{n|c}^{i})] \tag{9}$$

where N is the number of merging or diverging vehicles.

The Log-Likelihood in the multinomial logit model may be expressed by

$$l(\beta; y, X) = \sum_{i=1}^{N} [-\ln(1 + \exp(x_i\beta)) + y_i x_i \beta] \tag{10}$$

Given data and the number of classes, the objective is to estimate the structural parameter vectors, $\boldsymbol{\beta}_c$, and the class probability parameter by maximizing the LL function, can be solved by maximum likelihood estimation.

The maximum-likelihood estimate is used to predict the lane-changing probability of the maneuvering vehicle. The maximum likelihood method entails finding model parameters which maximize the likelihood of the observed selections conditional on the model. Namely, to maximize the likelihood that the sample was made from the model with the designated parameter values. The method for maximum likelihood estimation involves two major steps: First, developing a joint probability density function of the observed sample (called the likelihood function), and second, estimating parameter values which maximize this likelihood function.

The gap acceptance model is estimated using the maximum likelihood method with the vehicle trajectory dataset. The explanatory variables affect the drivers' lane-changing behaviors in the congested situation of the subset. Not all the parameters are significant so only significant parameters are identified in the upcoming tables. All available categorical and continuous variables were chosen to develop a model complex enough to identify driver heterogeneity in lane-changing behavior.

Unlike Ordinal logistic regression, Binary logistic regression uses maximum likelihood to estimate model parameters. Maximum likelihood estimation is an iterative process aimed at arriving at population (parameter) values that most likely produced the observed (sample) data. In general, this estimation approach assumes larger samples and, aside from issues of power, smaller sample sizes can create problems with model convergence and estimation of model parameters. As a side note, with smaller samples, exact logistic regression or firth procedure using Penalized Maximum Likelihood can be used. Unfortunately, these options are not commonly available in statistics programs. When a parameter is not considered to be a factor statistically to the model, you can consider removing it. Yet, you should be cautious of removing parameters that are known to contribute by some underlying system, regardless of the statistical significance of a hypothesis test, and understand that removing a variable can change the effect of others.

With the evaluation of model fit, as also with standard ordinal logistic regression, occurs on two levels. The first level involves evaluating the fit of the full model (containing the full set of predictors), which is done using a likelihood-ratio chi-square test

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{11}$$

which compares the full model with a null, or intercept-only and the Hosmer-Lemeshow test results. The Hosmer-Lemeshow test was omitted from this study due to its incapability to correctly test the model, coming from its strictness and it inconsistency. Moreover, overall model fit is often assessed using “pseudo-r-squared” indices and evaluation to the degree in which the model is able to classify individuals into groups on the dependent variable (Smith & Mckenna, 2013). The following one evaluates the individual predictors for their contribution to overall model fit. This is done by using either Wald tests or likelihood ratio tests.

The likelihood ratio test, which is used to compare log likelihood functions for models of interest, can be used to compare the estimated models.

$$LR = -2ln\left(\frac{L(m_1)}{L(m_2)}\right) = 2(loglik(m_2) - loglik(m_1)) \tag{12}$$

Likelihood ratio test involves comparing the full model with all the predictors against a reduced model with a given predictor removed. The statistic -2LogL (minus 2 times the log of the likelihood) is a badness-of-fit indicator. Large numbers denote poor fit of the model to the data. When selected from large samples, the difference between two values of -2LogL is delivered through the chi-square value.

The most important and difficult step in building multinomial logit model is to determine C , the number of classes. Since C is not a parameter, the hypotheses on C cannot be verified directly. Numerous criteria could be used to determine this, such as BIC and AIC (Akaike 1974; Schwarz 1978; Biernacki et al. 2000; Greene and Hensher 2003). The Bayesian Information Criterion (BIC) is an index used in Bayesian statistics to choose between two or more alternative models. Akaike's information criterion (AIC) compares the quality of a set of statistical models to each other. Although the AIC will choose the best model from a set, it won't say anything about absolute quality. It has been stated in Allenby (1990) that the AIC tends to excessively favor a model with more number of classes for large data samples and the BIC corrects this situation because it takes into account the sample size used in the study. Therefore, in this study, we will determine C by using BIC criteria:

$$BIC_{\text{model}} = -2LL + \log(N)\gamma \tag{13}$$

Where LL is the log-likelihood value, γ is the number of free parameters to be estimated, N is the number of observations in the data. A lesser BIC value indicates a better model. Note that BIC, more strictly penalizes the adding of parameters than AIC. This implies that when the sample size is adequately large, a complex model (with more parameters) may not always be the greatest one. Even so, as confirmed by specific studies (e.g., Konishi and Kitagawa 2008, Shmueli 2010, and Cavanaugh 2012), BIC is suggested when the primary goal of building the model is to describe the most meaningful factors.

5.0 Results & Summary of Findings

5.1 Model Estimation Results

The gap acceptance model is estimated using a maximum likelihood estimation procedure as described in the previous section. In this section, there is an estimation result of the gap rejection model: a model for US-101 is presented and discussed.

Table 3. *AIC & BIC values for the multinomial logit model*

MNL Model Fitting Information						
		Model Fitting Criteria			Likelihood Ratio Tests	
Site	Time Frame	AIC	BIC	-2 Log Likelihood	Chi-Square	df
US-101	8:05AM – 8:20AM	25397.44	26180.40	25265.44	9565.54	64

5.1.1 Multinomial Logit Model Estimation

In the likelihood ratio tests of the multinomial model, the present model had a chi-square value that is statistically significant at the 95% confidence level. The degree of freedom is 64. The AIC, BIC, & -2 Log Likelihood values for the final model were all less than the intercept models, showing that the model is functional. The Pseudo R-Square values for the multinomial logistic regression model has McFadden’s R-Square values along with Cox & Snell and Nagelkerke. These values have shown a range of variance from 0.9% as the minimum to 27.8% as the maximum. For the classification tables, in the prediction of gap acceptance of the trajectory dataset, the percentage predicted as gap acceptance was around 72.2%.

Table 4. *Multinomial Logit R-Square Values*

MNL Psuedo R-Square				
Site	Time Frame	Cox & Snell R Square	Nagelkerke R Square	McFadden
US-101	8:05AM – 8:20AM	0.009	0.278	0.275

5.2 Summary of Findings

5.2.1 Traffic Density Influence

The downstream density and upstream density of US-101 are 45.6 veh/km/lane and 46.8 veh/km/lane. Therefore, US-101's dataset may not be demonstrative of exceedingly congested conditions.

Table 5. *Traffic Density Influence Table*

Site & Variables	Mean	STD	Median	Minimum	Maximum
US-101					
Average Downstream Density (veh/km/lane)	45.6	15.2	46.7	0	93.3
Average Upstream Density (veh/km/lane)	46.8	15.7	46.7	0	100
Average Velocity (m/sec)	9.8	3	10.3	2.8	17.4

5.3 Gap Acceptance

5.3.1 Gap Acceptance in the Multinomial Logit Models

The parameters of the gap acceptance models are estimated using vehicle trajectory data. Estimation results show that the estimated model is affected by traffic conditions such as average speed in the mainline, interactions with lead and lag vehicles, and urgency of the merge.

US-101 Multinomial Logit Models

In the multinomial logit model of US-101 from 8:05AM to 8:20AM, in the initial gap and nested gap grouping, Vehicle Angle, Vehicle Acceleration, Immediate Distance Travelled, Lane Identification 2, Previous Lane 5, 6, & 7, and Vehicle Class 1 are all statistically significant. The smallest odds being vehicle angle and acceleration. Immediate distance travelled had increasing odds of gap selection at 85.7% per one unit increase and trucks showed to have a 92.8% more likely chance to accept gaps. In the nested gap grouping, Lane identification 3 appeared, showing that vehicles in lane 2 and 3 were more likely to change their lanes. It also showed that vehicles that previously performed mandatory lane changes from lane 5 and 6 were more likely to maneuver again. This may be due to the congested traffic conditions and the will to get downstream. The statistically significant model parameter numbers can be found in the appendices.

What it means for the statistically significant variables in the model to impact traffic flow and How to integrate this method to Traffic Flow, Operations, and Management

On a multi-lane roadway, systematic lane-changes can seriously impact traffic flow and cause capacity reductions. If we refer back to the interference effect of lane changing that I mentioned before, it is more prominent when heavy vehicles change lanes. When the vehicle class of trucks is shown to be 92.8% more likely to accept gaps we start to see the physical effects that heavy vehicles inflict on surrounding traffic. Vehicle velocity was a significant parameter in our model showing that as vehicles decreased their speeds, they would be more likely to change lanes by a factor of 1.473 and 1.498 log-odds. As mentioned before, heavy vehicle drivers normally change into the slower lanes to avoid obstructing the fast moving vehicles that approach from the rear. This causes a slowdown and even bottlenecks in certain areas of the roadway. For US-101's lane-changing area it is nearby a merging junction and is upstream to a diverging junction. Bottlenecks and accidents have a higher probability to happen in this section.

The relationship between immediate distance travelled and density of vehicles on the road have direct relations to each other and traffic flow. As a vehicle increases its immediate distance travelled, it either increases the traffic density or decreases it, depending on the location and surrounding vehicles. Overall, flow denotes vehicles crossing a point or line on the roadway, for example the end of the observable area. Therefore, density and flow refer to diverse measurement bases: density over space at a point in time and flow over time at a point in space. Traffic density is the amount of vehicles per unit of distance of the roadway, whereas traffic flow is the amount of vehicles passing a location point per unit of time. When traffic density increases, traffic flow also rises until it gets to a tipping point afterwards in which the flow commences to dip due to congestion. The tipping point is identified as the critical density. Traffic flow is most influenced by sudden stops, in which many could be avoided with more extensive use of efficient traffic-management strategies. Techniques such as speed harmonization, which comprises of alerting motorists and modulating speed limits in real time to decrease sudden stops that resonate through the roadway, should be used. Speed harmonization can reduce serious crashes and associated traffic congestion outcomes by at least 25 percent. Techniques such as ramp metering can help immensely with speed harmonization by using traffic signals at highway entrance ramps to more evenly insert vehicles into the flow of traffic. In congested traffic, a lane-changing vehicle impacts the traffic flow on both its previous lane and its target lane. These techniques would be highly efficient if implemented to US-101.

5.4 Classes from the Multinomial Logit Model

Applying BIC as the indicator to check if the models are appropriate for defining classes of drivers, the US-101 trajectory dataset had a BIC value of 26180.40. Lower BIC indicates better models. According to the characteristics shown, we have a class from our qualifying models, the US-101 dataset model class will be called the “gap precision class”. In our next section we will seek to define a class and analyze its behaviors.

5.4.1 Defining the Class

“Gap Precision Class”:

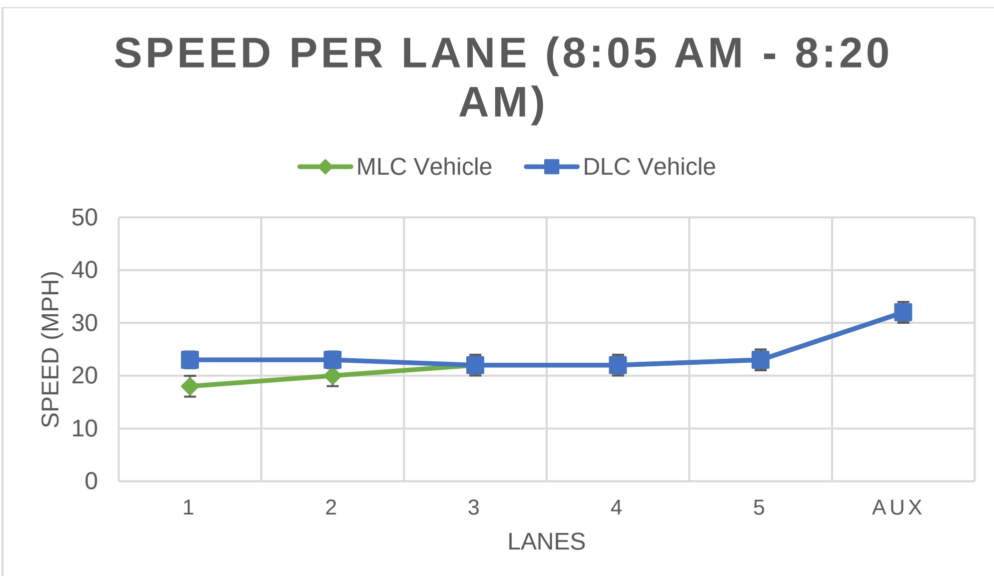
In the multinomial logit model of US-101 for the all timeframes dataset, lane IDs 5 through 7 and global y differences were all statistically significant in late gap acceptance. This shows that outside of MLCs, drivers from this class were more selective with were they decided to maneuver to. Vehicle velocity was a minimal influence to these drivers, barely showing 5% change per unit increase. Immediate global x & y difference was a highly influential parameter in gap acceptance for this class. These drivers would choose to maneuver more downstream in traffic rather than try to maneuver into another lane earlier and take the risk of slowing down. These drivers would accept nested gaps, later down the line. Vehicle angle was consistent amongst these drivers, showing that when they increased their angles, they would most likely merge.

Table 6. *Table of average values for lane-changing drivers at US-101*

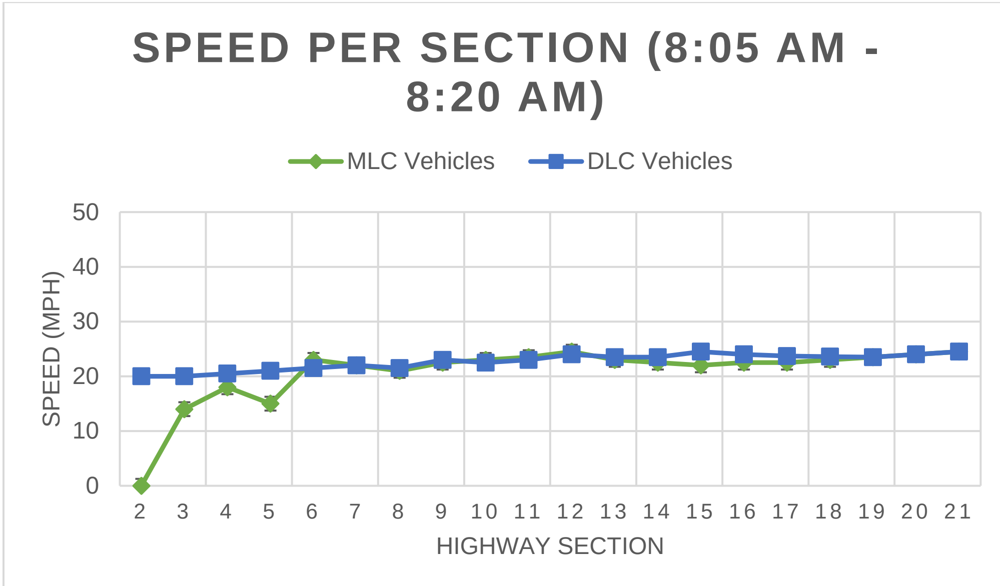
US-101 Mean (Median) Values per Maneuver

Maneuver type	Velocity (ft/s)	Acceleration (ft/s^2)	Local X (ft)	Local Y (ft)	Remaining Distance (ft)	Angle	Distance Travelled (ft)	Headway (sec)	Spacing (ft)
Merge	40.96	-1.05	57.25	911.00	1314.70	-36.27	4.12	1.53	63.58
Diverge	43.86	2.00	57.53	812.60	1412.00	-45.54	4.43	3.05	135.00
Discretionary	45.45	0.37	42.48	1152.20	1072.20	-40.13	4.58	1.26	71.02

Graph 1a. *Speed per lane for MLC & DLC vehicles from the dataset*



Graph 1b. *Speed per Highway section for MLC & DLC vehicles from the dataset*



6.0 Conclusion

6.1 Concluding Statement

Firstly, this study was done to progress the build on gap selection models considering the heterogeneity among drivers' characteristics and merging and diverging strategies. Secondarily, the influence of traffic density was observed, measured, and verified. In this study, different driver characteristics were identified. The objective to prove that the multinomial logit model could produce results to show gap acceptance behavior in drivers' lane-changing was shown. We detected what is a possible class of driving behaviors in the lane-changing process and determined what is to be statistically significant on the 95% confidence interval level. The study uncovered a class that is composed of drivers who possess a distinct preference for certain gaps during certain circumstances. The aim was to operate the study in a way which is consistent with findings in previous studies in literature.

There are many reasons that this research is timely starting with the need to go into a deeper understanding microscopic traffic flow. We can learn different methodologies and understand how they work to achieve many goals on a wider scale. Improving the basic knowledge in this area of research is imperative to attain that goal. This aids with future traffic operations and management. The fact is that every day we are adding more transportation innovations on the common road. That affects future traffic safety in total. This process could be possibly used as a basis for new traffic operations and management techniques. Traffic density has been included in one of these models too. The findings of this study will add to the understanding of driving behaviors in microscopic traffic flow. The model results can help to understand the impact of trucks on traffic flow and provide a theoretical basis for improving road traffic proficiency and safety. It would be interesting to see how connecting a multinomial logit lane-changing model to a multinomial logit zone crash data model for future research. This may be able to aid the improvement of traffic safety, operation, and simulation.

Centered on the suggested methodology, there are innovations in the present research study which are described here:

(1) Construction of a Multinomial Logit model based on the extracted NGSIM trajectory data to describe gap acceptance behavior in lane-changing behavior.

(2) Ways to apply the results of a multinomial lane-changing model to traffic operations and management.

(3) Examined results of the model and determined characterization of the lane-changing behavior.

6.2 Limitations

6.2.1 Why I didn't use the backwards elimination method

Methods such as forward, backward, and stepwise selection are available, but, in logistic as in other regression methods, are not to be recommended. They give incorrect estimates of the standard errors and p-values, can delete variables that are critical to include, and, perhaps most important, allow the researcher not to think (Harrell, 2001). It is much better to compare models based on their results, reasonableness, and fit (as measured, by the AIC & BIC) — note that a lower AIC indicates better fit). A good text on this is Burnham and Anderson (2002).

6.2.2 Why I didn't use the nested logit model or latent class model

SPSS does not currently have a procedure that will easily fit nested logit models. SPSS Statistics also currently does not have a procedure or module designed for latent class analysis. An enhancement request has been filed with SPSS Development.

6.2.3 Problems unifying datasets

The detector datasets were at different intervals. US-101's detector data was recorded at an interval of 5 minutes. Needless to say, it was difficult to unify the detector data with the trajectory data that was recorded at intervals of $1/10^{th}$ of a second. These attempts were saved and recorded for future analysis.

6.3 Recommendations

This study pushed the boundaries of building a complex model which was a good thing and a bad thing at the same time. It was good because it was able to microscopically observe heterogeneity in driver behavior as planned, but it was also bad because of complexities in the ability to process these models. Even though we made an attempt to extract data that well represented discretionary lane-changing, extracted discretionary lane-changing data included some non-discretionary lane-changing. This is because drivers often have explicit, desired traveled lanes and, consequently, change their lanes to reach those lanes. Non-discretionary lane-changing may also occur outside of the mandatory lane-changing area, but at the current moment we can only assume that cars that are switching lanes towards the off-ramp and actually diverge to the off-ramp are performing them. Also there is the fact that different vehicle types may result in different drivers' vision angles. This ought to be verified in future research.

7.0 References

1) Ahammed, M.A., Hassan, Y., Sayed, T.A., 2008. Modeling driver behavior and safety on freeway merging areas. Journal of Transportation Engineering 134 (9), 370–377

2) Ahmed, K.I. & Ben-Akiva, Moshe & Koutsopoulos, Haris & Mishalani, R.G.. (1996). Models of freeway lane changing and gap acceptance behavior. Proceedings of the 13th International Symposium on the Theory of Traffic Flow and Transportation. 501-515.

3) Ahn, S., Vadlamani, S., & Laval, J. (2013). A method to account for non-steady state conditions in measuring traffic hysteresis. Transportation Research Part C: Emerging Technologies, 34, 138–147. doi: 10.1016/j.trc.2011.05.020

4) Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716-723.

5) Alexiadis, V., Colyar, J., Halkias, J., Hranac, R., & McHale, G., 2004. The next generation simulation program. Institute of Transportation Engineers. ITE Journal, 74(8), 22.

6) Analyse-it Software, Ltd. (n.d.). Parameter estimates. Retrieved from https://analyse-it.com/docs/user-guide/fit-model/linear/parameter-estimates.

7) Ardakani, M. K., Yang, J., & Sun, L., 2016. Stimulus response driving behavior: an improved General Motor vehicle-following model. Advances in Transportation Studies, 39, 23-36.

8) Balal, E., Cheu, R. L., Gyan-Sarkodie, T., & Miramontes, J. (2014). Analysis of Discretionary Lane Changing Parameters on Freeways. International Journal of Transportation Science and Technology, 3(3), 277–296. doi: 10.1260/2046-0430.3.3.277

9) Bham, & Hussain, G. (1970, January 1). [PDF] Estimating Driver Mandatory Lane Change Behavior on a Multi lane Freeway: Semantic Scholar. Retrieved from https://www.semanticscholar.org/paper/Estimating-Driver-Mandatory-Lane-Change-Behavior-on-Bham-Hussain/3d324a31b49880e9b5d76da3afa0381e0a9b6245.

10) Bloomberg, L., & J. Dale., 2000. A comparison of the VISSIM and CORSIM traffic simulation models on a congested network. Transportation Research Record 1727, 52-60.

11) Cassidy, M., Rudjanakanoknad, J., 2005. Increasing the capacity of an isolated merge by metering its on-ramp. Transportation Research Part B 39 (10), 896–913.

12) Cheng, S., Xu, Y., Zong, R., & Wang, C. (2019). A fast decision making method for mandatory lane change using kernel extreme learning machine. International Journal of Machine Learning and Cybernetics, 10(12), 3363–3369. doi: 10.1007/s13042-019-00923-8

13) Daamen, W., Hoogendoorn, S. P. & Loot, M., 2010. Empirical analysis of merging behavior at freeway on-ramp. Transportation Research Records: Journal of Transportation Research Board 2188, 108-118.

14) Daganzo, C., Cassidy, M. J., and Bertini R. L. (1999). Possible explanations of phase transitions in highway traffic. Transportation Research Part A, 33(5) 365–379.

15) Deo, N., & Trivedi, M. M. (2018). Multi-Modal Trajectory Prediction of Surrounding Vehicles with Maneuver based LSTMs. 2018 IEEE Intelligent Vehicles Symposium (IV). doi: 10.1109/ivs.2018.8500493

16) Du, R., Peng, Z., & Lu, Q. (2012). Comparison of SMS Calculation Methods Based on NGSIM Data for UAV Detection. 2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics. doi: 10.1109/ihmsc.2012.123

17) Elefteriadou, L., Roess, R.P. & McShane, W.R., 1995. The probabilistic nature of 42 breakdowns at freeway-merge junctions. Transportation Research Record: Journal of the Transportation Research Board 1484, 80-89.

18) Game Theory in Lane-Changing. (2015, September 16). Retrieved from https://blogs.cornell.edu/info2040/2015/09/16/game-theory-in-lane-changing/.

19) Gen, L. & Sun, L. (2018). Characterizing Heterogeneity in Drivers' Merging Maneuvers Using Two-Step Cluster Analysis, Journal of Advanced Transportation, Volume 2018, Article ID 5604375

20) Grace-Martin, K. (2019, March 8). Logistic Regression Models for Multinomial and Ordinal Variables. Retrieved from https://www.theanalysisfactor.com/logistic-regression-models-for-multinomial-and-ordinal-variables/.

21) Grace-Martin, K. (2019, March 11). How to Decide Between Multinomial and Ordinal Logistic Regression Models. Retrieved from https://www.theanalysisfactor.com/decide-between-multinomial-and-ordinal-logistic-regression-models/.

22) Greene, W.H. & Hensher, D.A., (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B 37 (8), 681-698.

23) He, Zhengbing. (2017). Research based on high-fidelity NGSIM vehicle trajectory datasets: A review. 10.13140/RG.2.2.11429.60643/1.

24) Hastie, T., Tibshirani, R., Friedman, J., 2001. The elements of statistical learning. NY Springer.

25) Hidas, P., 2002. Modelling lane changing and merging in microscopic traffic simulation.Transportation Research Part C 10, 351-371.

26) Hoogendoorn, S. P., Bovy, P. H. L. Gas-kinetic modeling and simulation of pedestrian flows. Transpn Res. Rec., 2000, 1710, 28–36.

27) Interstate 80 Freeway Dataset, FHWA-HRT-06-137. (2006, December). Retrieved from https://www.fhwa.dot.gov/publications/research/operations/06137/index.cfm.

28) Jin, W.-L. (2010). A kinematic wave theory of lane-changing traffic flow. Transportation Research Part B: Methodological, 44(8-9), 1001–1021. doi: 10.1016/j.trb.2009.12.014

29) Jin Woo Ro, Partha S. Roop, Avinash Malik, Prakash Ranjitkar, A formal approach for modelling and simulation of human car following behavior, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS

30) Kay, N., Ahuja , S., Cheng, T. N., Vuren, T. van, & MacDonald, M. (2006). Estimation and Simulation Gap Acceptance Behaviour at Congested Roundabouts. Association for European Transport and Contributors, 2006. Retrieved from https://stuff.mit.edu/afs/athena/course/11/11.951/oldstuff/albacete/Other_Documents/Europe Transport Conference/traffic_engineering_an/estimation_and_sim1560.pdf

31) Kim, J.W. & Mahmassani, H.S., 2011. Correlated parameters in driving behavior models: carfollowing example and implications for traffic microsimulation. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board. Washington, DC.

32) Kita, H., 1993. Effects of merging lane length on the merging behavior at expressway onramps. In: Proceedings of the 12th International Symposium on the Theory of Traffic Flow and Transportation.

33) Kita, H., 1999. A merging-giveway interaction model of cars in a merging section: a game theoretic analysis. Transportation Research Part A 33, 305-312.

34) .Knoop, VL., Duret, A., Buisson, C., & van Arem, B. (2010). Lane distribution of traffic near merging zones : Influence of variable speed limits. In s.n. (Ed.), Proceedings 13th international IEEE conference on intelligent transportation systems (pp. 485-490). IEEE Society.

35) Kourkouss, S., Motomura, H., Emura, K., & Ohdachi, E. (2018). Anticipating Driver Actions via Deep Neural Networks and New Driver Personalization Technique through Transfer Learning. Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems. doi: 10.5220/0006669002690276

36) Krajewski, R., Bock, J., Kloeker, L., & Eckstein, L. (2018). The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. 2018 21st International Conference on Intelligent Transportation Systems (ITSC). doi: 10.1109/itsc.2018.8569552

37) Laval, J.A., & Daganzo, C.F. (2006). Lane-Changing in Traffic Streams. Transportation Research Part B-methodological, 40, 251-264.

38) Lee, G., 2006. Modeling Gap Acceptance in Freeway Merges. MSc Thesis, Department of Civil and Environmental Engineering, MIT.

39) Li, C., Chiang, A., Dobler, G., Wang, Y., Xie, K., Ozbay, K., … Wang, D. (2016). Robust vehicle tracking for urban traffic videos at intersections. 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). doi: 10.1109/avss.2016.7738075

40) Li, J., Dai, B., Li, X., Xu, X., & Liu, D. (2019). A Dynamic Bayesian Network for Vehicle Maneuver Prediction in Highway Driving Scenarios: Framework and Verification. Electronics, 8(1), 40. doi: 10.3390/electronics8010040

41) Ma, X. & Andreasson, I., 2007. Statistical analysis of driver behavior data in different regimes of the car-following stage. Transportation Research Records: Journal of Transportation Research Board 2018, 87-96.

42) Mahmassani, H. & Sheffi, Y., 1981. Using Gap Sequences to Estimate Gap Acceptance Functions, Transportation Research Part B 15, 143-148.

43) Mannering, Fred, Jodi Koehne and Soon-Gwan Kim. (1995). Statistical Assesssment of Public Opinion Toward Conversion of General-Purpose Lanes to High-Occupancy Vehicle Lanes. Transportation Research Record #1485 pp. 168-176. National Academy of Sciences.

44) Marczak, F., Daamen, W. & Buisson, C. (2013). Merging behavior: Empirical comparison between two sites and new theory development, Transportation Research Part C 36, 530-546.

45) Mauch, M. and Cassidy M. J. (2002). Freeway traffic oscillations: Observations and predictions. In: M. Taylor (ed.) Transportation and Traffic Theory, Elsevier, pp. 653-673.

46) McCartt, A. T., Northrup, V. S., & Retting, R. A., 2004. Types and characteristics of ramp related motor vehicle crashes on urban interstate roadways in Northern Virginia. Journal of Safety Research, 35, 107–114.

47) Milam, R. T., Choa, F. (1998). A comparison of partial and full partial cloverleaf interchange operations using the CORSIM micro-simulation model. In: ITE District 6 Annual Conference. SanJose, California.

48) Mizanur Rahman, Mashrur Chowdhury, Senior Member, IEEE, Yuanchang Xie, and Yiming He. (2013), Review of Microscopic Lane-Changing Models and Future Research Opportunities, IEEE Transactions on Intelligent Transportation Systems, Vol. 14, No. 4, December 2013

49) Munigety, Caleb & V. Mathew, Tom. (2016). Towards Behavioral Modeling of Drivers in Mixed Traffic Conditions. Transportation in Developing Economies. 2. 10.1007/s40890-016-0012-y.

50) Oppenlander, J. C., & Dawson, R. F., 1970. Interchanges. In Traffic control and roadway elements-their relationship to highway Safety/Revised). Washington, DC: Highway Users Federation for Safety and Mobility.

51) Ossen, S. & Hoogendoorn, S.P., 2005. Car-following behavior analysis from microscopic trajectory data. Transportation Research Record: Journal of the Transportation Research Board 1934, 13-21.

52) Ossen, S., Hoogendoorn, S.P. & Gorte, B.G.H., 2006. Inter driver differences in car-following. A vehicle trajectory-based study. Transportation Research Records: Journal of Transportation Research Board 1965, 121–129.

53) Ossen, S. & Hoogendoorn, S.P., 2011. Heterogeneity in car-following behavior: theory and empirics. Transportation Research Part C 19, 182–195.

54) Pan, Y. & Sun, L. (2012). Characterizing Heterogeneity in Vehicular Traffic Speed Using Twostep Cluster Analysis, Journal of Southeast University, 28(4), 480-484

55) Papadimitriou, S., & Choudhury, C. F. (2017). Transferability of Car-Following Models Between Driving Simulator and Field Traffic. Transportation Research Record: Journal of the Transportation Research Board, 2623(1), 60–72. doi: 10.3141/2623-07

56) Park, M., Jang, K., Lee, J., & Yeo, H. (2015). Logistic regression model for discretionary lane changing under congested traffic. Transportmetrica A Transport Science 11, 333-344.

57) Penn State University. (n.d.). AIC vs. BIC. Retrieved from https://www.methodology.psu.edu/resources/AIC-vs-BIC/.

58) Project Level Analyses. (n.d.). Retrieved from https://www.fhwa.dot.gov/ENVIRONMENT/air_quality/conformity/research/project_level_analyses/pla02.cfm.

59) PTV, 2004. User Manual: VISSIM 4.0.

60) Rao, A., 2006. Modeling Anticipatory Driving Behavior. MSc Thesis, Department of Civil and Environmental Engineering, MIT.

61) Robin, P. (2014). Use on multinomial logistic regression in work zone crash analysis for Missouri work zones analysis for Missouri work zones. Missouri S&T: Scholars' Mine. Retrieved from http://scholarsmine.mst.edu/cgi/viewcontent.cgi?article=8340&context=masters_theses

62) Rpa. (2019, December 18). Reduce highway congestion without adding new lanes. Retrieved from http://fourthplan.org/action/highway-congestion.

63) Sara Moridpour, Majid Sarvi, Geoff Rose, (2010), Lane Changing Models: A Critical Review, Department of Civil Engineering, Monash University, Melbourne, VIC, Australia

64) Sarvi, M., and Kuwahara M., 2007. Microsimulation of freeway ramp merging processes under congested traffic conditions, IEEE Transactions on Intelligent Transportation Systems, 8(3): 470-479

65) Sasoh, A. (2002): Impact of Unsteady Disturbance on Multi-lane Traffic flow. J. Phys. Soc. Japan 71, 989–996

66) Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6(2): 461-464.

67) Shaaban, Khaled & Radwan, Essam. (2005). A Calibration and Validation Procedure for Microscopic Simulation Model: A Case Study of SimTraffic for Arterial Streets. 10.13140/RG.2.1.3804.6565.

68) Sharma, Anshuman & Ali, Yasir & Saifuzzaman, Mohammad & Zheng, Zuduo & Haque, Md. Mazharul. (2018). Human Factors in Modelling Mixed Traffic of Traditional, Connected, and Automated Vehicles. 262-273. 10.1007/978-3-319-60591-3_24.

69) SIAS, 2005. S-Paramics 2005 Reference Manual (Edinburgh: SIAS Ltd).

70) Sopasakis, A., & Katsoulakis, M. (2016). Information metrics for improved traffic model fidelity through sensitivity analysis and data assimilation. Transportation Research Part B: Methodological, 86, 1–18. doi: 10.1016/j.trb.2016.01.003

71) Stephanie. (2017, October 31). Akaike's Information Criterion: Definition, Formulas. Retrieved from https://www.statisticshowto.datasciencecentral.com/akaikes-information-criterion/.

72) Stephanie. (2018, March 17). Bayesian Information Criterion (BIC) / Schwarz Criterion. Retrieved from https://www.statisticshowto.datasciencecentral.com/bayesian-information-criterion/.

73) Taboga, M. (n.d.). Logistic classification model - Maximum likelihood estimation. Retrieved from https://www.statlect.com/fundamentals-of-statistics/logistic-model-maximum-likelihood.

74) Tien Dung Chu, Dr.Eng.; Tomio Miwa; and Takayuki Morikawa, (2017), Discrete Choice Models for Gap Acceptance at Urban Expressway Merge Sections Considering Safety, Road Geometry, and Traffic Conditions, ASCE

75) T.L. Pan , William H.K. Lam , A. Sumalee , R.X. Zhong, (2016), Modeling the impacts of mandatory and discretionary lane-changing maneuvers, Transportation Research Part C 68 403–424

76) Toledo, T, (2003). Integrated Driving Behavior Modeling PhD Thesis, Department of Civil and Environmental Engineering, MIT.

77) Toledo, T., Koutsopoulos, H. N., & Ben-Akiva, M., (2009). Estimation of an integrated driving behavior model. Transportation Research Part C: Emerging Technologies, 17(4), 365-380.

78) U.S. Department of Transportation Intelligent Transportation Systems Joint Program Office, & U.S. Department of Transportation Federal Highway Administration. (2018, August 27). Next Generation Simulation (NGSIM) Vehicle Trajectories and Supporting Data: Department of Transportation - Data Portal. Retrieved from https://data.transportation.gov/Automobiles/Next-Generation-Simulation-NGSIM-Vehicle-Trajector/8ect-6jqj.

79) US Highway 101 Dataset, FHWA-HRT-07-030. (2007, January). Retrieved from https://www.fhwa.dot.gov/publications/research/operations/07030/.

80) Vechione, M., Balal, E., & Cheu, R. L. (2018). Comparisons of mandatory and discretionary lane changing behavior on freeways. International Journal of Transportation Science and Technology, 7(2), 124–136. doi: 10.1016/j.ijtst.2018.02.002

81) Wall G.T. and Hounsell N.B. (2005). Microscopic modeling of motorway diverges, European Journal of Transportation Infrastructure Research, 5, (3), pp. 139-158.

82) Wan, X., Jin, P. J., Yang, F., & Ran, B. (2016). Merging Preparation Behavior of Drivers: How They Choose and Approach Their Merge Positions at a Congested Weaving Area. Journal of Transportation Engineering, 142(9), 05016005. doi: 10.1061/(asce)te.1943-5436.0000864

83) Wan, X., Jin, P. J., Gu, H., Chen, X., & Ran, B. (2017). Modeling Freeway Merging in a Weaving Section as a Sequential Decision-Making Process. Journal of Transportation Engineering, Part A: Systems, 143(5), 05017002. doi: 10.1061/jtepbs.0000048

84) Wang, J., 2005. A simulation model for motorway merging behavior. Transportation and traffic theory 16,

85) Wang, Q., Li, Z., & Li, L., 2014. Investigation of discretionary lane-change characteristics using next-generation simulation datasets. Journal of Intelligent Transportation Systems, 18(3), 246-253.281-301.

86) Wei, Y., Avcı, C., Liu, J., Belezamo, B., Aydın, N., Li, P. T., & Zhou, X. (2017). Dynamic programming-based multi-vehicle longitudinal trajectory optimization with simplified car following models. Transportation Research Part B: Methodological, 106, 102–129. doi: 10.1016/j.trb.2017.10.012

87) Weng, J. & Meng, Q., (2011). Modeling speed–flow relationship and merging behavior at work zone merging areas. Transportation Research Part C 19(6), 985–996.

88) Winnie Daamen, Martijn Loot, and Serge P. Hoogendoorn, (2010) Empirical Analysis of Merging Behavior at Freeway On-Ramp

89) XiaWan, Ph.D., S.M.ASCE; Peter J. Jin, Ph.D.; HaiyanGu; Xiaoxuan Chen, S.M.ASCE; and Bin Ran, Ph.D. (2016), Merging Preparation Behavior of Drivers: How They Choose and Approach Their Merge Positions at a Congested Weaving Area, ASCE, J. Transp. Eng., 142(9): 05016005

90) XiaWan, Ph.D., S.M.ASCE; Peter J. Jin, Ph.D.; HaiyanGu; Xiaoxuan Chen, S.M.ASCE; and Bin Ran, Ph.D. (2017), Modeling Freeway Merging in a Weaving Section as a Sequential Decision-Making Process, ASCE, J. Transp. Eng., Part A: Systems, 143(5): 05017002

91) Xiong, W., Sun, L. & Zhou, J., 2010. Spline-based multi-regime traffic stream models, Journal of Southeast University, 26(1) 122-125.

92) Xu, H., Liu, H., & Gong, H. (2013). Modeling the asymmetry in traffic flow (a): Microscopic approach. Applied Mathematical Modelling, 37(22), 9431–9440. doi: 10.1016/j.apm.2013.04.037

93) Yang, Q., Koutsopoulos, H.N., 1996. A microscopic traffic simulator for evaluation of dynamic traffic management systems. Transportation Research Part C 4, 113–129.

94) Yeo, H., Skabardonis, A., Halkias, J., Colyar, J., & Alexiadis, V. (2008). Oversaturated Freeway Flow Algorithm for Use in Next Generation Simulation. Transportation Research Record: Journal of the Transportation Research Board, 2088(1), 68–79. doi: 10.3141/2088-08

95) Zheng, Z., (2014). Recent developments and research needs in modeling lane changing. Transportation Research Part B 60(1), 16-32.

96) Zheng, L., Jin, P. J., Huang, H., Gao, M., & Ran, B. (2015). A vehicle type-dependent visual imaging model for analysing the heterogeneous car-following dynamics. Transportmetrica B: Transport Dynamics, 4(1), 68–85. doi: 10.1080/21680566.2015.1055618

8.0 Appendices

MNL Classification Tables

Site	Time Frame	Initial Gap Rejection Predicted Correctly	Initial Gap Rejection Predicted as Nested	Initial Gap Rejection Predicted as Gap Acceptance
US-101	8:05AM – 8:20AM	1532	0	0

Site	Time Frame	Nested Gap Rejection Predicted Correctly	Nested Gap Rejection Predicted as Initial Gap Rejection	Nested Gap Rejection Predicted as Gap Acceptance
US-101	8:05AM – 8:20AM	1046260	0	73

Site	Time Frame	Gap Acceptance Predicted Correctly	Gap Acceptance Predicted as Initial Gap Rejection	Gap Acceptance Predicted as Nested Gap Rejection
US-101	8:05AM – 8:20AM	512	10	187

Site	Time Frame	Initial Gap Rejection Percentage Correct	Nested Gap Rejection Percentage Correct	Gap Acceptance Percentage Correct
US-101	8:05AM – 8:20AM	100.00	99.99	72.20

Statistically Significant Model Estimation Parameters

Multinomial Logit Model

MNL US-101 8:05 AM - 8:20 AM Variables in the Equation

Variable	B	S.E.	Wald	df	Sig.	Exp(B)
Initial Gap Grouping						
Vehicle Angle	-0.014	0.004	12.171	1	0.000	0.986
Vehicle Acceleration	-0.024	0.009	6.372	1	0.012	0.976
Distance Travelled	-1.942	0.277	49.314	1	0.000	0.143
[Lane Identification=2]	-46.399	19.641	5.581	1	0.018	7.07E-21
[Lane Identification=2]	-46.399	19.641	5.581	1	0.018	7.07E-21
[Previous Lane=5]	-42.541	12.346	11.872	1	0.001	3.35E-19
[Previous Lane=6]	-82.590	14.413	32.835	1	0.000	1.35E-36
[Previous Lane=7]	-173.279	12.362	196.489	1	0.000	5.57E-76
[Vehicle Class=1]	-2.631	0.827	10.112	1	0.001	0.072
Vehicle Velocity	0.387	0.055	48.826	1	0.000	1.473
Global Y Diff	2.997	0.320	87.735	1	0.000	20.017

[Lane Identification=5]	47.199	12.366	14.568	1	0.000	3.149E+20
[Lane Identification=6]	85.787	14.426	35.365	1	0.000	1.81E+37
[Lane Identification=7]	173.949	12.369	197.782	1	0.000	3.51E+75
[Previous Lane=2]	50.991	19.612	6.760	1	0.009	1.396E+22
			Nested Gap Grouping			
Vehicle Angle	-0.015	0.002	53.712	1	0.000	0.985
Vehicle Acceleration	-0.025	0.007	14.117	1	0.000	0.975
Distance Travelled	-2.048	0.455	20.271	1	0.000	0.129
[Lane Identification=2]	-62.486	0.835	5594.467	1	0.000	7.29E-28
[Lane Identification=3]	-34.118	0.649	2766.744	1	0.000	1.52E-15
[Previous Lane=5]	-28.551	0.000	.	1	.	3.99E-13
[Previous Lane=6]	-69.591	0.000	.	1	.	5.99E-31
[Previous Lane=7]	-153.826	0.000	.	1	.	1.56E-67
[Vehicle Class=1]	-2.637	0.467	31.823	1	0.000	0.072
Remaining Distance	0.078	0.035	4.961	1	0.026	1.081
Local Y	0.078	0.035	4.925	1	0.026	1.081
Vehicle Velocity	0.404	0.019	448.546	1	0.000	1.498
Global Y Diff	3.096	0.354	76.605	1	0.000	22.118
[Lane Identification=5]	33.858	0.296	13076.071	1	0.000	5.063E+14
[Lane Identification=6]	73.373	0.200	134666.994	1	0.000	7.337E+31
[Lane Identification=7]	154.869	0.151	1048163.602	1	0.000	1.82E+67
[Previous Lane=2]	67.691	0.000	.	1	.	2.499E+29
[Previous Lane=3]	39.380	0.000	.	1	.	1.266E+17

[Previous Lane=4]	5.559	0.000	.	1	.	259.544