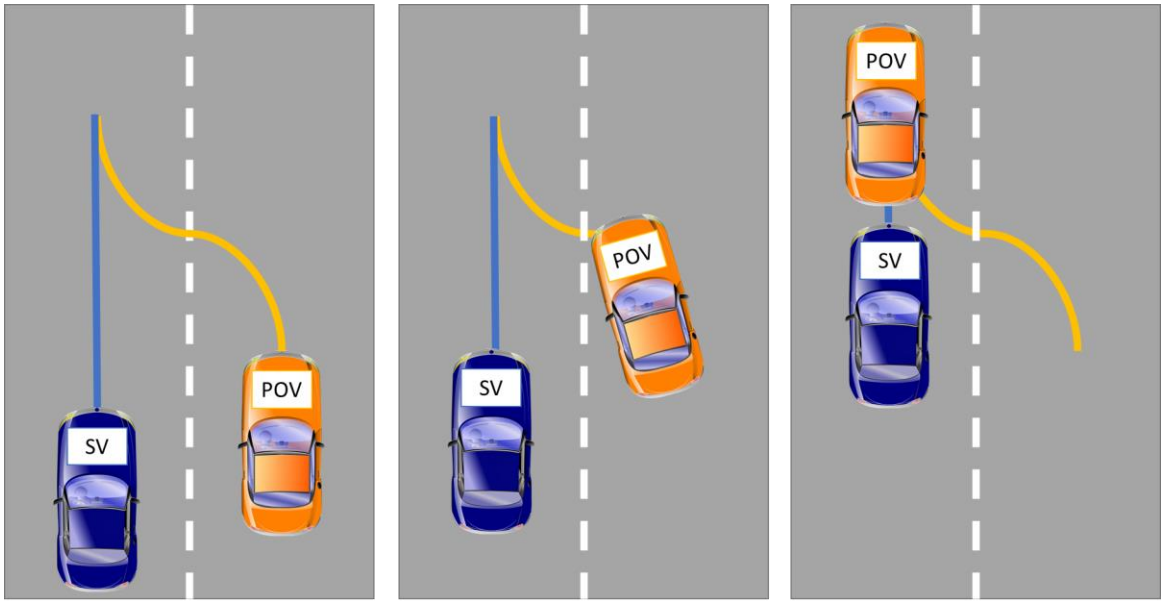
CHALMERS
UNIVERSITY OF TECHNOLOGY



Statistical modelling of critical cut-in trajectories based on naturalistic driving data

Master's thesis in Automotive Engineering

Sainath Kishore Josyula

DEPARTMENT OF Mechanics and Maritime Sciences

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2021
www.chalmers.se

MASTER'S THESIS IN AUTOMOTIVE ENGINEERING

Statistical modelling of critical cut-in trajectories based on naturalistic driving data

Sainath Kishore Josyula

Department of Mechanics and Maritime Sciences
Division of Vehicle Safety
Unit of Crash Analysis and Prevention
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2021

Statistical modelling of critical cut-in trajectories based on naturalistic driving data

Sainath Kishore Josyula

Master's Thesis 2021:37
Department of Mechanics and Maritime Sciences
Division of Vehicle Safety
Unit of Crash Analysis and Prevention
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone: + 46 (0)31-772 1000

Cover:
Critical cut in manoeuvre starting from right side. By the end of the manoeuvre (Right most figure), the gap between SV and POV is small, causing a sudden evasive manoeuvre of the SV (Figure 2 in Introduction).

Department of Mechanics and Maritime Sciences
Göteborg, Sweden 2021-09-30

Statistical modelling of critical cut-in trajectories based on naturalistic driving data

Master's thesis in Automotive Engineering

Sainath Kishore Josyula
Department of Mechanics and Maritime Sciences
Division of Vehicle Safety
Unit of Crash Analysis and Prevention
Chalmers University of Technology

Abstract

Cut-in maneuvers are events when a vehicle changes lane and moves close to another vehicle in the adjacent lane. This phenomenon is quite common on highways and has adverse impact on traffic safety. Statistics from the Fatality Analysis Reporting System (FARS) by National Highway Traffic Safety Administration (NHTSA) show that there have been more than 290,000 traffic crash injuries associated with cut-in maneuvers (including rear-end, angle or sideswipe collision) between years 2015 and 2019. Active safety systems and autonomous vehicles are being developed to achieve safe driving and should be able to detect potentially dangerous scenarios like critical cut-ins, and act to avoid them. Statistical models of cut-in scenario trajectories are useful for developing and evaluating both active safety systems and autonomous vehicles. This thesis aims to increase our understand of and model cut-in trajectories of vehicles performing cut-in maneuvers, using SHRP2 (The Second Strategic Highway Research Program) naturalistic driving data. To conduct the study, the SHRP2 event data has been manually categorized. Thereafter, a dataset of kinematic variables, which were extracted using a video annotation tool, has been prepared to enable studying the trajectories in detail. Specifically, this study uses a quintic polynomial of time to model lateral and longitudinal trajectories of the vehicle that cuts in (or principle another vehicle; POV). One of the required inputs is the event duration which is calculated by identifying maneuver start and end times . The event durations follow a normal distribution and range from 2.1 to 6.4 seconds. Then, two linear models of event duration are built to calculate the remaining two variables required for a polynomial trajectory model, which are initial POV lateral acceleration and final POV longitudinal position. Using these three variables a range of trajectories have been generated, representing SHRP2 naturalistic driving for right to left single lane change cut-ins. This thesis also uses a probabilistic regression model to calculate the distribution of parameters of quintic polynomial of lateral position of POV. Using this model, new (other than training data) trajectories have been generated. Both these generative models of trajectories can be used as one of the inputs to simulations used in the design and evaluation of active safety systems and automated driving systems.

Key words:

Autonomous Driving, Cut-in maneuvers, Trajectory models, Naturalistic driving data, Probabilistic regression.

Contents

Preface

This study is a master thesis work carried out at the Department of Mechanics and Maritime Sciences, Division of Vehicle Safety, Unit of Crash Analysis and Prevention, Chalmers University of Technology, Sweden. The work was performed during the MPAUT master's program. The work was developed from January to September 2021 in the Vehicle and Traffic Safety Centre at Chalmers - SAFER, in Göteborg.

Acknowledgement

First, I would like to express my gratitude to the supervisor of the project Giulio Bianchi Piccinini and to the examiner Jonas Bärgman, for your constant willingness to help and to support me in any situation. I would also like to thank everyone working in SAFER, for providing a supportive environment and resources to conduct this thesis.

I am very grateful to my partner and to my parents, that constantly supported me in this experience far from home and having trust in me in everything I was doing. Finally, thanks to all my friends, the ones I met in Sweden from all over the world for making this a joyful experience.

This work has been performed under a VTTI SHRP2 Data License Agreement. The findings and conclusions of this report are those of the authors and do not necessarily represent the views of VTTI, the Transportation Research Board, or the National Academies. The data used falls within the license agreement SHRP2-DUL-A-16-204 and the data has the DOI:10.15787/VTT1K013.

Göteborg September 2021-09-30

Sainath Kishore Josyula

Notations

Abbreviations

ISS	Intelligent Safety Systems
AD	Autonomous driving
ADAS	Advance Driver Assistance Systems
AEB	Autonomous Emergency Braking
NDD	Naturalistic Driving Data

List of Figures

1 Introduction

From the year 2000 to 2016 the road traffic death rate marginally decreased from 18.8 to 18.2 death per 100,000 population, while the number of deaths in absolute terms increased from 1.15 to 1.35 million (World Health Organization, 2018). Road injuries are among the top 10 leading causes of death in lower-middle and upper-middle income countries (WHO, 2019). Road safety is a global issue and hence it has been considered as an important part of global sustainable development goal (Stockholm Declaration, 2020). Because of its impact, many nations have been trying to improve road safety for a long time. For example, the vision zero concept, introduced in 1995 in Sweden, is aimed at reducing the road safety by promoting a holistic view on road safety. This view shifts the focus of research related to improve road safety from "accident prevention to the goal that no one should die or be severely injured in traffic" (Trafikverket, 2020). One of the major contributors for road injuries are human factors which contributed to up to 90% of road crashes (Pakgohar et al., 2011). Intelligent Safety Systems (ISS) are developed to reduce human error in driving activities and thus reduce road crashes. ISS are systems which have the ability to sense the vehicle's environment and either warn the driver or control the vehicle in safety critical situations (Hannan et al., 2010).

Bärgman et al. (2017) points out that, during last few decades, the focus of the automotive industry has shifted from development of passive safety systems to development of ISS where ISS include both Autonomous Driving (AD) and Advance Driver Assistance Systems (ADAS). Nowadays, there is a continuous development of such ISS systems with an aim to increase the safety benefit of the system for all road users and achieve a system capable of handling the entire driving task on its own (NHTSA, 2021).

The Society of Automotive Engineers (SAE) standard J3016_202104 (SAE International, 2021) provides the classification and definitions for terms related to motor vehicle automation systems. According to this standard there are six levels of automation as shown in Figure 1, starting from no automation (Level 0) to full driving automation (Level 5).

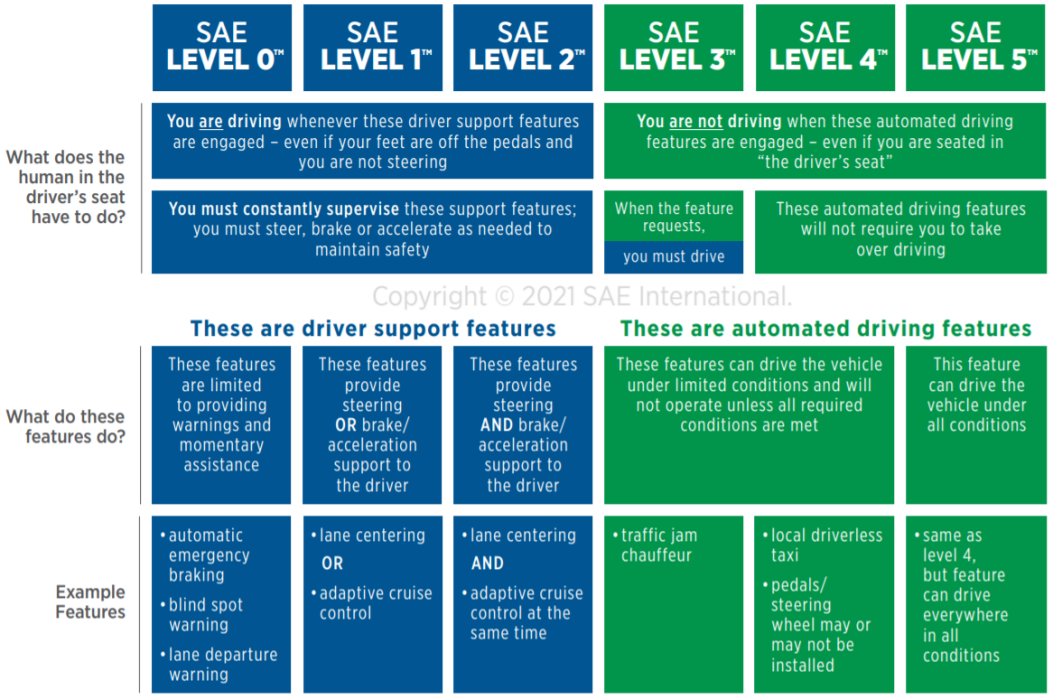


Figure 1: Levels of automation defined by SAE J3016 (SAE International, 2021)

According to this standard, crash mitigation and avoidance capability are part of automated driving functionality for levels 3 to 5. Apart from this, levels 0-2 also have the functions of assisting human driver through warnings or momentary assistance when the system detects a safety critical scenario in the driving environment. Automated driving can be divided into two sub domains: perception and localization, and planning and control (Murthy & Bharadwaj, 2020). The perception and localization domain senses the vehicle environment and identifies the location of vehicle in the environment. The planning and control domain plans the optimal path for the vehicle in the given environment and provides necessary control inputs to the vehicle.

The systems of automated driving, which replace the human driving shall identify potential crash scenarios while driving and take necessary action to prevent fatality or serious damage (Jin et al., 2020). Also, given the current lifespan of vehicles, autonomous vehicles must interact with manually driven vehicles for decades after their introduction (Shams El Din 2020, Zhao et al., 2017). As the level of automation increases, the AD systems have to deal with many uncertainties in real world including imperfect human driver behaviors (Zhao, et al., 2017). This makes understanding the driver behavior in manual vehicles an important factor for development of AD/ADAS systems.

Feng et al. (2019) reviewed the following approaches of designing test scenarios for evaluating ADAS systems, including their pros and cons. These five in the sequence of evolution are Naturalistic Field Operational Testing (N-FOTs), Monte Carlo Simulations (MCS), Accelerated evaluation methods (AE) and Test Matrix approach (TM). N-FOTs evaluated the system through real world driving but had low level of exposure to safety critical crashes/near-crashes. MCS has been used to develop driver models using stochastic/random parameters based on data collected through naturalistic studies and generate scenarios using these driver models. For example,

in Yang et al. (2008) a car following model was developed with driver distraction and lead vehicle deceleration as stochastic/random processes which produced scenarios similar to the field testing data. AE methods were proposed by Zhao, et al. (2016) to speed up the evaluation process of active safety systems by generating critical scenarios for simulation and validation of the systems. TM methods evaluate the systems by collecting factors such as positions, speeds and accelerations of the system being evaluated and generating testing scenarios using them. An example for the use of TM is the EuroNCAP standard for AEB evaluation where a TM with different configurations of the test vehicle and obstacle was presented. Feng et al. (2019) identified that one of the common limitations of these methods was lower availability of safety critial scenario data compared to non critical driving data in the available data sources like naturalistic driving data.

To overcome the limited availability of critical scenarios extensive research is being done using naturalistic driving data (NDD). NDD is a data collection method that provides information about the normal driving behavior of people on roads. This data is collected by equipping vehicles with data acquisition systems which have sensors like cameras, RADARs to continuously collect data about the driver, vehicle, and the surrounding environment. This type of data has the potential to add significant value to understanding crash causation, providing a large database for analyzing crash and crash related behaviors (Schagen & Sagberg, 2012). This thesis uses second Strategic Highway Research Program (SHRP2) naturalistic driving data and the L3Pilot classification scheme for different types of cut-in maneuvers. SHRP2 is the world's largest naturalistic database, collected by Virginia Tech Transportation Institute (VTTI) with more than 3000 participants, for a duration of two years, in the United States. The dataset consists of 8769 crash and near-crash events. According to SHRP2 data definitions, crash and near-crash are defined as "Any contact that the subject vehicle has with an object, either moving or fixed, at any speed in which kinetic energy is measurably transferred or dissipated.", and "Any circumstance that requires a rapid evasive maneuver by the subject vehicle or any other vehicle, pedestrian, cyclist, or animal to avoid a crash", respectively (Transportation Research Board of the National Academies of Science, 2013). L3Pilot is a large-scale pilot study of SAE level 3 and level 4 automated driving functions. It started in 2017 and is expected to finish by the end of 2021.

Different types of critical scenarios are observed in NDD analysis which lead to crashes and near-crashes, and the focus of this thesis is on one of such scenarios which occurs due to a vehicle performing dangerous lane changing maneuver. According to Lee et al. (2016), lane changing causes negative shockwaves on roads. That is one car brakes and so other cars should subsequently brake which leads to traffic safety issues. In the U.S., cut-ins account for approximately 5% of all crash fatalities (Hou et al., 2015). According to statistics of SHRP2 data from the SHRP2insight website (Virginia Tech Transportation Institute, 2020), there are 1426 critical events (crashes and near crashes) out of 8769 events which were caused due to another vehicle changing lane. Critical cut-ins are maneuvers where a vehicle (Principal Other Vehicle, POV) changes lane and moves very close to the vehicle (Subject Vehicle, SV) in the adjacent lane (Wang et al., 2019) causing the SV to

perform a sudden evasive maneuver like braking or steering, or both. This is shown in Figure 2.

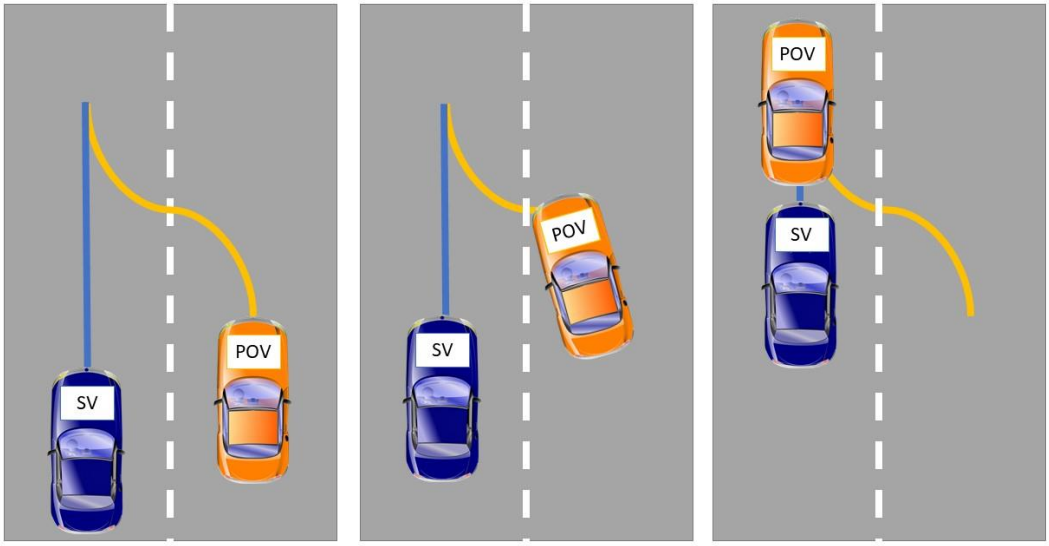


Figure 2: Critical cut in manoeuvre starting from right side. By the end of the manoeuvre (Right most figure), the gap between SV and POV is small, causing a sudden evasive manoeuvre of the SV.

Yang et al. (1996) classified lane changes as either Mandatory Lane Change (MLC) or Discretionary Lane Change (DLC). In a mandatory lane change, the driver executed a compulsory lane change to stay on the desired path – for example, taking the exit ramp to exit highway. In a discretionary lane change the driver execute a voluntary lane change to maintain the desired driving conditions. According to SHRP2 there were crashes and near-crashes due to both MLC and DLC lane changes.

Lane changing models were extensively researched during the last three decades (Gipps, 1986; Toledo et al., 2007). In Toledo et al. (2003), lane change maneuvers were modelled in two phases:

1. Lane change decision phase
2. Lane change execution phase

In the decision phase the motivating factors (traffic conditions, driver intensions, etc.) for the driver to perform a lane change were modelled. In the execution phase the kinematics of the vehicle movement were modelled. Different types of lane change decision models were compared and suggested that for future development of lane change models' large datasets of driver lane changing trajectories are required (Moridpour et al., 2010). Bernard et al. (2012) also highlighted the importance of trajectory models in the context of improving road safety by stating that trajectory models were required to understand the interactions between vehicles, drivers and road infrastructure in detail, which in turn enable development of warning and intervention functions of ADAS systems.

According to Bernard et al. (2012) a trajectory of vehicle is the representation of the position, velocity, and accelerations of the vehicle as functions of time. Models of cut-in trajectories are important as they represent the critical lane changing events

and are useful for overcoming the limited availability of critical scenario data necessary for evaluation of AD/ADAS systems that were mentioned by Feng et al. (2019). An example of a cut-in trajectory is shown in Figure 2, represented as blue and orange curves. Three types of modelling approaches have identified for lane change trajectories and are shown in Figure 3.

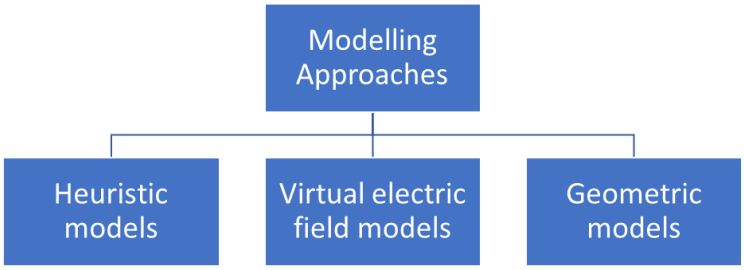


Figure 3: Three modelling approaches for lane changing trajectories found in literature review

Heuristic methods are typically based on Artificial Intelligence (AI) logic which use techniques like random sampling and search-based methods (Zhang et al., 2018, McNaughton et al., 2011). These methods have been used to solve path planning problems for autonomous driving applications. In models based on virtual electric field the vehicle movement is attributed to a force which is defined as the negative gradient of a virtual electric field. The minimum electric field potential between the vehicle and obstacle objects are considered as a basis for collision avoidance and path optimization problems. According to Peng, et al. (2020), this type of models have high computational complexity. Geometric methods generate trajectories based on parametric geometric curves such as splines, clothoids, polynomials, etc. According to Bai et al. (2017) and Wang et al. (2014), polynomial curves are simple, have continuous curvature and have better real time performance in case of AD applications.

This thesis focuses on a type of lane change maneuver known as critical cut-in. The goal of this thesis is to build statistical models for the POV trajectory in cut-in maneuvers by using the variables extracted using video annotation tool developed during last year thesis work (Shams El Din, 2020) and also by creating some classification variables, and aims to develop a generic trajectory model for a group of classified maneuvers that can be used to interpolate between and extrapolate from available naturalistic driving data from SHRP2. The dataset derived from SHRP2 by (Shams El Din, 2020) is used for this thesis. The details of the dataset, heuristics used to filter the required type of trajectories and variable extraction is explained in Section 2.

2 Method

This section has two parts. The first part explains the variables extracted from SHRP2 and new annotations done in this thesis, for classification of cut-ins resulting in a specific group of trajectories for further analysis. The second part describes two different trajectory models aimed at enabling interpolation and extrapolation of SHRP2 trajectories for a selected group of data.

2.1 Data

A total of 57 cut-in events' annotation data were used in this thesis. The events in this dataset are a mixture of different trajectories with respect to different POV speeds, different direction of lane change (right to left and left to right), different number of lane changes performed by the POV in the event, and they include Mandatory (MLC) and Discretionary (DLC) lane changes. This dataset consists of 75% DLC type lane changes and 25% of MLC type lane changes.

There are two types of variables associated with this data: kinematic variables and classification variables. The kinematic variables representing POV cut-in trajectories were extracted using a video annotation tool, developed by (Shams El Din, 2020), and manual annotation which is explained in Section 2.1.1. The kinematic variables obtained from the annotation tool are listed in Table 1 and other kinematic variables were provided by VTTI. The classification variables used for grouping the trajectories are derived through manual annotation and are explained in Section 2.1.2.

2.1.1 Kinematic variables

To process the video data from SHRP2 and extract kinematic variables of vehicles involved in cut-in scenarios, a python based video annotation tool was developed in a previous project by (Shams El Din, 2020). The User Interface (UI) of the tool enables the user to perform the following operations:

1. Select a video file along with the related data files
2. Identify and mark the POV doing the cut-in maneuver
3. Mark the lane markings

These markings are done for few frames in the video and based on these markings, kinematic variables of POV—position, velocities in x, y coordinates—are calculated as the output and the position output of the tool is aligned with the radar data to correct the annotation error as explained in (Shams El Din, 2020). The tool used the following data sources of SHRP2 to calculate the kinematic variables:

1. Video data Front camera – Data acquisition system
2. RADAR data – Data acquisition system
3. Speed and accelerations of SV – CAN data

All these variables were calculated in the coordinate system of Subject Vehicle (SV) and then transformed into the coordinate system of the lane center of the target lane. The final coordinate system is shown in Figure 4 and the list of variables are shown in Table 1.

Table 1: List of variables extracted from SHRP2 video data using video annotation tool.

Notation	Unit	Description	Source
$R_{Y,pov}$	m	Lateral range to POV	Available from tool output
$R_{X,pov}$	m	Longitudinal range to POV	Available from tool output
$Y_{LL,sv}$	m	Lateral distance from SV center to left lane	Available from tool output
$Y_{RL,sv}$	m	Lateral distance from SV center to right lane	Available from tool output
$Y_{LC,sv}$	m	SV lateral offset from lane center	Calculated using $Y_{LL,sv}$ and $Y_{RL,sv}$
$Y_{LC,pov}$	m	POV lateral offset from lane center	Calculated from $Y_{LC,sv}$ and $R_{Y,pov}$
$V_{X,sv}$	m/s	Longitudinal SV speed	Available from time-series data
$V_{Y,sv}$	m/s	Lateral SV speed in lane	Calculated using $\dot{Y}_{LC,sv}$
$S_{X,sv}$	m	Distance travelled by SV since start of event	Calculated using $\int V_{x,sv}$
$V_{X,pov}$	m/s	Longitudinal POV speed	Calculated using $\dot{S}_{X,pov}$
$V_{Y,pov}$	m/s	Lateral POV speed in lane	Calculated using $\dot{Y}_{LC,pov}$
$S_{X,pov}$	m	Distance travelled by POV since start of event	Calculated using $S_{X,sv}$ and $R_{X,pov}$

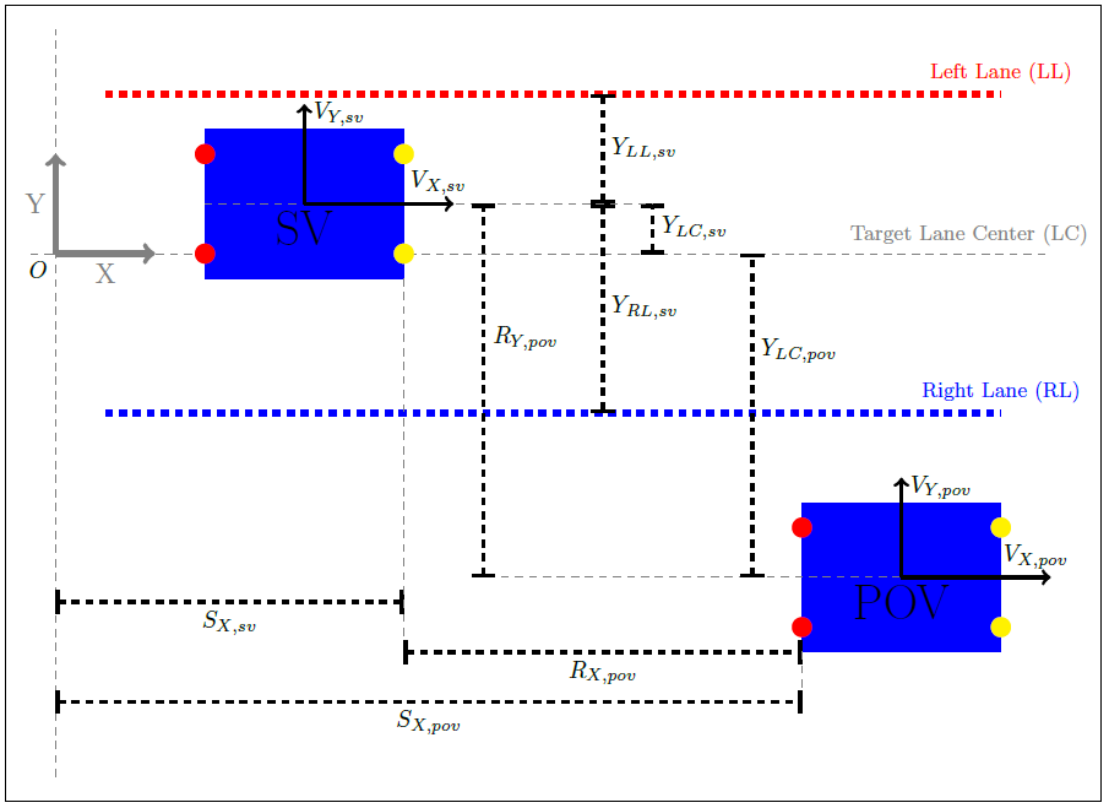


Figure 4: Reference frame with respect to target lane centre, shown in grey on top left of image (Shams El Din, 2020)

The trajectory of the POV for one event built using the extracted kinematic variable data is shown in Figure 5.

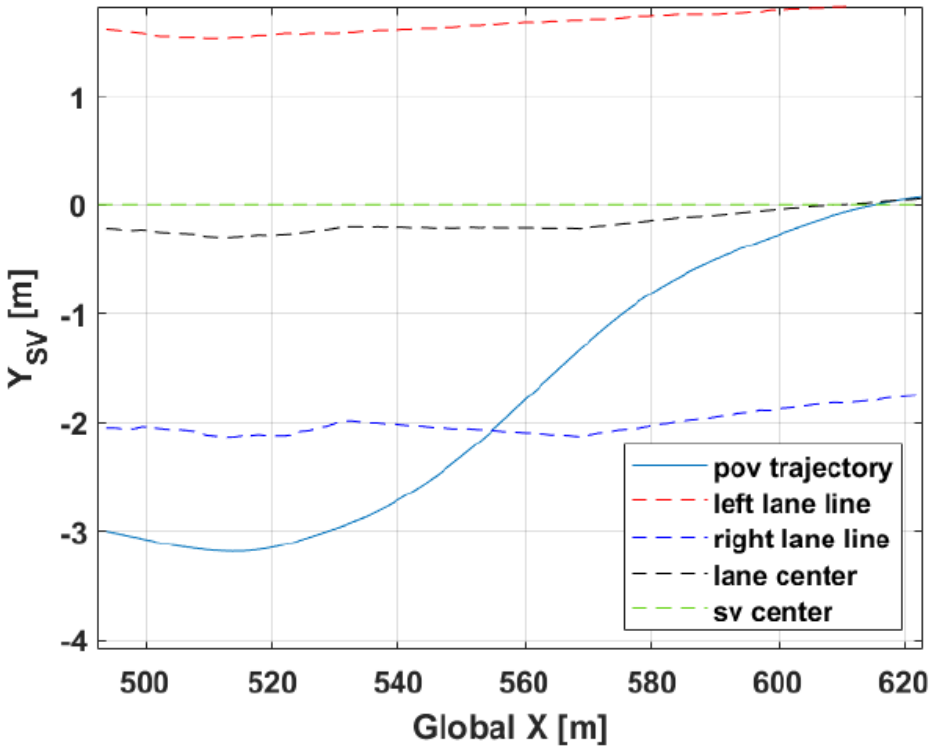


Figure 5: POV trajectory of one event built using the extracted kinematic variable data of the video annotation tool.

2.1.2 Classification variables

Apart from the kinematic variables of the event trajectories, classification of the events was necessary to identify maneuvers with similar characteristics. The idea behind such a classification is to build a generic trajectory model for each category of cut-in maneuver. The list of variables used for classification are summarized in Table 2. As seen in the table some of these variables are based on L3Pilot classification scheme (Hibberd, et al., 2018) and the others were identified by the author. All these variables are manually observed by watching the front video data of SHRP2 dataset.

Table 2: List of variables used for classifying annotation data into categories.

Variable Name	Reference
Cut-in motivation	L3Pilot
Starting side of cut-in	L3Pilot
Type of lane change	Author
Total number of lane changes	Author
Leading vehicle presence on target lane	L3Pilot
Leading vehicle – target lane influence	Author
Leading vehicle presence in origin lane	Author
Leading vehicle – origin lane influence	Author

The definitions of the cut-in classifications defined by L3Pilot project are explained below:

2.1.2.1 **L3Pilot classification variables**

- Cut-in motivation: This variable describes the motivation of POV driver to perform the cut-in maneuver.
  - Cut-in after overtaking the SV (example in Figure 6, Figure 7)
  - Overtake a slower moving vehicle (example in Figure 8)
  - Entry ramp - Entering from another road (before the cut-in, the vehicle was on another roadway) (example in Figure 9)
  - Exit ramp - Exiting the road (after the cut-in the vehicle will exit the roadway) (example in Figure 10)
  - Avoiding a work zone or other obstacle (static) (example in Figure 11)
  - Approaching an intersection in situations where the vehicle must change lanes to be in the correct lane (example in Figure 12)
  - Ending lane (the origin lane on which the cut-in vehicle was traveling is ending) (example in Figure 13)
  - Other (other motivation that is not included in the above categories).
- Starting Side of Cut-in:
  - Cut-in from left lane
  - Cut-in from right lane.
- Leading vehicle presence on target lane:
  - Yes
  - No.

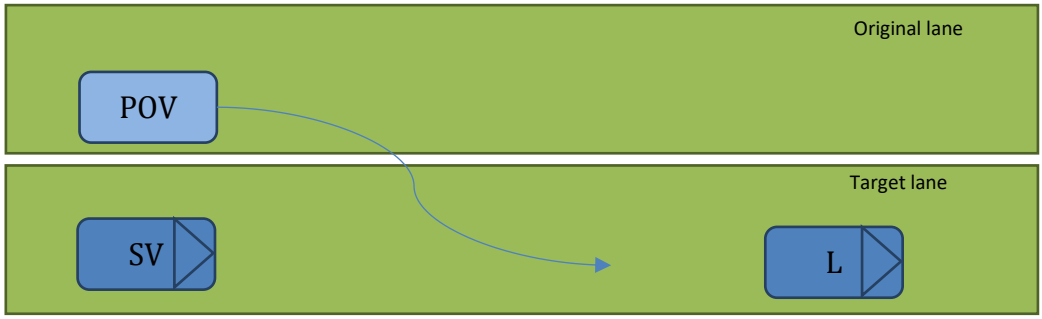


Figure 6: Cut-in after overtaking the ego vehicle. Single lane cut-in. Cut-in from left. Principal Other Vehicle (POV), Subject Vehicle (SV) and leading (L) vehicle. Leading vehicle can be present or not.

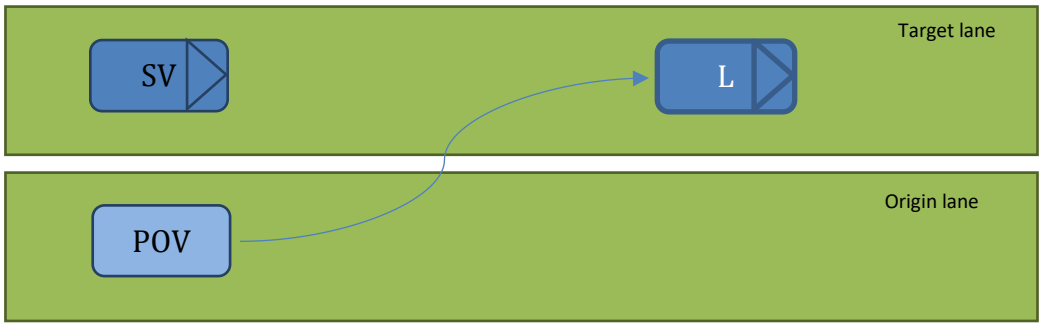


Figure 7: Cut-in after overtaking the ego vehicle. Single lane cut-in. Cut-in from right. Principal Other Vehicle (POV), Subject Vehicle (SV) and leading (L) vehicle. Leading vehicle can be present or not.

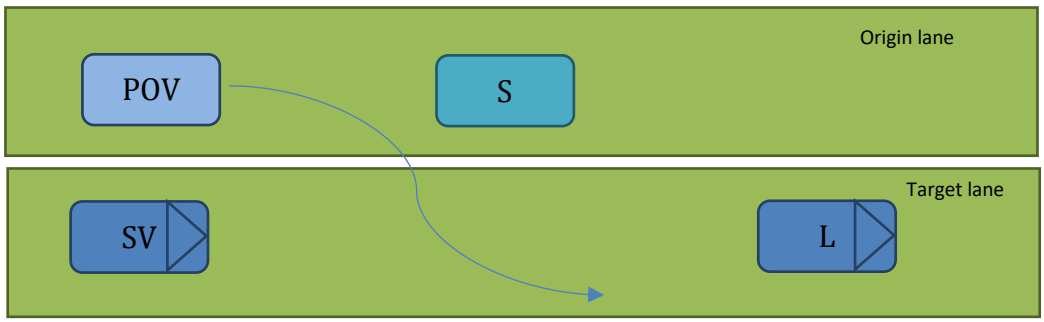


Figure 8: Overtake a slower moving vehicle. Principal Other Vehicle (POV), Subject Vehicle (SV), slower vehicle (S). Cut-in from left. Single lane cut-in. Leading vehicle can be present or not.

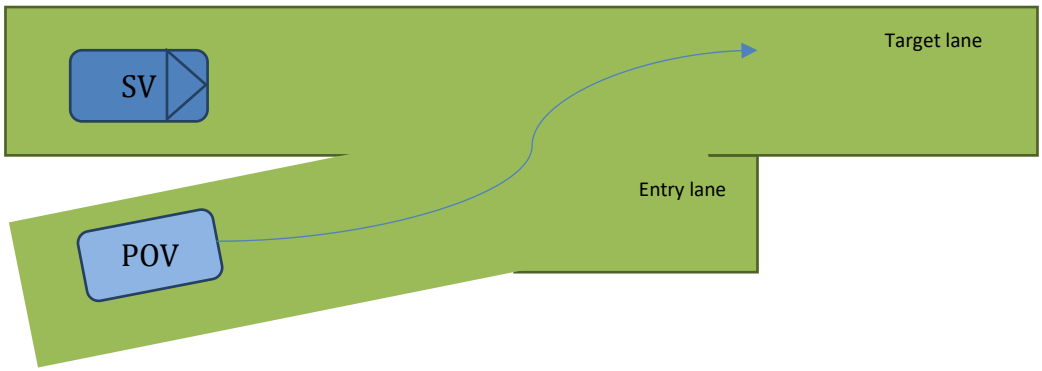


Figure 9: Entry ramp. Single lane cut-in. Cut-in from right. Principal Other Vehicle (POV), Subject Vehicle (SV). Leading vehicle can be present or not.

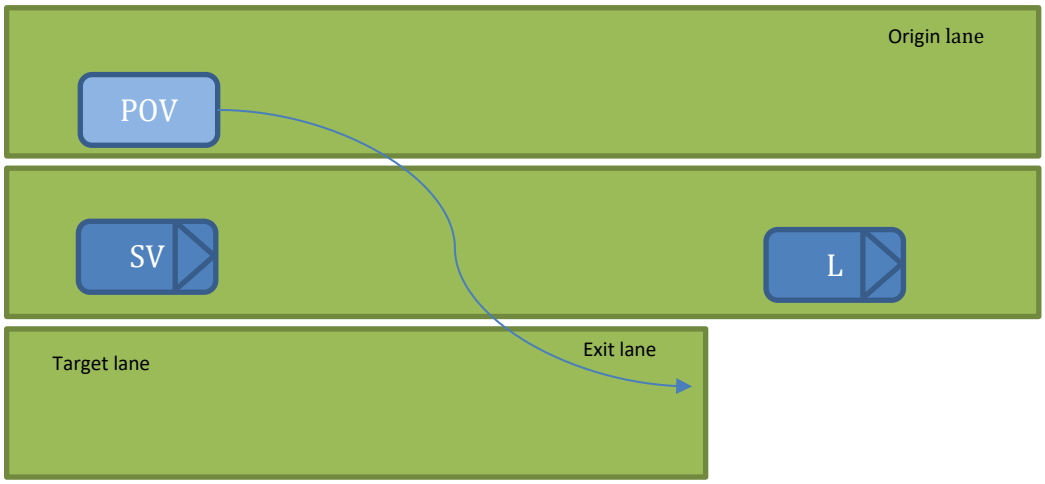


Figure 10: Exit ramp. Cut-in from left. Multiple lane cut-in. Principal Other Vehicle (POV), Subject Vehicle (SV). Leading vehicle can be present or not.

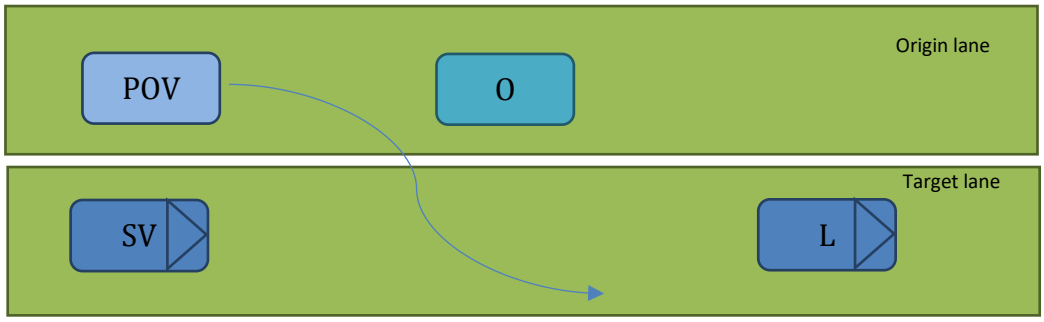


Figure 11: Avoiding a work zone or other obstacle. Principal Other Vehicle (POV)*, Subject Vehicle (SV), obstacle (O). Cut-in from left. Single lane cut-in. Leading vehicle can be present or not.*

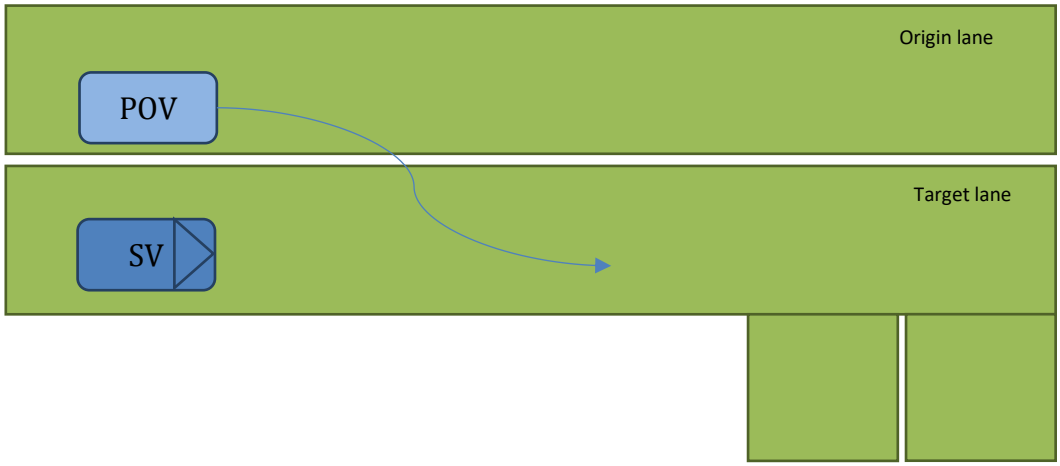


Figure 12: Approaching an intersection (cut-in vehicle plans to turn right at the intersection). Single lane cut-in. Cut-in from left. Principal Other Vehicle (POV), Subject Vehicle (SV) and leading (L) vehicle. Leading vehicle can be present or not.

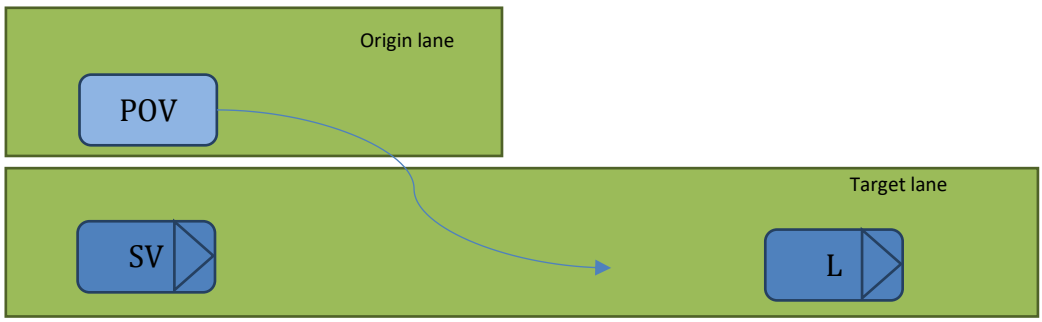


Figure 13: Ending lane. Single lane cut-in. Cut-in from left. Principal Other Vehicle (POV), Subject Vehicle (SV) and leading (L) vehicle. Leading vehicle can be present or not.

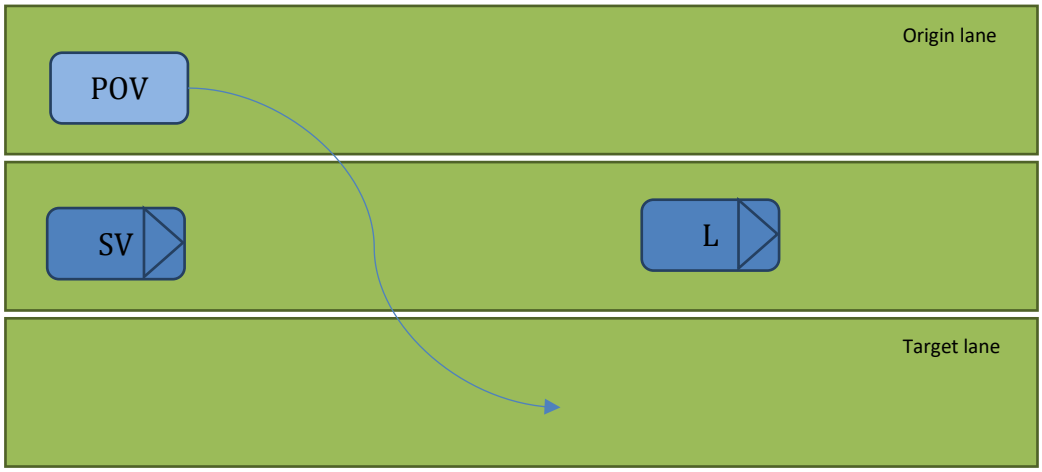


Figure 14: Multiple lane cut-in. Principal Other Vehicle (POV), Subject Vehicle (SV) and leading (L) vehicle. Cut-in from left. Leading vehicle can be present or not.

2.1.2.2 **Author defined classification variables**

- Type of lane change: This variable represents if the cut-in maneuver is Mandatory Lane Change (MLC) or Discretionary Lane Change (DLC).
  - MLC
  - DLC
- Total number of lane changes: This variable represents the total number of lane changes performed by the POV in a single cut-in maneuver.
  - N = number of lanes that the POV changes during the cut-in maneuver
- Leading vehicle – target lane influence: This variable captures the influence of a leading vehicle on the target lane. The influence is 'yes' if the POV changes acceleration during cut-in maneuver due to the leading vehicle on target lane.
  - Yes
  - No

- Leading vehicle presence on origin lane: This variable captures the presence of a leading vehicle on the original lane of POV.
    - Yes
    - No
- Leading vehicle - origin lane influence: This variable captures the influence of a leading vehicle on the target lane. The influence is 'yes' if the POV changes acceleration during cut-in maneuver due to the leading vehicle on original lane.
    - Yes
    - No

Now there are two datasets for trajectory modelling: a kinematic variables dataset (Section 2.1.1) and a classification variables dataset (Section 2.1.2).

2.2 Trajectory modelling

This section explains the trajectory modelling methods of this thesis. It starts with identification of start and end time of each maneuver, which is essential to calculate the independent variable for subsequent models: the event duration. Then, a polynomial model is built for interpolating the trajectories in the input dataset. Finally, a probabilistic model of lateral trajectory of POV built for extrapolation of trajectories from input dataset is explained.

Using the classification variables derived in previous section, the cut-in maneuvers occurring from right to left in which POV performs a single lane change are considered for building a trajectory model. 33 out of 57 of the events were in this category. The data of these 33 inputs is the input dataset for the trajectory model.

2.2.1 Start and end points of the maneuver

The start and end points of the maneuver are required for calculating the event duration for different trajectories. Event duration is further used as an independent variable in models explained in further sections. In the ideal case where the SV is travelling straight in the target lane and the POV is cutting in onto SV's lane, the POV is initially travelling in its original lane without any lateral movement, changes lane and stabilizes on the target lane. In this case the lateral range rate of POV looks like the example shown in Figure 15 where the lateral range rate starts from $0\ m/s$, increase to a maximum value (in case of cut-in from right to left lane) and then end at $0\ m/s$.

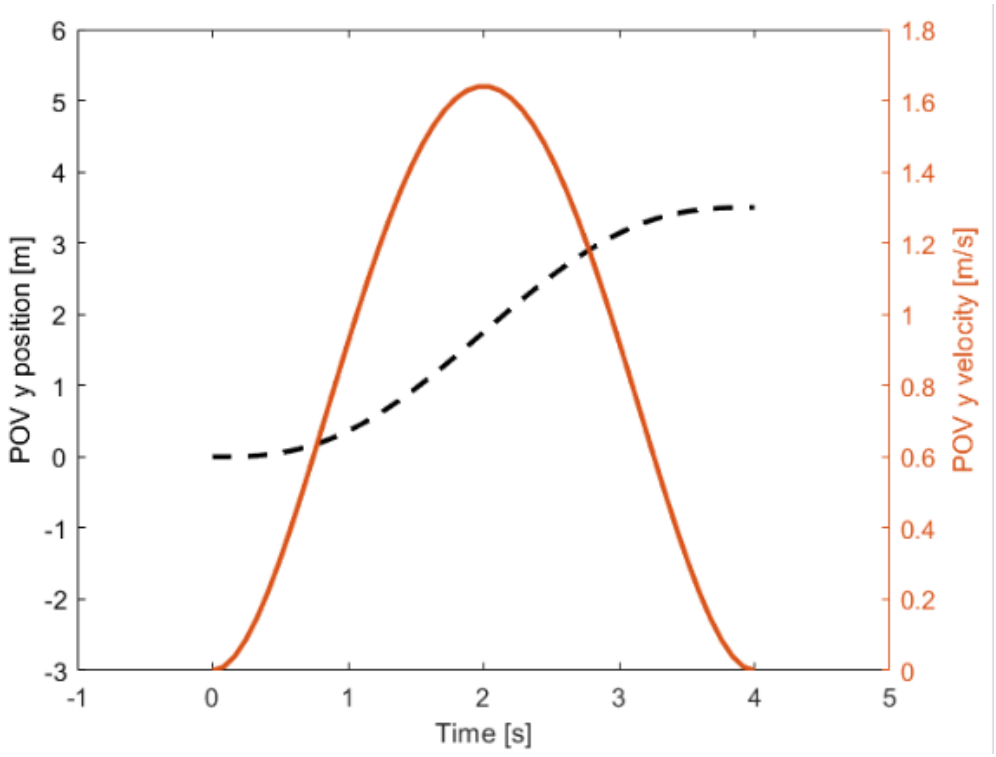


Figure 15: POV lateral range (dotted black curve) and lateral range rate (Orange bell shaped curve) for an ideal cut-in manoeuvre.

However, in the data derived from the annotation tool two differences were observed. First, the lateral range rate profile of POV is not a smooth bell-shaped curve like the ideal trajectory and had multiple peaks as shown in Figure 16. Secondly, the start and end points of the data are not at 0 m/s as shown below. These issues were because of manual annotation procedure employed in the video annotation tool (Shams El Din, 2020).

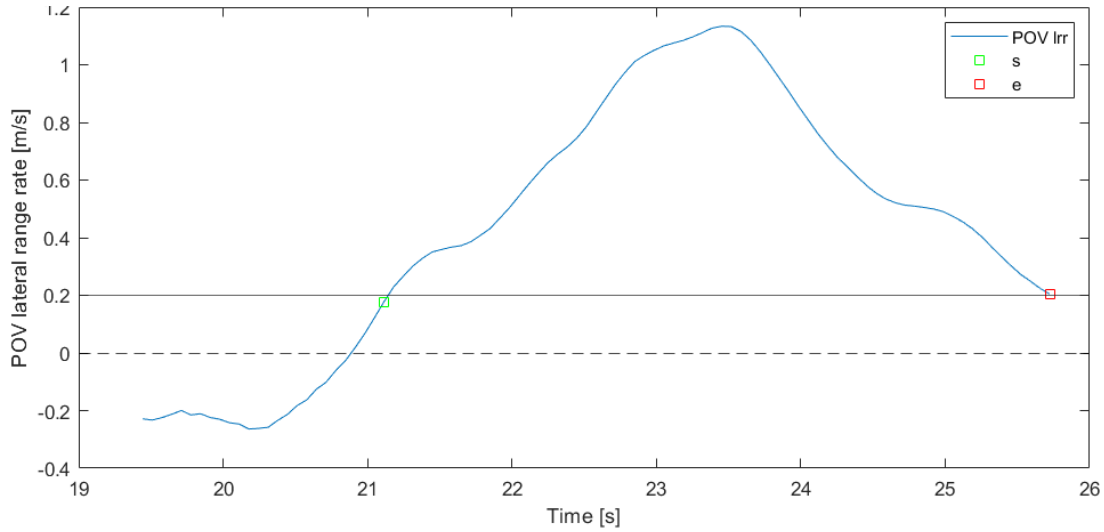


Figure 16: POV lateral range rate profile showing multiple peaks.

To overcome the first issue, the start and end points are identified by first identifying the maximum peak in POV lateral range rate profile and scanning on either side of the peak to determine the points. And the second issue is addressed by considering a

threshold value ($\pm 0.2\ m/s$; positive for cut-in maneuvers from right side to left side) to identify start and end points while scanning from the peak. A similar approach has been used in Wang et al. (2014) and Shams El Din (2020), and according to Wang et al. (2014), changing the threshold value changes the average value of the event duration but the distribution trend of the event duration is not much affected. The start and end points identified through this method are shown for one event in Figure 17.

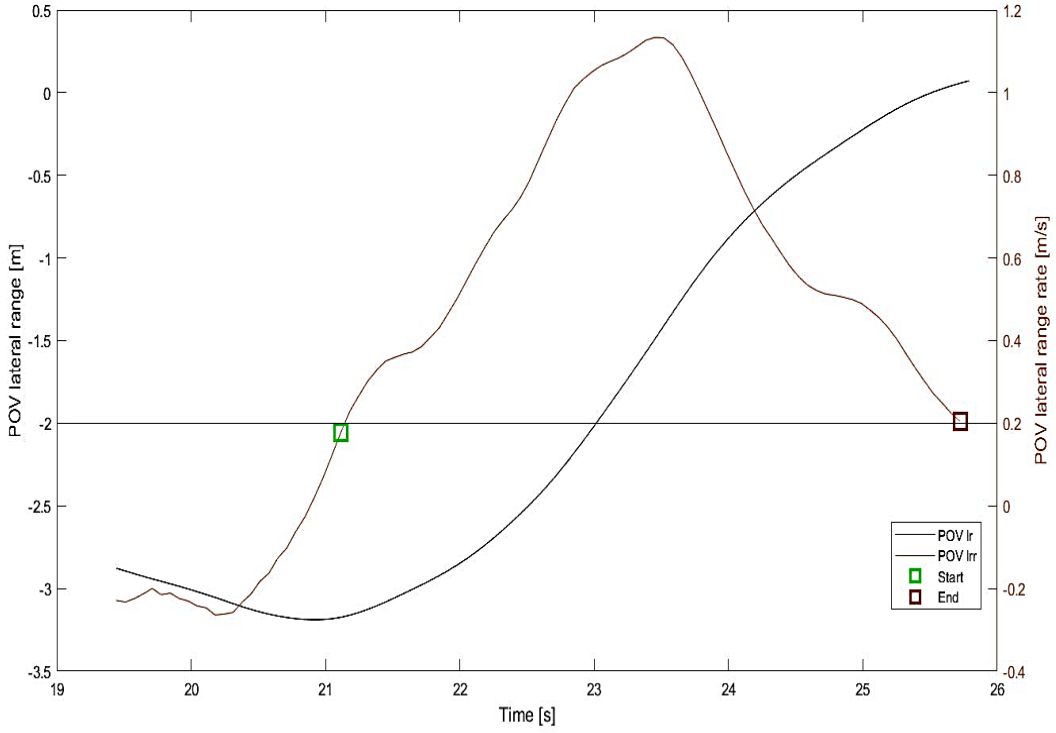


Figure 17: POV lateral range on left y axis and POV lateral range rate profile on right y axis, against event duration on x axis of one event. Start and end points marked on either side of the lateral range rate peak.

Once the start and end points are calculated, the trajectories are plotted together to verify their shape. The lateral and longitudinal components of POV trajectories are shown in Figure 18 and 19 respectively.

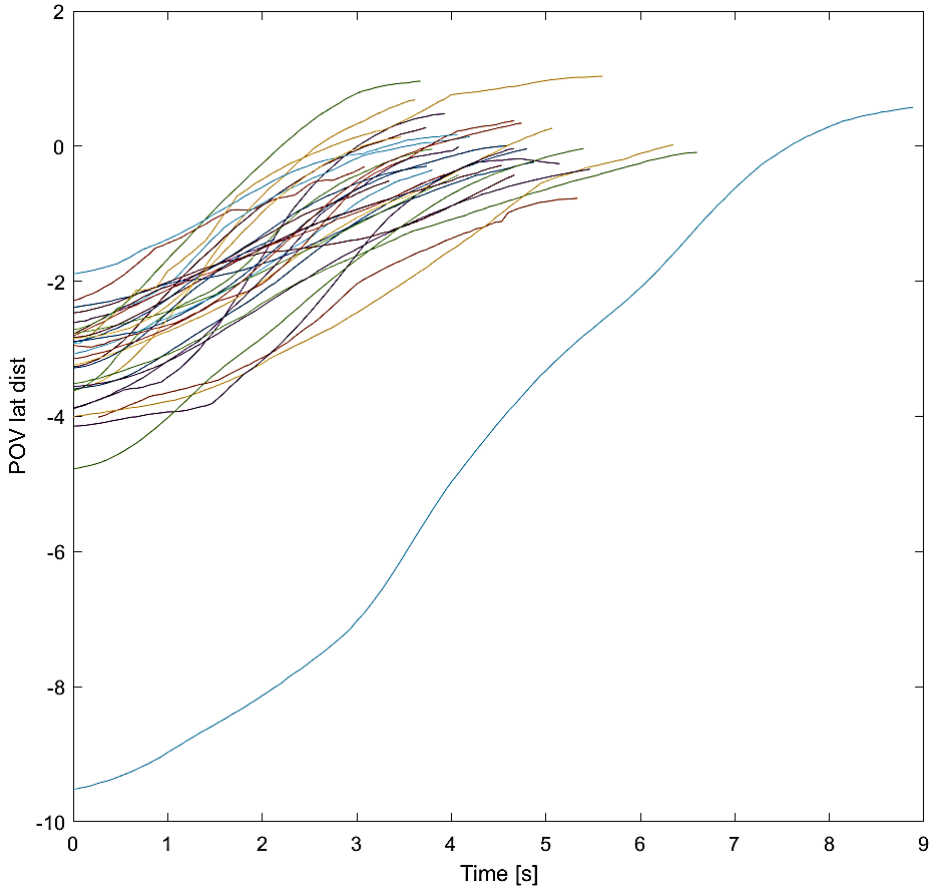


Figure 18: Lateral trajectories of cut-in manoeuvres occurring from right to left in which POV performs a single lane change.

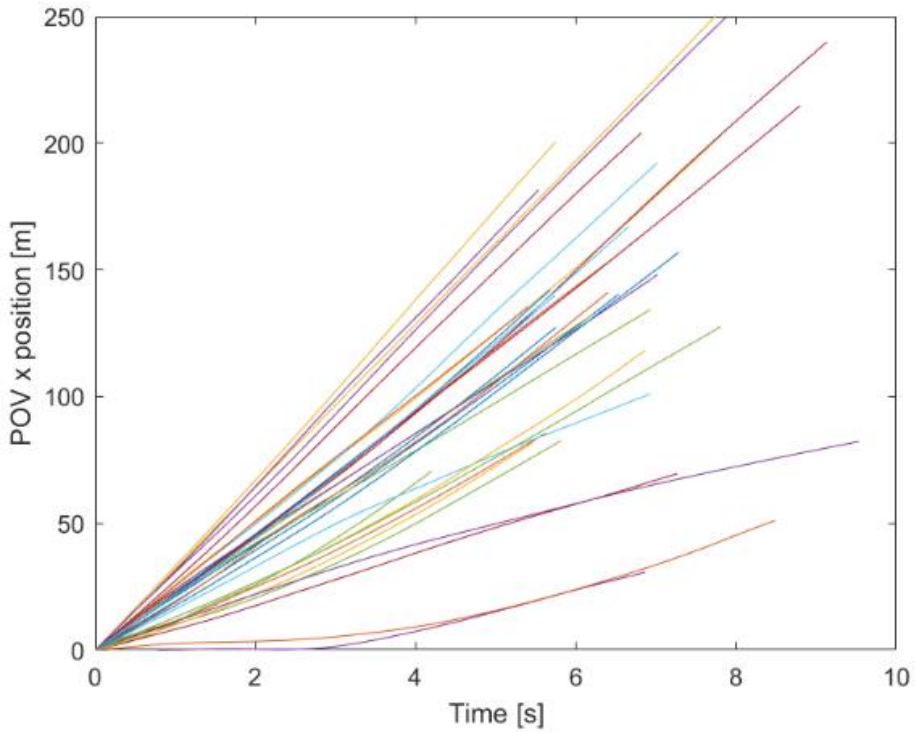


Figure 19: Longitudinal trajectories of cut-in manoeuvres occurring from right to left in which POV performs a single lane change.

Now the event durations of the trajectories are calculated as:

$$Event\ duration = end\ time - start\ time$$

After obtaining the event duration for all the events, the distribution for the duration is found by fitting 10 different distributions in the JMP Pro software and the best fit is found using Akaike Information Criterion (AIC) as a measure of goodness of fit.

2.2.2 Polynomial model for interpolation of trajectories

The trajectory of the POV can be defined by any of the kinematic variables - position, speed, and acceleration in longitudinal and lateral directions of POV. These variables can be expressed as functions of time as shown in equations 1 to 6.

$$x_{POV} = f(t) \quad (1)$$
$$\dot{x}_{POV} = \dot{f}(t) \quad (2)$$
$$\ddot{x}_{POV} = \ddot{f}(t) \quad (3)$$
$$y_{POV} = g(t) \quad (4)$$
$$\dot{y}_{POV} = \dot{g}(t) \quad (5)$$
$$\ddot{y}_{POV} = \ddot{g}(t) \quad (6)$$

Based on the literature study, it was found that a lane change maneuver can be expressed in the form of a quintic polynomial in time. The trajectory of the POV is expressed as the function $f(x, y, t)$ and is built based on the initial state (S_s) and final state (S_e) of the POV, represented by equations (19), (20).

$$S_s = (x_s, \dot{x}_s, \ddot{x}_s, y_s, \dot{y}_s, \ddot{y}_s) \quad (19)$$
$$S_e = (x_e, \dot{x}_e, \ddot{x}_e, y_e, \dot{y}_e, \ddot{y}_e) \quad (20)$$

Where $x_s, \dot{x}_s, \ddot{x}_s$ are the longitudinal displacement, velocity and acceleration respectively, and $y_s, \dot{y}_s, \ddot{y}_s$ are the lateral displacement, velocity and acceleration respectively. The equations of quintic polynomials of lateral and longitudinal displacements in terms of time, equations (1) and (2) can be rewritten as below.

$$y = a_5 t_n^5 + a_4 t_n^4 + a_3 t_n^3 + a_2 t_n^2 + a_1 t_n + a_0 \quad (21)$$

$$x = b_5 t_n^5 + b_4 t_n^4 + b_3 t_n^3 + b_2 t_n^2 + b_1 t_n + b_0 \quad (22)$$

Similarly, the equation for velocities and accelerations in lateral and longitudinal directions can be obtained by taking derivatives of equations (21) and (22). So, in total there are six unknown parameters for each of the directions (lateral and longitudinal) and six equations which can be represented as the following set of linear equations. The equations for lateral and longitudinal components are shown below:

$$\begin{bmatrix} t_s^5 & t_s^4 & t_s^3 & t_s^2 & t_s & 1 \\ t_e^5 & t_e^4 & t_e^3 & t_e^2 & t_e & 1 \\ 5t_s^4 & 4t_s^3 & 3t_s^2 & 2t_s & 1 & 0 \\ 5t_e^4 & 4t_e^3 & 3t_e^2 & 2t_e & 1 & 0 \\ 20t_s^3 & 12t_s^2 & 6t_s & 2 & 1 & 0 \\ 20t_e^3 & 12t_e^2 & 6t_e & 2 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} a_5 \\ a_4 \\ a_3 \\ a_2 \\ a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} y_s \\ y_e \\ \dot{y}_s \\ \dot{y}_e \\ \ddot{y}_s \\ \ddot{y}_e \end{bmatrix} \quad (23)$$

$$\begin{bmatrix} t_s^5 & t_s^4 & t_s^3 & t_s^2 & t_s & 1 \\ t_e^5 & t_e^4 & t_e^3 & t_e^2 & t_e & 1 \\ 5t_s^4 & 4t_s^3 & 3t_s^2 & 2t_s & 1 & 0 \\ 5t_e^4 & 4t_e^3 & 3t_e^2 & 2t_e & 1 & 0 \\ 20t_s^3 & 12t_s^2 & 6t_s & 2 & 1 & 0 \\ 20t_e^3 & 12t_e^2 & 6t_e & 2 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} b_5 \\ b_4 \\ b_3 \\ b_2 \\ b_1 \\ b_0 \end{bmatrix} = \begin{bmatrix} x_s \\ x_e \\ \dot{x}_s \\ \dot{x}_e \\ \ddot{x}_s \\ \ddot{x}_e \end{bmatrix} \quad (24)$$

To reduce the number of inputs in equations (23) and (24), the following assumptions are made:

- The start time, initial longitudinal and lateral displacements are zero as the starting point of POV was considered as zero.
- The final lateral displacement is equal to the lane width (i.e., the movement between the center of two lanes).
- The initial and final longitudinal velocity of POV is constant and hence the initial and final longitudinal accelerations are zero. This was because the observed longitudinal trajectories of the 33 events appeared to be straight lines with a constant slope (as shown in section 2.2.1).
- The initial and final lateral velocities and final lateral acceleration are zero.

With these assumptions' equations (23) and (24) can be rewritten as:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ t_e^5 & t_e^4 & t_e^3 & t_e^2 & t_e & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 5t_e^4 & 4t_e^3 & 3t_e^2 & 2t_e & 1 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 \\ 20t_e^3 & 12t_e^2 & 6t_e & 2 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} a_5 \\ a_4 \\ a_3 \\ a_2 \\ a_1 \\ a_0 \end{bmatrix} = \begin{bmatrix} 0 \\ y_e \\ 0 \\ 0 \\ \ddot{y}_s \\ 0 \end{bmatrix} \quad (25)$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ t_e^5 & t_e^4 & t_e^3 & t_e^2 & t_e & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 5t_e^4 & 4t_e^3 & 3t_e^2 & 2t_e & 1 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 \\ 20t_e^3 & 12t_e^2 & 6t_e & 2 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} b_5 \\ b_4 \\ b_3 \\ b_2 \\ b_1 \\ b_0 \end{bmatrix} = \begin{bmatrix} 0 \\ x_e \\ \dot{x}_s \\ \dot{x}_e \\ 0 \\ 0 \end{bmatrix} \quad (26)$$

Now there are four unknown variables for building trajectories using above equations and they are:

1. End time, which is equal to the event duration, since the start time is zero
2. Initial lateral acceleration of the POV
3. Final longitudinal displacement of the POV
4. Longitudinal velocity of the POV

The event duration was calculated in section 2.2.1 and its distribution was determined.

The initial lateral acceleration of the POV was first obtained by differentiating the lateral range rate. However when the trajectories were generated using this acceleration values they were showing a different shape compared to the actual trajectory and hence initial lateral acceleration of POV was determined by minimizing a cost function shown in Equation (27), which is calculated as the sum of squares of the difference in distances between original and generated trajectories at each timestep.

$$Cost = \sum_{i=1}^{N} \left(y_{pov,n} - y_{pov,polynomial,n}\right)^2 \qquad (27)$$

Where $y_{pov,n}$ is the lateral position of the POV at each time step n and $y_{pov,polynomial,n}$ is the lateral position of the POV generated by equation (25). The value of initial lateral acceleration is searched in the range of $-4\ m/s^2$ to $6\ m/s^2$ and cost is calculated. Then, the minimum cost acceleration is used for generating trajectories. The comparison of trajectories generated using differentiated acceleration and the low-cost acceleration can be seen in Figure 20 and 21.

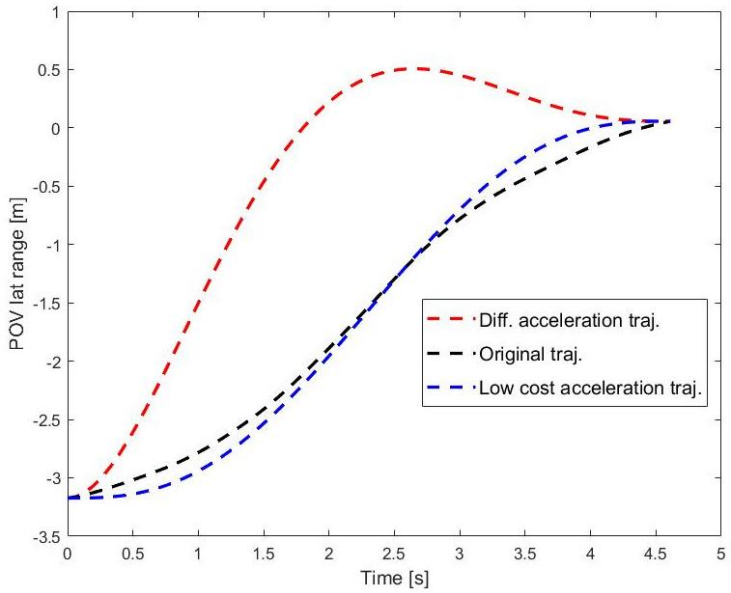


Figure 20: Comparing trajectories generated from differentiated acceleration and low-cost acceleration Event1.

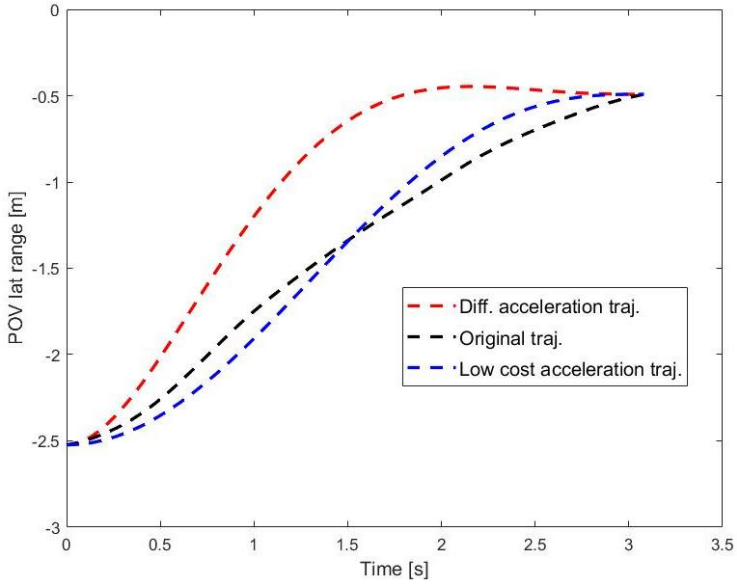


Figure 21: Comparing trajectories generated from differentiated acceleration and low-cost acceleration Event2.

Using the minimum cost lateral acceleration values of all the events, a linear model is built using MATLAB's '$fitlm()$' function to determine the initial lateral acceleration, using the event duration as independent variable. This function creates a linear model which is used to calculate initial lateral acceleration values of POV for any event duration. The general form of such linear model is $y = c_1 x + c_0$, where y is the dependent variable, lateral acceleration in our case, and x is the independent variable, which is event duration in our case and c_0, c_1 are coefficients of the linear model which are calculated using the input data of minimum cost acceleration values of all events. Likewise, another linear model is built for final longitudinal displacement of POV using event duration as the independent variable.

The longitudinal velocity of POV is calculated as

$$\dot{x}_s = \dot{x}_e = \dot{x} = \frac{x_e - x_s}{t_e - t_s} = \frac{x_e}{t_e} \qquad (28)$$

Finally, a range of interpolated trajectories are generated by following sequence mentioned below and any number of trajectories could be generated in the range of cut-in durations:

1. Select an event duration from the event duration distribution
2. Calculate initial POV lateral acceleration, final POV longitudinal position using the linear models of event duration and POV longitudinal velocity using equation (28)
3. Generate trajectories by using equations (25) and (26)

This is further presented in section 3.1.2.1.

2.2.3 Probabilistic model for extrapolation of trajectories

In addition to calculating the quintic polynomial parameters for lateral and longitudinal trajectories of POV, a probability distribution of the parameters

calculated using the input dataset was built. According to Rogers & Girolami (2016) textbook, probabilistic regression models can be used as generative models which calculate the probability distribution of the model parameters using training data.

A model built for lateral trajectory of POV is explained here. In this model, the POV lateral position is the output variable which is represented as the quintic polynomial of time (input variable). In a probabilistic model, a random noise variable is added to the input variable to include the uncertainty of input measurements into the model. The basic form of the model is shown below:

$$y_{POV} = f(t) + \epsilon \qquad (29)$$

Here, $f(t)$ is the quintic polynomial shown in equation (21) and ϵ is an additive noise term. The noise is modelled as a random variable with gaussian distribution for ease of mathematical computation. This is shown in equation (30) along with the parameters of gaussian distribution.

$$\epsilon = \mathcal{N}(0, \sigma^2) \qquad (30)$$

The mean of the noise variable is zero and variance is σ^2. For any given time step n, the model will be as shown in equation (31).

$$y_{n,POV} = c_5 t_n^5 + c_4 t_n^4 + c_3 t_n^3 + c_2 t_n^2 + c_1 t_n + c_0 + \epsilon_n \qquad (31)$$

In vector form, equation (31) can be written as:

$$y_{n,POV} = \boldsymbol{c}^T \boldsymbol{t}_n + \epsilon_n \qquad (32)$$

Where $\boldsymbol{c} = [c_0, c_1, \dots, c_5]^T$ and $\boldsymbol{t}_n = [1, t_n, t_n^2, \dots, t_n^5]^T$.

Stacking all responses into single matrix, $\boldsymbol{y}_{POV} = [y_1, y_2, \dots, y_N]^T$ and all the inputs into single matrix, $\boldsymbol{T} = [\boldsymbol{t}_1, \boldsymbol{t}_2, \dots, \boldsymbol{t}_N]$, we get the following equation for one event for entire event duration. Here, N is the total number of timesteps or length of one event from start to end.

$$\boldsymbol{y}_{POV} = \boldsymbol{T}\boldsymbol{c} + \boldsymbol{\epsilon} \qquad (33)$$

The parameters of the above model are represented by the vector $\boldsymbol{c}$ and the probability distribution of $\boldsymbol{c}$ is to be calculated to generate the additional trajectories. This is possible by using the Bayes rule of probability. The Bayes equation for the above model can be expressed as shown in equation (34)

$$P(\boldsymbol{c}|\boldsymbol{y}_{POV}, \boldsymbol{T}, \sigma^2, \Delta) = \frac{P(\boldsymbol{y}_{POV}|\boldsymbol{c}, \boldsymbol{T}, \sigma^2, \Delta) * P(\boldsymbol{c}|\Delta)}{P(\boldsymbol{y}_{POV}|\boldsymbol{T}, \sigma^2, \Delta)} \qquad (34)$$

The numerator in the right part of equation (12) has two terms – likelihood $P(\boldsymbol{y}_{POV}|\boldsymbol{c}, \boldsymbol{T}, \sigma^2, \Delta)$, and prior $P(\boldsymbol{c}|\Delta)$. Where Δ represents some set of parameters to define prior over $\boldsymbol{c}$ as explained in the following paragraphs.

The likelihood is the probability distribution of $\boldsymbol{y}_{pov}$ calculated from the input data. In equation (33), addition of gaussian random variable noise also makes $\boldsymbol{y}_{POV}$ a gaussian random variable. Hence the equation for likelihood term becomes:

$$P(\boldsymbol{y}_{POV}|\boldsymbol{c},\boldsymbol{T},\sigma^2,\Delta) = \mathcal{N}(\boldsymbol{T}\boldsymbol{c},\sigma^2\boldsymbol{I}_N) \qquad (35)$$

Where $\boldsymbol{I}_N$ is an identity matrix of size N.

According to Section 3.1 of Rogers & Girolami (2016), a likelihood-prior pair is said to be conjugate if they result in a posterior which is in the same form of prior. To get the exact expression for posterior term, which is explained in next paragraph, a prior is chosen such that it forms a conjugate pair with the gaussian likelihood. This results in a gaussian prior shown below.

$$P(\boldsymbol{c}|\Delta) = \mathcal{N}(\mu_0,\Sigma_0) \qquad (36)$$

Hence Δ represents the parameters μ_0,Σ_0, which are the mean and covariance of the gaussian distribution of prior over $\boldsymbol{c}$. The mean μ_0 is a null matrix since there is no prior information about the parameter values. The covariance Σ_0 is an identity matrix as the assumption is that the individual parameters are independent and have equal variances.

The left part of equation (34) is known as the posterior term and this is what is to be calculated. This term represents the probability distribution of parameters of the quintic polynomial, $\boldsymbol{c}$, given the input data of lateral positions $\boldsymbol{y}_{POV}$, the time $\boldsymbol{T}$**,** and variance of the noise σ^2. This can be used to generate new sets of values of parameters which can further be used to generate extrapolated trajectories for new time values using equation (37).

$$P\left(\boldsymbol{y}_{POV,new}\middle|\boldsymbol{x}_{new},\boldsymbol{T},\sigma^2,\Delta\right) = \int P\left(\boldsymbol{y}_{POV,new}\middle|\boldsymbol{x}_{new},\boldsymbol{c},\sigma^2\right) * P(\boldsymbol{c}|\boldsymbol{y}_{POV},\boldsymbol{T},\sigma^2,\Delta) \quad (37)$$

By choosing a conjugate likelihood-prior pair, the form of posterior is known which is a gaussian. The expression of posterior and its gaussian parameters are shown below.

$$P(\boldsymbol{c}|\boldsymbol{y}_{POV},\boldsymbol{T},\sigma^2,\Delta) = \mathcal{N}(\mu_c,\Sigma_C) \qquad (38)$$

$$\Sigma_c = \left(\frac{1}{\sigma^2}\boldsymbol{T}^T\boldsymbol{T} + \Sigma_0^{-1}\right)^{-1} \qquad (39)$$

$$\mu_c = \Sigma_c\left(\frac{1}{\sigma^2}\boldsymbol{T}^T\boldsymbol{y}_{POV} + \Sigma_0^{-1}\mu_0\right) \qquad (40)$$

3 Results

This section explains the results obtained for trajectory modelling starting from data classification and then results from the polynomial modelling of trajectory are presented, followed by the probabilistic model for trajectory generation.

3.1.1 Data Classification

As explained in section 2.1.2, classification variables were collected via manual annotation of SHRP2 video data. These classification variables were then used to identify the similarities in cut-in trajectories.

The initial data of 57 trajectories was plotted using the kinematic variables from video annotation tool. These trajectories look as shown below:

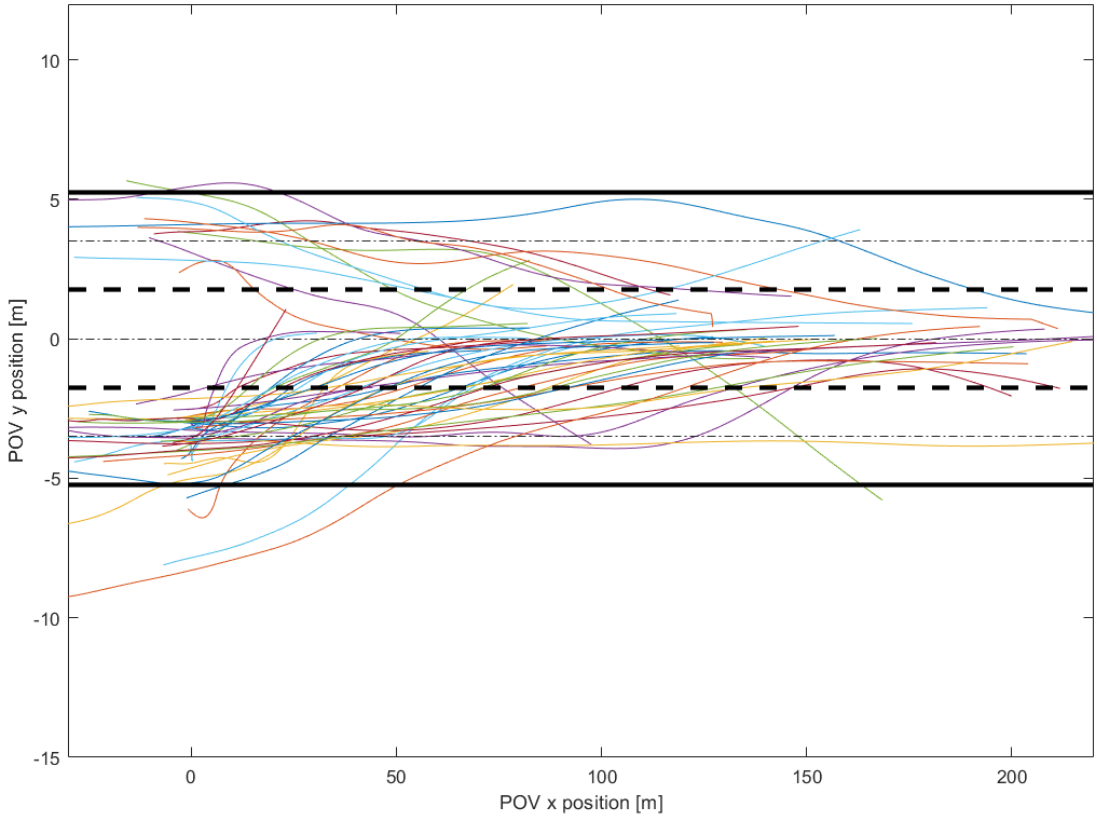


Figure 22: Plot of 57 cut-in trajectories from SHRP2. Zero on y-axis represents the target lane centre.

From this image it can be seen that there are cut-ins starting from left side and right side and that there are more starting from the right. Also, the cut-in maneuvers were classified as either MLC or DLC as explained in section 2.1.2. The distributions of the type of lane change and side of lane change is shown below:

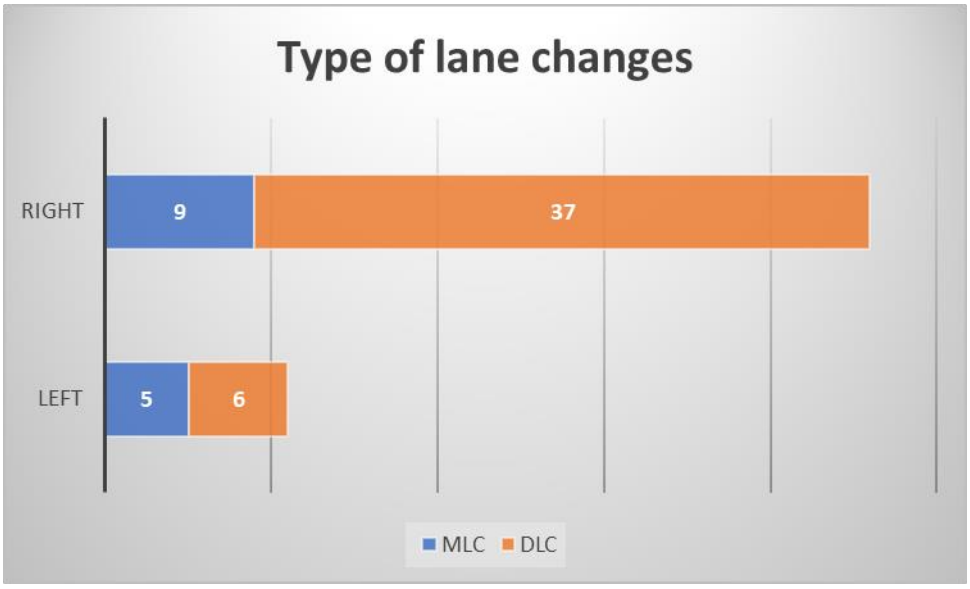


Figure 23: Distribution of types of cut-ins and starting side of cut-in.

It can be seen that there are more cut-ins starting from right side and also that majority of them are discretionary type lane changes (DLC). The distribution of the cut-in motivation is shown below:

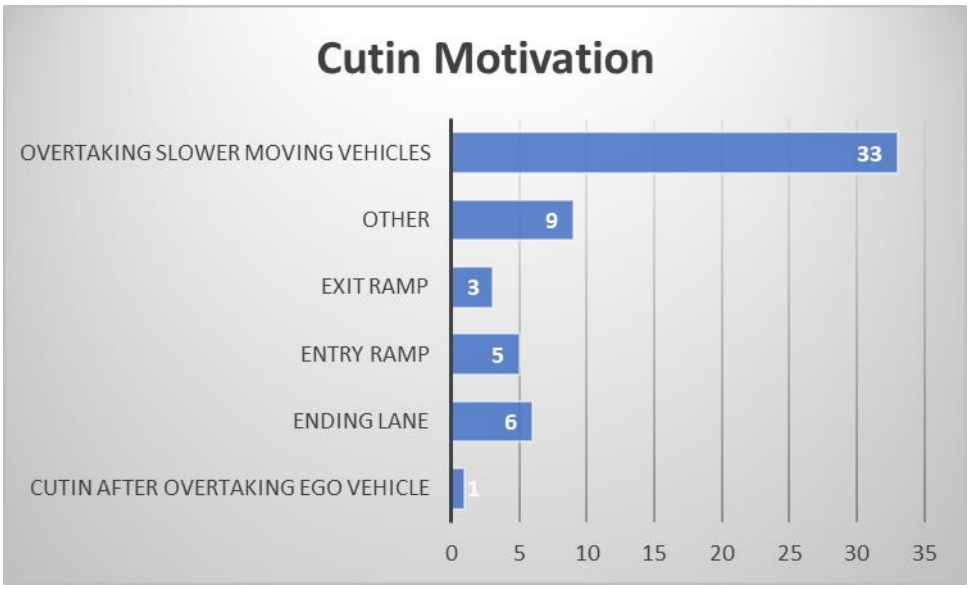


Figure 24: Distribution of events based on cut-in motivation.

It can be seen that the major motivation for cut-ins is a slow-moving vehicle in front of the POV. Another interesting result is the number of lane changes POV performed in a cut-in manoeuvre. The number of lane changes can be classified based on starting side of cut-in manoeuvre and type of lane change. These two distributions are shown below:

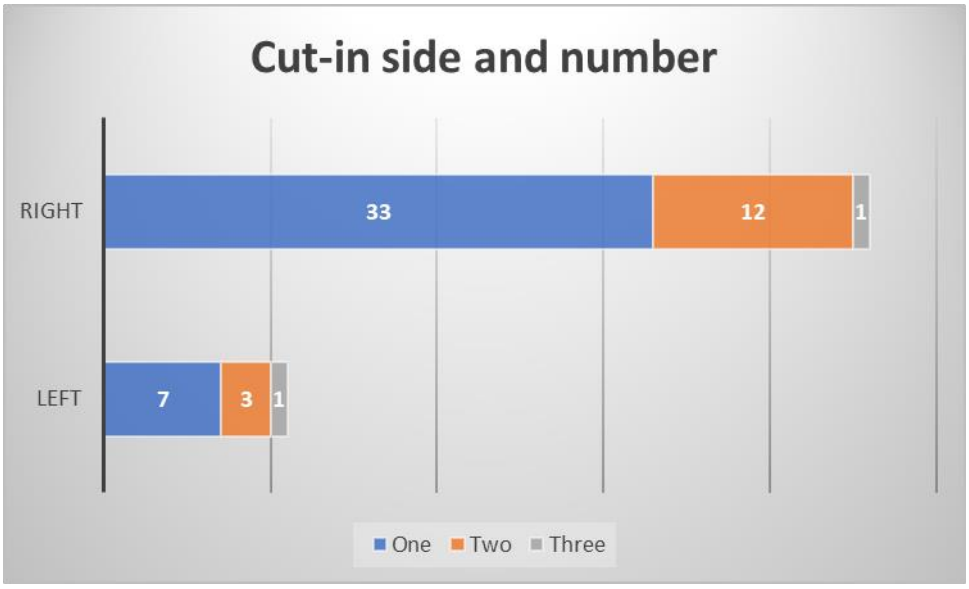


Figure 25: Distribution of number of lane changes performed by POV in a cut-in manoeuvre grouped based on starting side of cut-in.

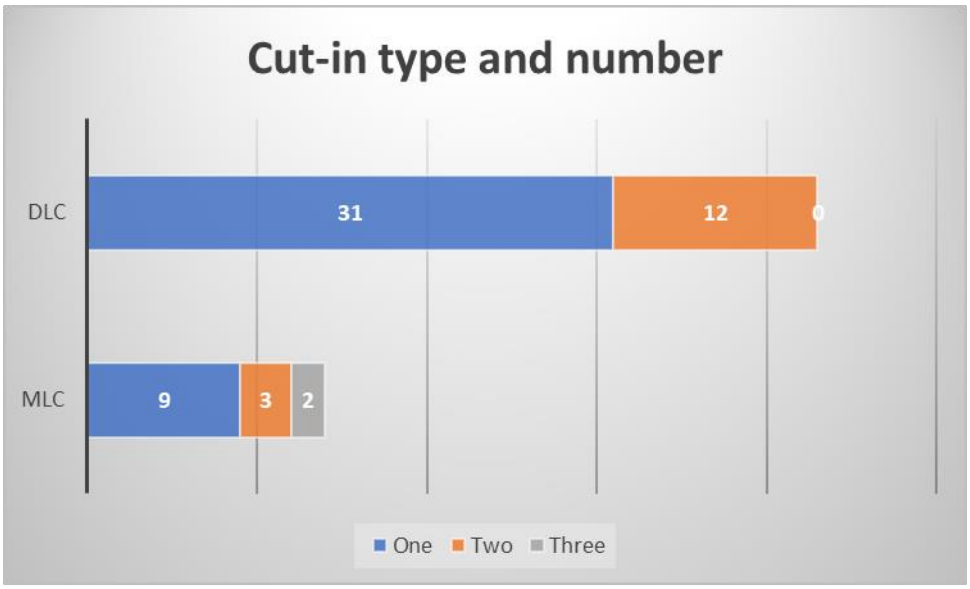


Figure 26: Distribution of number of lane changes performed by POV in a cut-in manoeuvre grouped based on type of cut-in.

From the above results it can be seen that single lane changes/cut-ins are more and are selected for further analysis which resulted in 33 events available for further analysis. The lateral and longitudinal trajectories of cut-ins starting from right side and having single lane change are shown in Figures

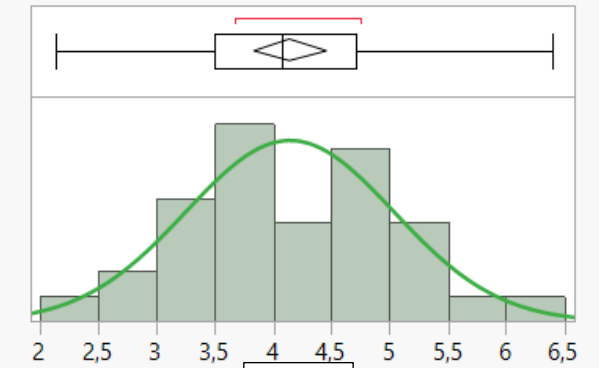
The above trajectories show that the lateral trajectory follows a S shaped curve and longitudinal trajectories follow a straight line with constant slope for the selected group of manoeuvres. These were further used to build trajectory models, as explained in below sections.

3.1.2 Trajectory modelling

This section shows the results of trajectory models built to be used for trajectory interpolation and extrapolation. These models used the kinematic data of the 33 events grouped as explained in previous section.

3.1.2.1 Start and end points, event duration

As explained in section 2.1.1, event duration of the trajectories has been calculated using the start and end points of the trajectory. Then, different distributions have been considered to model the event durations and it was found that the normal distribution fits best. The comparison of AIC score for different distributions is shown below along with the distribution curve of event durations. The distribution modelling has been done in JMP Pro software from SAS using the *distribution* function. To know more about this function refer to (Distributions (jmp.com))



Compare Distributions

Distribution	AICc	AICc Weight
Normal	89,442219	0,3178
Gamma	89,614931	0,2915
Lognormal	90,542809	0,1833
Weibull	90,852204	0,157
SHASH	94,226154	0,0291
Johnson Sb	95,038055	0,0194
Normal 2 Mixture	100,33301	0,0014
Cauchy	102,10801	0,0006
Normal 3 Mixture	110,28316	9,5e-6
Exponential	161,98067	0

Figure 27: Top: Plot showing the distribution of event duration and the probability density curve of normal distribution. Bottom: Comparison of AIC values for different distributions. Results obtained using (JMPPro15).

The probability density function of normal distribution is

$$f(t|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{\frac{-(t-\mu)^2}{2\sigma^2}\right\}$$

Where t is the event duration, μ and σ are normal mean and variance of the normal distribution. This thesis found that $\mu = 4.14\ s$ and $\sigma = 0.89\ s$ describe the normal

distribution of the cut-in duration for the 33 SHRP2 cut-in near-crashes used in this study and described in section 3.1.1.

3.1.2.1 Polynomial model for interpolation of trajectories

The polynomial trajectory built using the equations (23) and (24) for one event is plotted along with the original trajectory of the event extracted from the annotation tool in figure below.

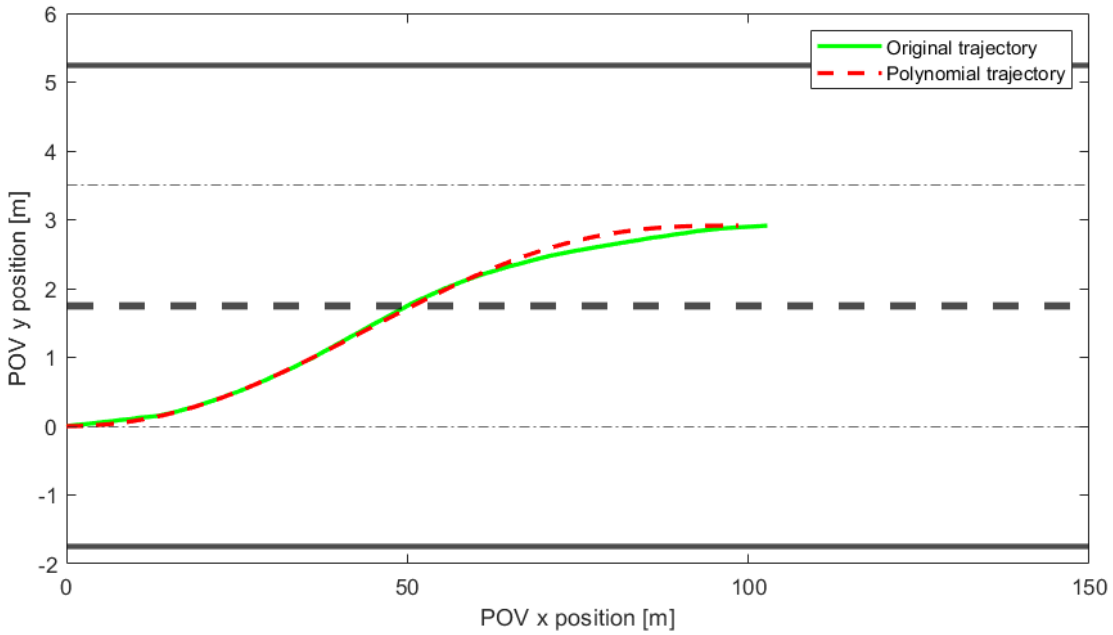


Figure 28: Figure showing data of one event. The original trajectory extracted from annotation tool is shown as green line and the red line is the trajectory derived using the polynomial model.

It can be seen that the polynomial model generates a trajectory with similar shape of the trajectory extracted from the annotation tool. However, the model only fits well the trajectories with 'S' shape and is not sufficiently suited to represent trajectories with different shape (e.g., linear shape). An example is shown below:

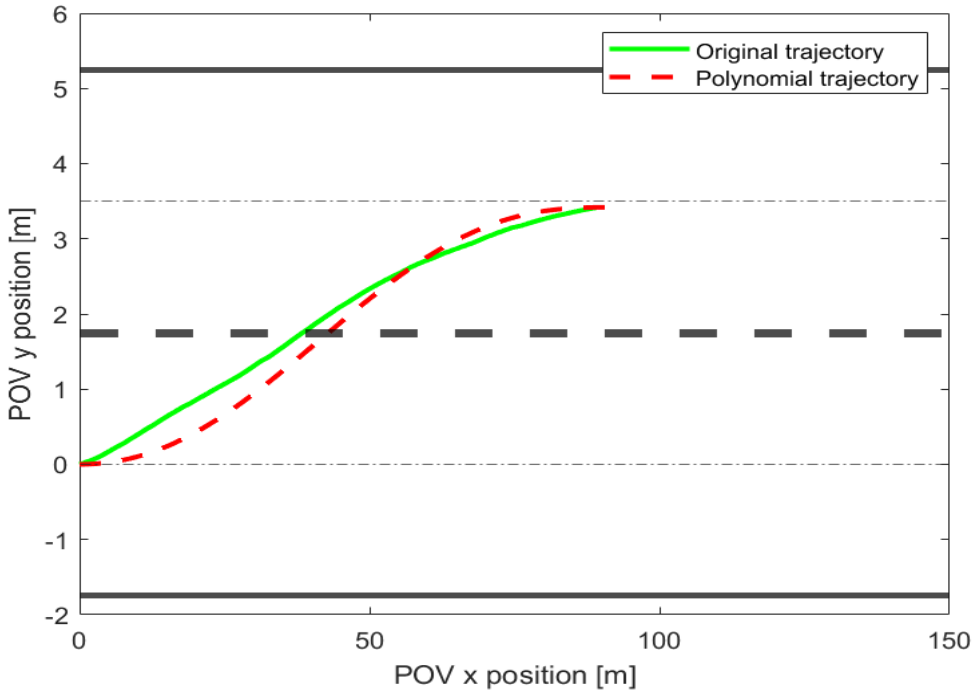


Figure 29: Figure showing data of one event in which the polynomial model does not fit well the data. The original trajectory extracted from annotation tool is shown as green line and the red line is the trajectory derived using the polynomial model.

As mentioned in section 2.2, the polynomial trajectory model shown above was used to generate additional trajectories from the original 33 maneuvers extracted from the SHRP2 naturalistic dataset. Also, as mentioned in section 2.2, two linear models have been made to interpolate the values of POV initial lateral acceleration and final longitudinal position for the range of event duration obtained from the distribution shown in Section 3.1.2.1. As described in the Method section, the initial lateral acceleration and the final longitudinal position determined with the linear models are used for the generation of trajectories. The linear models are shown below:

```
y_lm =
Linear regression model:
    POV_lat_acc_start ~ 1 + Event_duration

Estimated Coefficients:
                       Estimate      SE        tStat      pValue
                       ________    _______    _______    _________

    (Intercept)         4.1439      1.1996     3.4542    0.0016197
    Event_duration     -0.7584     0.28312    -2.6787     0.011718


Number of observations: 33, Error degrees of freedom: 31
Root Mean Squared Error: 1.43
R-squared: 0.188,  Adjusted R-Squared: 0.162
F-statistic vs. constant model: 7.18, p-value = 0.0117
```

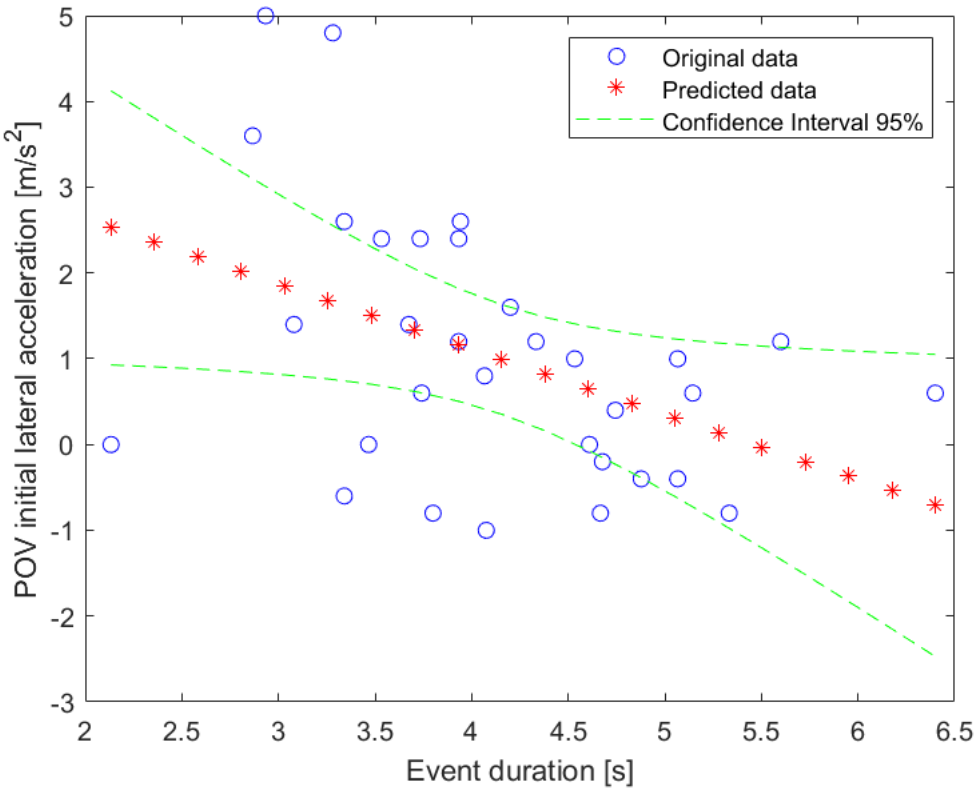


Figure 30: Linear regression model of POV initial lateral acceleration and event duration. Blue data points are from SHRP2 data set. Red points are the predictions of linear model along with the confidence interval shown as green lines.

```
y_lm1 =
Linear regression model:
    POV_long_displ_end ~ 1 + Event_duration

Estimated Coefficients:
                       Estimate      SE       tStat       pValue
                       ________    ______    ________    _________

    (Intercept)        -5.3355     28.822    -0.18512      0.85434
    Event_duration      22.537     6.8021      3.3132    0.0023536


Number of observations: 33, Error degrees of freedom: 31
Root Mean Squared Error: 34.3
R-squared: 0.262,  Adjusted R-Squared: 0.238
F-statistic vs. constant model: 11, p-value = 0.00235
```

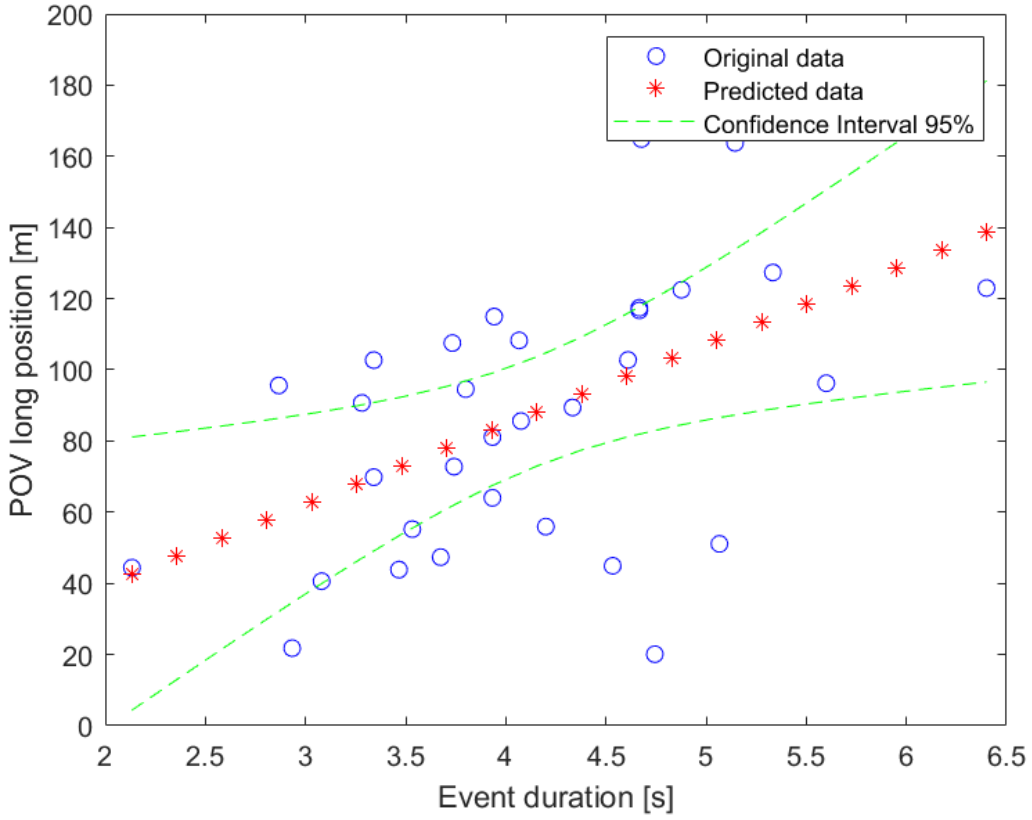


Figure 31: Linear regression model of POV final longitudinal position and event duration. Blue data points are from SHRP2 data set. Red points are the predictions of linear model along with the confidence interval shown as green lines.

It can be seen that the trends in these linear models show that POV lateral acceleration is higher for manoeuvres with short duration and decreases as the duration increases, and final longitudinal position of POV increases with the event duration. However, these models could only capture 18% and 26% of variation in input data and hence is not fitting well with the data. These linear models are used to interpolate the required inputs for the polynomial trajectory model. For a given event duration, the other required variables for polynomial model—final longitudinal displacement of POV, initial lateral acceleration and longitudinal velocity—are obtained from the linear models, and then trajectories are generated. This method can be used to generate any

number of trajectories within the range of event duration. A sample of 25 interpolated trajectories is shown below.

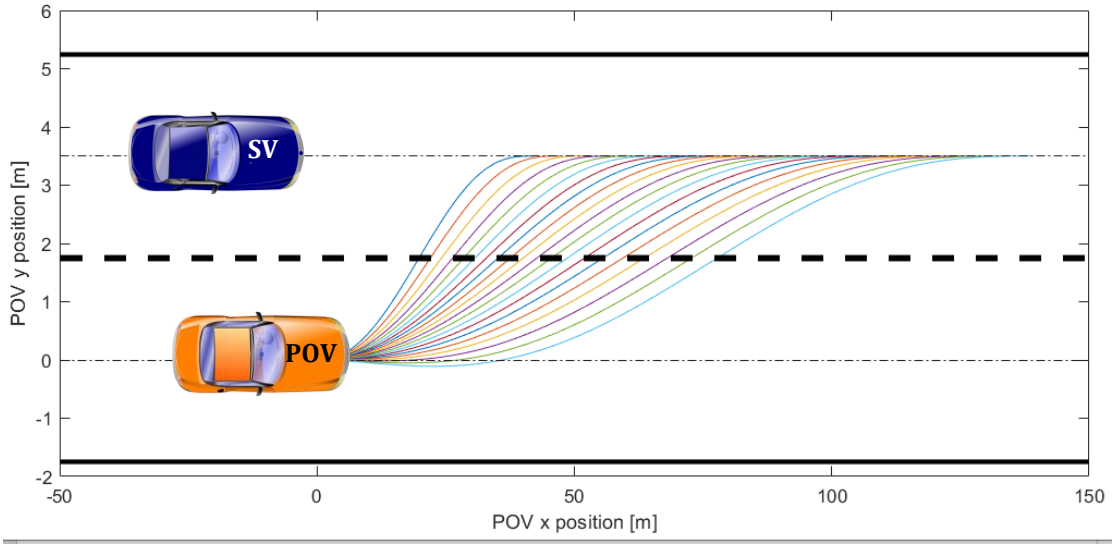


Figure 32: Trajectories of POV generated using polynomial model.

3.1.2.2 Probabilistic Regression model for extrapolation of trajectories

To build the probability model, the data of one event annotated by 5 different annotators in the previous project (Shams El Din, 2020) is considered and shown below:

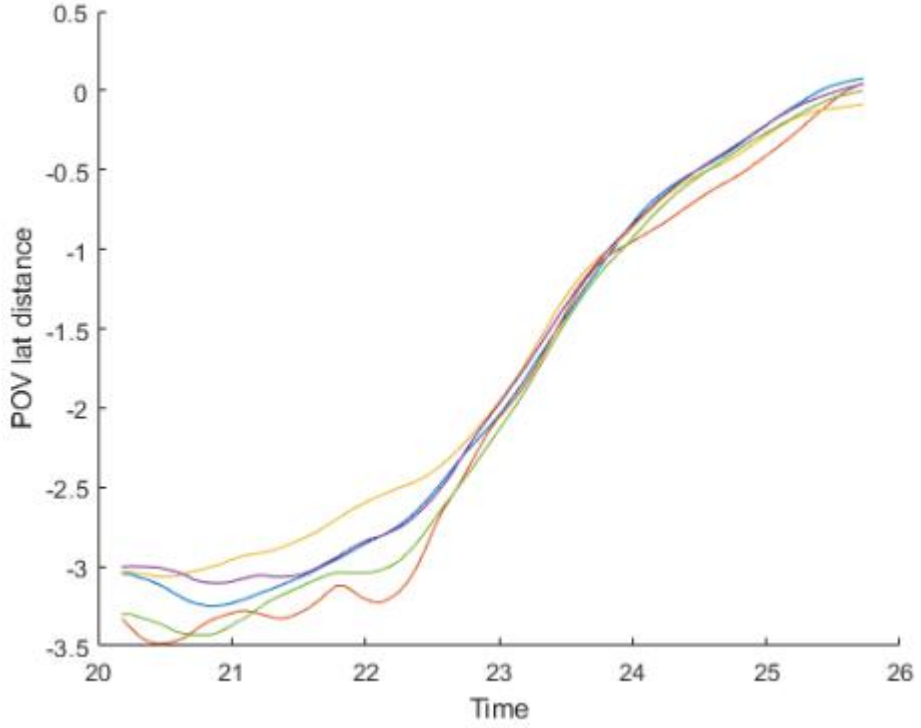


Figure 33: Figure showing the POV lateral distance data for one event. Five coloured curves represent the annotation data from 5 annotators.

Using this input data, the probability distribution of the quintic polynomial trajectory coefficients were calculated using equations (37) to (40). Then, a random set of these parameters were drawn using the calculated distribution. Any number of trajectories can be generated using these parameter sets. A sample of 50 trajectories generated for this event using the randomly generated parameters from posterior distribution is shown in Figure 34 below.

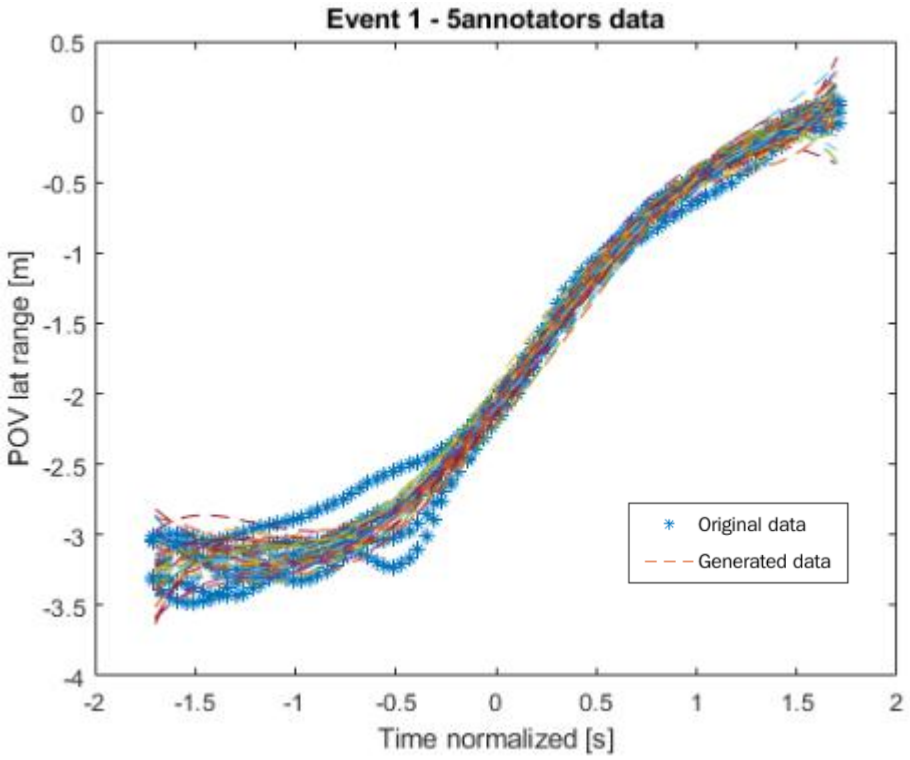


Figure 34: Figure showing the generated trajectories from the posterior distribution. Blue stars represent raw data of 5 annotators. Coloured dotted lines represent generated trajectories.

4 Discussion and Conclusion

AD/ADAS systems have a great potential to improve road safety in the future and one of the main factors necessary for designing such systems is understanding human driver behavior and predicting correct behavior of other road participants in a given scenario. However, the data of critical scenarios is scarce because of low chances of occurrence and due to the limitations of current data collection systems. One of the ways to overcome this challenge is to develop models of driver trajectories using the limited available samples and then interpolate and extrapolate new trajectories. Two models in this context have been studied in this thesis – a polynomial model for interpolation of trajectories and a probabilistic model for extrapolation of trajectories. The limitations and assumptions of these models are explained in the following text and some of the potential ways to improve the accuracy of these models is also mentioned.

The aim of the thesis were: 1) to build statistical model(s) of the trajectory of a vehicle (POV) performing a critical cut-in in front of another vehicle; 2) to use the model for generating new trajectories representing a larger number of cut-in maneuvers. These trajectory models can be used as part of the development and evaluation of AD/ADAS systems (Bernard & Violette, 2012). To achieve the objectives of the thesis, first a dataset was prepared using annotated data collected in previous project (Shams El Din, 2020) and based on the SHRP2 naturalistic driving study (Transportation Research Board of the National Academies of Science, 2013).

In the previous project a total of 1191 cut-in event data were collected from SHRP2 out of which only 86 were found suitable for the study as mentioned in Section 3.1.2 of Shams El Din (2020). Out of this 86-events annotation, data was available for 57 events and this was used as the input dataset for this thesis. This data included kinematic variables of SV and POV and initial plots of the trajectories revealed that there were different shapes of trajectories with some similarities. Additional annotations were performed in this thesis project to provide a more complete description about the cut-in maneuvers. Wang et al. (2014) used a clustering method to identify similar shaped trajectories and identified that the subgroups of these shapes would have been explained with physical meaning if data regarding driver characteristics was available. In this thesis some of the driver characteristics were used as classification parameters and hence grouping of trajectories could be done without using clustering algorithms. As seen in Wang et al. (2014), clustering is useful for preliminary separation of similar shaped trajectories specifically for large dataset.

A sample of 20 generated trajectories using the polynomial trajectory model were shown in results Section 3.1.2.1. From these results it can be seen that the interpolated trajectory shapes follow an ideal shape compared to the input trajectories from the annotation tool. Since there is no ground truth trajectory data of the POV, it is difficult to comment on the accuracy of the generated data, by only using annotated data. However, the shapes of the interpolated trajectories are different from the annotation tool outputs because of the model assumptions stated in Section 2.2.2. One of the major drawbacks of polynomial model was the fixed lateral acceleration of POV whereas in original trajectories the acceleration of POV is

not fixed and changed depending on the environmental factors like surrounding vehicle positions. Also, the input variables of polynomial model (POV lateral acceleration, longitudinal displacement and velocity) have been modelled using linear regression. These linear models generated trajectories that do not fit well the annotated trajectories. Hence, more complex models can be considered in future research, with the aim to model these variables, which, in turn, likely will generate more accurate trajectory shapes. For example, to overcome the limitation of fixed acceleration of polynomial models, dynamic models could be explored (Xu, Liu, Ou, & Song, 2012). Xu et al. (2012) found that the dynamic lane change trajectory model had better fit compared to the polynomial model and one of the major difference in these models was that the lateral acceleration of the POV was not fixed and was modelled as a linear function of lateral velocity and lateral range.

The probability model was built to extrapolate new trajectories based on SHRP2 data. Looking at the results in section 3.1.2.2, the generated trajectories have different shapes during the start and end of the maneuver, compared to the original trajectories. This is because of the assumptions made on likelihood and prior distributions: that they are gaussian distributed, considering the coefficients of the quintic polynomial to be independent and selecting the noise variance through trial and error. Also, the assumption that the noise is a gaussian random variable made the variation of lateral position at each timestep random. In future research, the noise variable can be modelled using different distributions which can represent the uncertainty of measurement of the trajectory more accurately. Also, techniques using variational Bayesian methods has in previous literature been found to generate more accurate trajectory samples, in terms of their ability to represent actual human driving trajectories, from the distributions of trajectory parameters (Ding, Xu, & Zhao, 2020). Ding et al. (2020) developed a safety critical data synthesizing framework based on variational Bayesian methods. In their model, first the road map data is separated from trajectory data, then the trajectory sequences are encoded into latent space. In latent space the key parameters required to define the trajectory points are identified based on similar patterns and features of the trajectories using the convolutional later of a neural network. These latent space variables are then interpolated, and new trajectories are generated using a decoder. They used two input datasets, one with safe driving data and other with collision data. The data from these datasets was used to make the distributions of latent space variables using conditional probability and finally critical/near-miss trajectories were sampled from this mixed distribution. Ding et al. (2020) concluded that this model could generate realistic interpolated trajectory data in between the safe trajectory and collision trajectory. However, no direct comparison has been made with any naturalistic dataset of near-miss events. The probabilistic model in this thesis is directly trained on the critical cut in data occurring on similar road environment and hence a simple probabilistic regression model has been used which is less computationally expensive compared with models used by Ding et al. (2020). However, the shapes of the generated trajectories did not match well with the naturalistic trajectory data obtained from video annotation tool. An exact root cause for this has not been found in this thesis, but the video annotation tool uncertainty could have been a probable cause. Hence, though this model is computationally inexpensive, it does not produce accurate results when compared to Ding's model.

For further development of this probabilistic model feature reduction and latent space interpolation could increase the accuracy of the model.

4.1 Limitations & Future work

The initial plan for the thesis was to improve the annotation tool developed in previous project and to annotate more events and increase the available data. However, due to changes in the original plans during the thesis it was not possible to work further on the annotation tool, which resulted in limited availability of cut-in data. Hence, future work should increase the available annotation data, which will help to identify more categories of cut-in trajectories and build more statistically sound models.

Due to the limited time available, the longitudinal trajectory data of the POV could not be analyzed in detail, leading to the assumptions considered for building polynomial models. Due to this, the polynomial model developed in this thesis did not fit all the events as well as expected. This can be considered for future work which can improve the model fit for different events.

Initially the probabilistic model was also considered to build a generic trajectory model representing all the events in the data set. However, it was not possible because the trajectories of different events had substantially different durations and hence a generic model was not possible. Such a model can be considered in the future when more trajectory data will be available for each duration value. Hence, the probabilistic model was built for generating trajectories of an event within the uncertainty region caused by the annotation tool. In future research, approximation methods could be explored to improve the Bayesian model. Also, non-parametric supervised machine learning models like the gaussian process could be considered to build a predictive model of trajectories (Liu et al., 2020).

5 References

Bai, H., Shen, J., Wei, L., & Feng, Z. (2017). Accelerated Lane-Changing Trajectory Planning of Automated Vehicles with Vehicle-to-Vehicle Collaboration. *Journal of Advanced Transportation*, 1-11. doi:10.1155/2017/8132769

Bärgman, J., Boda, C. N., & Dozza, M. (2017). Counterfactual simulations applied to SHRP2 crashes: The effect of driver behavior models on safety benefit estimations of intelligent safety systems. *Accident Analysis & Prevention*, 165-180.

Bernard, J., & Violette, E. (2012). Vehicle Trajectory Analysis: An Advanced Tool for Road Safety. *Procedia - Social and Behavioral Sciences, 48*, 1805-1814. Retrieved from https://doi.org/10.1016/j.sbspro.2012.06.1155

Ding, W., Xu, M., & Zhao, D. (2020). Cmts: A conditional multiple trajectory synthesizer for generating safety-critical driving scenarios. *IEEE International Conference on Robotics and Automation*, 4314-4321.

Feng, G., Jianli, D., Yingdong, H., & Zilong, W. (2019). A Test Scenario Automatic Generation Strategy for Intelligent Driving Systems. *Mathematical Problems in Engineering, vol. 2019*, 10. doi:10.1155/2019/3737486

Gipps, P. G. (1986). A model for the structure of lane-changing decisions. *Transportation Research Part B: Methodological*, 403-414.

Hannan, M. A., Hussain, A., & Samad, S. (2010). System Interface for an Integrated Intelligent Safety System (ISS) for Vehicle Applications. *Sensors*, 1141-1153.

Hibberd, D., Louw, T., Aittoniemi, E., Brouwer, R., Dotzauer, M., Fahrenkrog, F., . . . Zerbe, A. (2018). From Research Questions to Logging Requirements: L3Pilot Deliverable D3.1.

Hou, Y., Edara, P., & Sun, C. (2015). Situation assessment and decision making for lane change assistance using ensemble learning methods. *Expert Systems with Applications*, 3875-3882.

Jin, H., Duan, C., Liu, Y., & Lu, P. (2020). Gauss mixture hidden Markov model to characterise and model discretionary lane-change behaviours for autonomous vehicles. *IET Intelligent Transport Systems, Volume 14, Issue 5*, 401-411.

Lee, J., Park, M., & Yeo, H. (2016). A probability model for discretionary lane changes in highways. *KSCE J Civ Eng 20*, 2938–2946. Retrieved from https://doi.org/10.1007/s12205-016-0382-z

McNaughton, M., Urmson, C., Dolan, J. M., & Lee, J. W. (2011). Motion planning for autonomous driving with a conformal spatiotemporal lattice. *IEEE International Conference on Robotics and Automation*, 4889-4895.

Moridpour, S., Majid, S., & Geoff, R. (2010). Lane changing models: A critical review. *Transportation Letters: The International Journal of Transportation Research, Vol.02*, 157-173.

Murthy, A. S., & Bharadwaj, P. S. (2020). *Human Behaviour-Based Trajectory Planning for Autonomous Overtaking Maneuver.* Gothenburg: Chalmers tekniska högskola / Institutionen för mekanik och maritima vetenskaper.

NHTSA. (2021). *Automated vehicles safety*. Retrieved from www.nhtsa.gov: https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety

Pakgohar, A., Tabrizi, R. S., Khalili, M., & Esmaeili, A. (2011). The role of human factor in incidence and severity of road crashes based on the CART and LR

regression: a data mining approach. *Procedia Computer Science, 3*, 764-769.
Peng, T., Su, L., Zhang, R., Guan, Z., Zhao, H., Qiu, Z., & Xu, H. (2020). A new safe lane-change trajectory model and collision avoidance control method for automatic driving vehicles. *Expert Systems with Applications*, 141.
Rogers, S., & Girolami, M. (2016). *A first course in machine learning.* Chapman and Hall/CRC.
SAE International. (2021). *SAE J3016.* Retrieved from SAE: https://www.sae.org/binaries/content/assets/cm/content/blog/sae-j3016-visual-chart_5.3.21.pdf
Schagen, I., & Sagberg, F. (2012). The Potential Benefits of Naturalistic Driving for Road Safety Research: Theoretical and Empirical Considerations and Challenges for the Future. *Procedia - Social and Behavioral Sciences, Vol. 48*, 692-701.
Shams El Din, A. (2020). *Statistical modelling of critical cut-ins for the evaluation of autonomous vehicles and advanced driver assistance systems.* Gothenburg: Chalmers tekniska högskola / Institutionen för mekanik och maritima vetenskaper.
Shimoda, S., Kuroda, Y., & Iagnemma, K. (2007). High-speed navigation of unmanned ground vehicles on uneven terrain using potential fields. *Robotica*, 409-424.
Stockholm Declaration. (2020, 02 20). *Stockholm Declaration.* Retrieved from Third Global Ministerial Conference on Road Safety: Achieving Global Goals 2030: https://www.roadsafetysweden.com/contentassets/b37f0951c837443eb9661668d5be439e/stockholm-declaration-english.pdf
Toledo, T., Koutsopoulos, H., & Ben-Akiva, M. (2003). Modeling Integrated Lane-Changing Behavior. *Transportation Research Record Journal of the Transportation Research Board*. Retrieved from https://www.researchgate.net/publication/228904295
Toledo-Moreo, R., Zamora-Izquierdo, M. A., & Gómez-Skarmeta, A. F. (2007). Multiple model based lane change prediction for road vehicles with low cost GPS/IMU. *IEEE Intelligent Transportation Systems Conference*, 473-478.
Trafikverket. (2020, 01 21). *This is Vision Zero.* Retrieved from https://www.trafikverket.se/en/startpage/operations/Operations-road/vision-zero-academy/This-is-Vision-Zero/
Transportation Research Board of the National Academies of Science. (2013). Retrieved from The 2nd Strategic Highway Research Program Naturalistic Driving Study Dataset.: Available from the SHRP2 NDS InSight Data Dissemination web site: https://insight.shrp2nds.us.
Virginia Tech Transportation Institute. (2020). *Data*. Retrieved from Insight SHRP2 NDS: https://insight.shrp2nds.us/data/index
Virginia Tech Transportation Institute. (2020). *Event detail table*. Retrieved from Insight SHRP2 NDS: https://insight.shrp2nds.us/data/category/events#/table/38
Wang, Q., Li, Z., & Li, L. (2014). Investigation of Discretionary Lane-Change Characteristics Using Next-Generation Simulation Data Sets. *Journal of Intelligent Transportation*, 246-253.

Wang, X., Yang, M., & Hurwitz, D. (2019). Analysis of cut-in behavior based on naturalistic driving data. *Accident Analysis and Prevention 124*, 127-137.

WHO. (2019). Retrieved from The top 10 causes of death: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

World Health Organization. (2018). *Global status report on road safety: Summary.* Geneva. Retrieved from https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/

Xu, G., Liu, L., Ou, Y., & Song, Z. (2012). Dynamic modeling of driver control strategy of lane-change behavior and trajectory planning. *IEEE Transactions on Intelligent Transportation Systems*, 1138-1155.

Yang, H. -H., Peng, H., Gordon, T. J., & Leblanc, D. (2008). Development and validation of an errorable car-following driver model. *American Control Conference*, 3927-3932. doi:10.1109/ACC.2008.4587106

Yang, Q., & Koutsopoulos, H. N. (1996). A microscopic traffic simulator for evaluation of dynamic traffic management systems. *Transportation Research Part C: Emerging Technologies*, 113-129.

Zhang, F., Gonzales, J., Li, S. E., Borrelli, F., & Li, K. (2018). Drift control for cornering maneuver of autonomous vehicles. *Mechatronics*, 167-174.

Zhao, D., Lam, H., Peng, H., Bao, S., LeBlanc, D. J., Nobukawa, K., & Pan, C. S. (2017). Accelerated Evaluation of Automated Vehicles Safety in Lane Change Scenarios based on Importance Sampling Techniques. *IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 3*, 595-607.

Zhao, D., Peng, H., Bao, S., Nobukawa, K., LeBlanc, D., & Pan, C. (2016). Accelerated evaluation of automated vehicles using extracted naturalistic driving data. In *The Dynamics of Vehicles on Roads and Tracks* (pp. 287-296).