CHALMERS

Quantifying Drivers' Behaviours when Overtaking Bicyclists on Rural Roads

A Study Using Naturalistic Driving Data from a Vehicle's Perspective

Master's thesis in Applied Mechanics

GUSTAV NERO

Department of Applied Mechanics
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2017

MASTER'S THESIS IN APPLIED MECHANICS

Quantifying Drivers' Behaviours when Overtaking Bicyclists on Rural Roads

A Study Using Naturalistic Driving Data from a Vehicle's Perspective

GUSTAV NERO

Department of Applied Mechanics
Division of Vehicle Safety
CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2017

Quantifying Drivers' Behaviours when Overtaking Bicyclists on Rural Roads
A Study Using Naturalistic Driving Data from a Vehicle's Perspective
GUSTAV NERO

Master's thesis 2017:71
ISSN 1652-8557
Department of Applied Mechanics
Division of Vehicle Safety
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone: +46 (0)31-772 1000

Cover:
"Rural Overtaking" Digital Art by Gustav Nero

Chalmers Reproservice
Göteborg, Sweden 2017

Quantifying Drivers' Behaviours when Overtaking Bicyclists on Rural Roads
A Study Using Naturalistic Driving Data from a Vehicle's Perspective
Master's thesis in Applied Mechanics
GUSTAV NERO
Department of Applied Mechanics
Division of Vehicle Safety
Chalmers University of Technology

ABSTRACT

Overtaking manoeuvres of vulnerable road users on rural roads have previously been found to be accident-prone events with severe or fatal outcomes for the involved vulnerable road users. Most of the previously conducted studies revolving around overtaking manoeuvres of vulnerable road users on rural roads have used either naturalistic driving data which was collected from the bicyclists' perspective or data from driving-simulators. The work described in this thesis involved quantifying data of cars overtaking vulnerable road users on rural roads. The thesis exemplified how to extract overtaking manoeuvre segments from naturalistic driving data and further demonstrated how to extract comfort zone boundaries. The data used was extracted from the database of the European Naturalistic Driving project UDRIVE and included CAN-data, Mobileye-data, and video data. The video data came from cameras capturing both the inside and the outside of the ego vehicle. The data was enriched via manual annotations and automatic derivation of signals using tools such as SALSA and MATLAB. Manual annotations also verified the data, since not all data extracted from the database contained overtaking manoeuvres. To keep the work manageable only events where a single vulnerable road user traveling in the same direction as the ego vehicle and where the vulnerable road user was in the outer-most lane as the ego vehicle were considered. The focus of the thesis has been method development, that is by primarily using ME data identifying overtaking segments from the UDRIVE database and then derive comfort zone measures such as time to collision, lateral clearance and minimum distance. However, due to issues with A) Subjectivity of the video annotations, B) the lack of a comprehensive quality check of the data (i.e. not comparing what the video-feed showed with what various signals implied), and C) error in a derived measure (i.e. the speed of the vulnerable road user), which in turn was used in several other derived measures, results were only compared to previous studies briefly. For future work the quality of the data (both raw and derived) should be considered to have a higher priority. In other words, a more comprehensive data validation should be performed to verify that extracted data is fit for analysis.

Keywords: Active Safety, Driver Behaviour, Comfort Boundaries, Overtaking, Rural Road, Cycling Safety, Naturalistic Driving Data

SAMMANFATTNING

Omkörningar av sårbara trafikanter på landsvägs har pekats ut som en olycksbenägen manöver med allvarligare eller dödliga konsekvenser för inblandade sårbara trafikanter. De flesta tidigare genomförda studier om omkörningar av sårbara trafikanter på landsväg har antingen använt naturalistisk data insamlad via cyklar utrustade med olika sorters sensorer eller så har de använt data insamlad i körsimulatorer. Arbetet beskrivet i denna rapport involverade kvantifiering av data från bilar som kör om sårbara trafikanter på landsväg men där datan var insamlad från bilistens perspektiv. Rapporten exemplifierar hur omkörnings segment kan extraheras från naturlistisk kör data och demonstrerade hur säkerhetsmått kan tas fram. Datan var extraherad från databasen tillhörande det europeiska projektet UDRIVE och inkluderade CAN-data, Mobileye-data och ett videoflöde som bestod av flera olika kameravinklar både i kabinen i bilen och på väg-miljön utanför bilen. Datan var ytterligare berikad via manuella annoteringar samt automatiskt härledda signaler via verktyg som SALSA och MATLAB. Annoteringer bidrog även till att verifiera datan, då inte all data som extraherades från databasen innehöll omkörningar. För att hålla arbetet inom hanterbara mått så användes bara omkörnings-manövrar som inkluderade enskilda cyklistar som hade samma färdriktning som fordonet och som dessutom befann sig i den yttersta körbanan. Metodutveckling har varit fokus i denna studie, mer specifikt hur identifering och extrahering av omkörningar av sårbara trafikanter på landsväg från en databas innehållande naturalistisk kördata kan genomföras, men på grund av A) Oklara instruktioner till annotörer (vilket innebar en risk för subjektiva klassifikationer), B) en avsaknad av en good kvalitetskontroll på datan (med andra ord ingen utförlig jämförelse mellan videoflöde och signaler gjordes), och C) ett fel i en härledd signal som påverkade ett antal andra betydelsefulla härledda signaler så genomfördes endast en ytlig jämförelse med tidigare genomförda studier. För framtida arbete så rekommenderas starkt att sätta högre prioritet på datakvaliteten (både rådata och härledd data) och därmed även en mer utförlig kvalitetskontroll av datan för att säkerställa att den är användbar till diverse analytiska jämförelser.

PREFACE

ACKNOWLEDGEMENTS

This master thesis was carried out at SAFER premises in Gothenburg

First and foremost, I would like to express my appreciation to my primary supervisor, Dr. Jonas Bärgman, for the useful comments and remarks which he provided in a calm manner throughout the course of this master thesis. Whenever I felt lost he guided me back on track.

I must express my gratitude towards my secondary supervisor, Jordanka Kovaceva, for her outstanding positive attitude and her willingness to help.

I would also like to thank Fabio Forcolin who did not hesitate the slightest to answer even the simplest (and dumbest) questions I had about the tools used in this thesis.

I must also express my appreciation for my fellow colleagues at Chalmers University of Technology and the staff of SAFER for providing a pleasant and sound environment to work and learn.

Finally, I would like to thank my wonderful mother Eva for her unconditional love, my generous and hardworking father Thomas for his financial support, and last but absolutely not least my late best friend Sture (who also happened to be my grandfather) for teaching me that showing appreciation makes the world a better place.

LISTS OF FIGURES AND TABLES

List of Figures

List of Tables

CONTENTS

1 Introduction

1.1 Background

Ask children what is essential when driving a car in traffic and the responses likely are related to what they have perceived from the back seat; pressing the foot pedals, yanking the gear stick and turning the steering wheel. Maybe even some honking and yelling. It is a good answer by all means, but not everything revolves around the actions within the vehicle. There is a constant interaction with other road-users, which mostly does not take the form of honking and yelling. Road-users communicate via other means, e.g. lane position, speed and the direction indicators in an attempt to keep traffic organized and avoid accidents. This applies to both vehicles and vulnerable road users (VRU) [12]. However, accidents do happen.

While the total number of road crashes in Europe is declining, the number of accidents involving bicyclists is not descending at a similar rate [32]. A majority of the accidents happen when the bicyclist is travelling in the same direction as the motorist, that is not facing each other [14]. Lane position, speed and the direction indicators are primarily visual signals. Whereas the lack of blinkers on a bicycle easily can be replaced by a raised arm, there is no real replacement for a rear mirror. No way to perceive what the intentions of the road-users approaching from behind are. Roads with high road speed limit and presence of VRU, e.g. rural roads, are prone to these accidents [20]. When travelling on rural roads the bicyclist can only hear the approaching vehicle's presence and await the execution of an overtaking-manoeuvre. The bicyclist must trust the motorist to perform the overtaking manoeuvre in a safe manner. Due to the combination of a large speed difference and the VRU more or less being unprotected, accidents related to overtaking-manoeuvres on rural roads tend to end in severe or even fatal injuries for the VRU [3].

1.2 Purpose and Goals

Whereas several studies has examined overtaking-manoeuvres of VRUs from naturalistic data collected via instrumented bicycles [35] [24] [8] [28] [9] [15], studies based on naturalistic driving data from motor-vehicles are sparse. This study aims to contribute to the field of traffic safety by studying naturalistic driving data collected from cars in the UDRIVE (EU FP7) project. In the small scope, this study aims at presenting a methodology to identify and extract bicycle overtaking manoeuvres in the UDRIVE database and derive comfort zone measures from these. In the big scope, by acquiring knowledge, in the form of comfort zone boundaries and other measures that can be used to understand how motorists perform overtaking-manoeuvres of VRUs and choose their safety margins, improvements can be made to road safety in the form of enhanced technology, e.g. driver assistance and autonomous vehicles, as well as guidelines (e.g. road design) and policies. However, knowledge is processed information. Thus, while understanding how events such as overtaking-manoeuvres of VRUs unfold is an important step in the process of making bicycle overtaking safer, there is a step that precedes it; Getting the relevant information, or more specifically for his study the extraction of data which contains the information about the overtaking manoeuvre of bicyclists on rural roads from the UDRIVE database.

1.3 Problem Definition

Problems that will have to be solved in this study are

- How to extract manageable data in the SALSA environment?
- What signals and properties should be used to enable accurate filtering and adequate data reduction?
- How to derive measures from raw data signals?
- How to analyze overtaking manoeuvres from naturalistic driving data?
- How to interpret results from estimations of safety measures such as comfort zone boundary and time to collision?
- How does 'road/traffic conditions', e.g. presence of oncoming traffic or leading vehicle, affect safety measures?

- How to compare results with previously conducted studies?

1.4 Limitations

This study focuses on cars overtaking a single bicyclist when they are travelling in the same direction and the outermost lane on a rural road. In addition there were some technical limitations related to the tools and data used, namely:

1. The SALSA environment has a limited flexibility when it comes to on-line analysis and data need to be extracted with a slow process that causes delay in the analysis.
2. UDRIVE data is rather novel and the quality of the data is to a certain degree undocumented requiring the analyst to implement and perform a sanitation process, again time-consuming and delicate (a small mistake may result in a biased sample for analysis).

Due to the time-frame of this study being 4 months and reprocessing the database in SALSA environment happens once a week (at best) the development of data-reduction-algorithms, i.e. filters, is fairly constrained time-wise in the sense that the feedback on how changes made in the algorithms affect the database are often delayed. The purpose of the filters was to narrow down hours upon hours of driving data to segments where overtaking-manoeuvres of VRUs were performed. Ideally the filtering should have a perfect hit-rate with zero false-positives and false-negatives, however this is likely impossible to achieve due to the simplistic nature of the algorithms used in this work (and likely with any algorithm). Additionally, the data is not 100% reliable. For instance the classification of objects in the Mobileye-data (ME) has a confidence value associated with them, implying that the signals have a degree of uncertainty -which also was confirmed reviewing the data. Furthermore, although a minor limitation, the units for the ME-data were missing. For instance, it was not known whether speed was displayed in miles per hour, kilometers per hour, meters per second or knot etc.

1.5 Outline

Chapter 2 – Literature Review Introduces research previously conducted in the fields of active safety and accident prevention.

Chapter 3 – Approach and Methodology Outlines the methodologies used in each phase of this project and gives implementation details of the corresponding methods respectively.

Chapter 4 – Experimental Results Gives an overview of the results obtained by the methods, mainly through various tables and figures.

Chapter 5 – Discussion Provides a comprehensive discussion where advantages and disadvantages for each part of this project are more thoroughly evaluated.

Chapter 6 – Conclusions & Future Work Gives a concluding summary of the author's thoughts on the obtained results and how the initial problems were solved. As well as some suggestions for future work.

2 Literature Review

In this section past research on the topic of motorists' interaction with VRUs is discussed. This includes both studies based on accident data records and (quasi-)naturalistic data. The literature discussed provided a general insight into why overtaking manoeuvres of VRUs are interesting and what variables in such overtaking manoeuvres usually are examined.

2.1 Defining the overtaking

What is an overtaking? Querying various dictionaries for a definition gives results such as "drive past someone or something" [31], "to come from behind another vehicle or a person and move in front of them" [4] and "Catch up with and pass while travelling in the same direction" [23]. These definitions provide a good grasp of what an overtaking means in the context of traffic actions. However, there are plenty of ways such a maneuver can be executed. For example, one vehicle overtaking another vehicle, one vehicle overtaking a multiple of other vehicles, as well as a vehicle overtaking a vehicle which in turn is overtaking a third vehicle [26]. Additionally, the overtakings can be differentiated by how they are performed; The flying overtaking where the overtaking vehicle's speed remains constant or near constant during the overtaking and the accelerating overtaking where the overtaking vehicle follows behind the soon-to-be-overtaken vehicle and by acceleration increases its speed to enable the overtaking. If one vehicle is overtaken by a multiple of other vehicles continuously in a row, the first overtaking vehicle is considered to be the lead vehicle, while the overtaking vehicle(s) behind the lead vehicle is considered to be performing a piggy backing overtaking. The lead vehicle adopts either the flying or accelerative strategy [19]. In this study the piggy backing strategy is considered to be a attribute of the accelerative- and flying overtaking manoeuvre strategies, in the same manner that the presence of oncoming traffic is.

How such a maneuver is executed can be further dissected, mainly by dividing the maneuver into a number of defined time-frames, that is phases (any further mentioning of phases refers to these time-frames). There have been numerous proposals on how to divide the overtaking maneuver, both regarding the number of phases and how to distinguish them. Examples of such proposals includes division into three phases (divert from lane, drive straight in adjacent lane, and return to lane) [30], four phases (approaching, pull out, passing, and returning) [9] and five phases (decide whether or not to overtake, prepare for overtaking, change lane, pass, and return to initial lane) [11]. The three-phase classification presented in the Shamir [30] paper is primarily intended for a motorist overtaking another motorist, whereas the classification in the Dozza et al. [9] paper is customized for a car overtaking a bicycle. Another categorization designed for the car overtaking bicycle case is the three-phase-classification used in the Chuang et al. [8] paper (before passing, while passing, and after passing), defining the first and third phases as time-offsets (±0.5 seconds) of the second phase. For this study the classification used is the one established in the Dozza et al. [9] paper, mainly to maintain continuity and allow for comparison with Master Theses previously conducted by Schindler and Bast [28], Fatahtooei Nejad [10] and Moretto [21]. The phases used in the Dozza et al. [9] paper were as previously mentioned

Phase 1 - Approaching phase; Starts the vehicle reaches the bicycle (within reading range for the sensor) from behind and ends when phase 2 begins.

Phase 2 - Steering Away/Pull Out phase; Starts when the vehicle start to steer away from the collision path and ends when phase 3 begins.

Phase 3 - Passing phase; Starts when the front of the vehicle is 2 meters behind the rear of the bicycle and ends when phase 4 begins.

Phase 4 - Returning phase; Starts when the rear of the vehicle is 2 meters in-front of the front of the bicycle and ends when the vehicle returns to the lane-position it had before the overtaking-manoeuvre was performed.

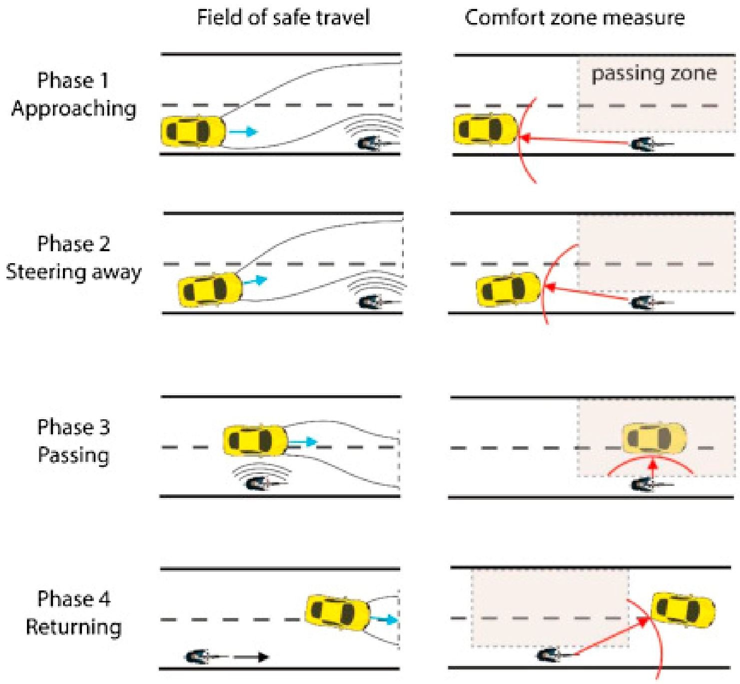


Figure 2.1: *The phases of an overtaking as defined by Dozza et al. [9] as well as a visualization of the comfort zone boundary.*

Source: Dozza et al. [9]. Used with authors' permission.

As can be seen in figure 2.1 there is also something called comfort zone measure, which also can be referred to as comfort zone boundary (CZB). The CZB can be viewed as the limit of where the driver is comfortable in everyday normal driving with regards to fellow road-users and the infrastructure. The CZB is dynamical and is constantly influenced by factors spanning from driver characteristics (e.g. risk-perception) to context (e.g. presence of oncoming traffic) [1]. However, it is worth noting that some slight adjustments were made in this study regarding how to determine the start and end of the phases, but that is covered in section 3.3.

2.2 Overtaking a VRU

Cars, buses and lorries. What do they all have in common? They all share the road with the least protected and most vulnerable; Bicyclists and pedestrians [34]. In urban settings accidents between a motorist and a bicycle most frequently occur in intersections [14] [3]. However, the severity of injuries caused in urban setting is fairly low when compared to that of rural setting [20]. For example a study conducted in Victoria, Australia, found urban accidents almost 19 times as common as rural ones, but rural accidents were roughly 25% more likely to result in severe injuries [3]. Furthermore, a study ran in North Carolina, USA, found that approximately 80% of the accidents occurred when the motorist and the bicyclist were sharing the same travel lane. Additional conclusions were that overtaking manoeuvres were the second most accident-prone motorist-bicycle-interaction (at 11.5%, while the most frequent, at 23%, was bicyclist turning or merging) and the severity of the injuries was greater when the bicyclist was at fault [14]. However, in Czech Republic a study concluded the opposite of the latter; The motorist was at fault in a majority (57%) of the accidents that lead to serious injuries and that these accidents were 1.62 times as likely to have a fatal outcome [2].

A common CZB measure is the minimal lateral distance between vehicle and VRU during the passing phase of an overtaking [35] [8], however the Dozza et al. [9] paper used the minimum distance between car and VRU during all phases as a measure for the drivers' CZB and found that the steering away, passing and returning phase were highly correlated. During overtaking manoeuvres of bicyclists the motorists have been found to leave an unsatisfying lateral clearance [15]. Thus, drivers seem to adjust their CZB when adjacent traffic is present by passing closer to the VRU [9]. Car-drivers have been found to perceive car-bicyclist-interactions less risky than the bicyclists [7]. With increasing speed and faster overtakings the bicyclist expects a greater lateral clearance, while the motorist instead is more influenced by visibility and adjacent traffic [9] [13]. If the bicyclists do not feel safe in traffic they may avoid cycling altogether as a coping strategy [6] and it does not matter whether the perceived feeling of being at risk is objectively accurate or not [38]. In a study where observers were static, predictions about bicyclists intent were found to depend on the bicyclist's head movement, position and speed [12]. However, for a motorist the bicyclist's speed is hard to perceive correctly and the the motorist's estimation of how long it takes for the longitudinal distance between vehicle and bicyclist to reach zero, that is time to collision (TTC), is influenced by pedaling frequency [29]. The motorist's behavior is also affected by the appearance of the bicyclist. Worn outfit suggesting a certain skill- and/or experience-level has been found to not affect the lateral distance [36]. However, the lateral distance between motorist and bicyclist increases when the bicyclist appears to be female [8], does not wear a helmet and is positioned closer to the shoulder/edge. Additionally, the type of vehicle also affects the lateral clearance, for instance large vehicles with low-acceleration passes closer to the bicyclists [35]. In a re-analysis of the data used in the Walker [35] paper Olivier and Walter [22] found that while helmet usage did affect the lateral passing distance, it did so in events where the distance gained was negligible. Other factors influencing the motorists' behavior during overtaking of bicyclists have been shown to be of infrastructural kind, such as road grade, road design speed, presence and width of the shoulder as well as the marked centerline [5]. The presence of bicycle facilities (e.g. on-road bike routes, on-road marked bike lanes, and off-road bike paths) has been shown to be associated with the lowest risk for bicyclists [25]. However, in the presence of a bicycle lane while the motorists ride within the constraints of their own marked lane, they may neglect the need of a comfortable passing distance to the bicyclists [24]. Additionally, the occurrence of oncoming traffic severely reduces the likelihood of an overtaking vehicle contacting the marked centerline [27] and thus decreases the lateral clearance [9].

Although studies has taken place in a diversity in geographical settings such as United Kingdom [35], Spain [15], North Carolina [14], Wisconsin [5], Taiwan [8], Australia [3][20], France [7] and Sweden [9][28], they all seem to conclude the same things, for instance studies based on accident statistic find that due to the high speed on rural roads and the presence of VRU the risk of serious or fatal injuries for the VRU are increased or that driver behaviour during overtakings is greatly affected by oncoming traffic the in adjacent lane [27] [9] . One disagreement is the previously mentioned difference in findings between the Kim et al. [14] paper and Bíl, Bílova, and Müller [2] paper. Whether or not a driver was reported to be at fault for an accident might be caused by the different geographical setting, culture and jurisdiction or simply by coincidence. However, that difference in findings does not influence this work due to nature of the studies. This study looks at how overtaking-manoeuvres are performed, not who is at fault when accidents occur.

2.3 Data Acquisition in Previously Conducted Studies

The studies performed by Walker [35], Parkin and Meyers [24], Chuang et al. [8], Schindler and Bast [28], Dozza et al. [9] and Llorca et al. [15] all used bicycles equipped with some sort of proximity measurement unit such as ultrasonic sensor, laser or LIDAR to collect data when overtaken by motorists. Schindler and Bast [28] also used an instrumented car. Other approaches includes extracting data from videos such as in the Hegeman, Brookhuis, and Hoogendoorn [11], Savolainen et al. [27] and Love et al. [16] studies or using simulations such as the Fatahtooei Nejad [10] and Moretto [21] studies.

3 Approach & Methodology

A block diagram of the methodology utilized in this work is presented in figure 3.1. The first step was to find where a VRU-objects disappears from the data and extracting the context around these disappearances. The next step was to evaluate if the disappearances where caused by the driver performing an overtaking, resulting in the segments being either classified as 'overtaking' or 'non-overtaking'. The segments classified as 'overtaking' were then manually verified via annotations. Additionally, CZB measures were derived from the data for each phase for the verified overtaking segments. Finally, the CZB was briefly compared to measures found in a previously conducted study which was based on simulator driving data.

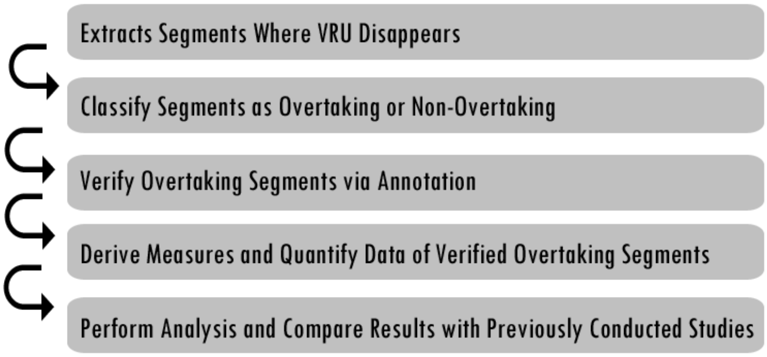


Figure 3.1: *Overview of methodology.*

3.1 Tools

In order to make analysis easier to perform on a large data, such as the database of UDRIVE, software tools which provides solutions to computer science challenges related to handling big data and processing of data are necessary.

3.1.1 MATLAB

MATLAB® is a software program and programming language developed by the corperation MathWorks® based in Natick, Massachusetts, United States. MATLAB provides an optimized environment to solve engineering problems and visualize data. For further information about MATLAB please refer to its documentation [17].

3.1.2 SALSA

Smart Automation for Large data Sets Analysis, SALSA, is a software tool developed by the non-profit organization Centre European Studies Safety And Analysis Des Risques, CEESAR, based near Paris in France. The tool was developed specifically for the UDRIVE-project and it runs in the MATLAB environment. Within SALSA it is possible to filter and query for data, visualize data, calculate new measures based on the acquired data and add annotations to data.

One of the advantages of using SALSA is that it provides an interface for performing the tasks previously mentioned (i.e. filter and query data etc.), however a drawback is that it (at least at time of writing) is impossible to evaluate MATLAB scripts/algorithms on a subset of the database within SALSA. The scripts/algorithms can be applied manually on individual data entries, whereas for them to be applied on the full database requires an update which runs all scripts/algorithms associated with the database. This is costly in time and thus for practical reasons can only be scheduled once a week which results in the development/analysis process being delayed.

The data model of SALSA as visualized in figure 3.2 where

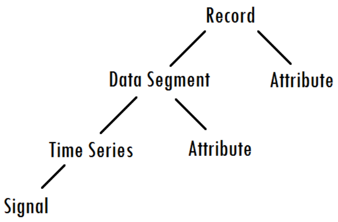


Figure 3.2: *Data model of SALSA*

Record - An entity in the database is referred to as a 'record', which represents a recording session of a vehicle. A record has a set of particular attributes associated to it including, but not limited to the date and time of the beginning and the end of the recording session and a unique 'Record Name'.

Segment - A record can also be decomposed into multiple segments. Segments which might encompass a certain type of event, for example overtaking manoeuvres. In other words, a segment is a fragment of a record defined by a beginning and an end time-stamp. Furthermore, two segments which encompass the same event can not overlap, that is the beginning time-stamp must be followed by an end time-stamp and not by an additional beginning time-stamp. Additionally, a segment can have its own user-defined attributes and time series.

Attribute - An attribute is a user-defined property of either a record or a segment which takes the form as a string, a Boolean or numerical scalar, or a reference; A reference is a user-defined table that holds a column with a numerical or logical values and a column with corresponding labels. For instance '0' and 'not present', which is a logical value and a label respectively.

Time Series - A time series is a sequence of time-stamps sorted in ascending order stored in a segment. It contains said time vector as well as its own signals.

Signal - Unlike an attribute, a signal is a time-dependent vector of data. It is stored in time series, meaning that each time-stamp in a time series has an assigned signal value.

For further information about SALSA please refer to UDRIVE Deliverable D242.1 [33].

3.2 Raw Data Available

The database consisted of records, and each of these records had a wide variety of attributes, signals and user-defined properties.

Mobileye Data ME-data is based on image processing. The ME-data included tracking of up to eight objects (of which up to four could be road users) on the road, as well as some other aspects such as road type. The data of each object included measures such as lateral distance, longitudinal distance, relative velocity, angle, type of object (car, truck, pedestrian, bicyclist, etc) etc. The ME have a limited field of view of approximately 50 degrees, that is 25 degrees left and fright from the middle of the vehicle, in front of the vehicle.

CAN Data CAN-data includes measures related to the ego vehicle such as velocity, steering wheel angle, pressure on pedals etc.

GPS Data In addition to only being GPS coordinates, there was map-matching data which contains information which can be derived from the location such as road-speed-limit.

Video Feed In addition to these attributes and signals there was also a video-feed displaying different sections of the vehicle such as the view in front of the vehicle (middle), the view in front of the vehicle (to the left), the cabin of the vehicle, the driver and steering wheel etc. This video feed did not include the video feed that the ME-data based on.

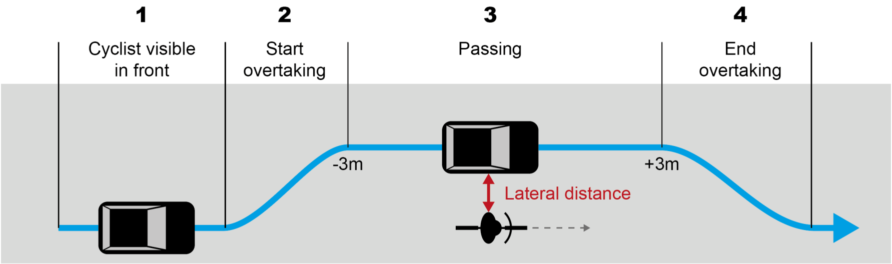


Figure 3.3: *Visualization of the phases of an overtaking.*

Source: Reinier Jansen, 2017 ©. Used with permission.

3.3 Defining the Phases of the Overtaking Manoeuvre

As mentioned earlier there were some slight modifications done to how to define the phases, when compared to the classification used in the Dozza et al. [9] paper. This was mainly due to there being a different set of signals available, observing the events from a different perspective. More specifically instead of LIDAR-data, there is CAN-data and ME-data which was collected from a motor-vehicle instead of a bicycle. Additionally, the time-resolution of the signals ($\leq 10Hz$) meant that the distance covered in between two samples when travelling at 70 km/h was roughly two meters. Seeing as a bicycle is often less than two meters long in a worst case scenario the phase three could last for one sample. In order to allow more data to be classified as phase three while still maintaining the concept of the phase three, the conditions for when phase three start and end were changed from two meters in longitudinal difference to three meters in longitudinal distance as explained in the list below and as depicted in figure 3.3.

Phase One - Approaching phase; Begins when the VRU is visible in the video-feed and ends when phase two begins.

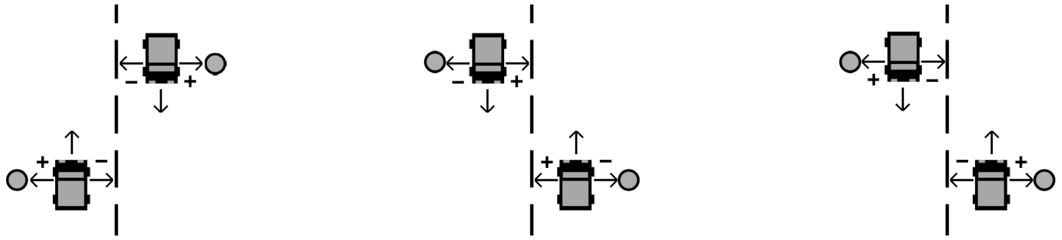
Phase Two - Steering Away/Pull Out phase; Begins when the vehicle start to steer away from the collision path and ends when phase three begins.

Phase Three - Passing phase; Begins when the rear of the VRU is less than three meters in-front of the rear of the ego vehicle and ends when phase four begins.

Phase Four - Returning phase; Begins when the front of the VRU is greater than three meters behind the rear of the ego vehicle and ends when the ego vehicle has travelled 50 meters since the start of phase four.

3.4 Extracting Context When a VRU Disappears

A signal that highlighted when a VRU disappear from the ME data, that is infront of the ego vehicle, was used as input, or more specifically the attributes begin-time and end-time of the segment, speed of the vehicle and longitudinal distance to the VRU. The output was a Boolean vector and a vector with corresponding timestamps. If the VRU disappeared within 50 m of the vehicle and the vehicle traveled above 20 km/h the conditions were met to extract the context. These thresholds were present mainly to remove instances where a VRU was disappearing far ahead, for instance a motorbike overtaking the vehicle, or a VRU disappearing when travelling slow, for example while the vehicle stood still at an intersection. Continuing, the context was set to be a offset of ten seconds before the begin-time and ten seconds a after the end-time. This was to make sure that the previously mentioned phases of the overtaking-manoeuvre were included. For example phase four ends when the ego vehicle has travelled 50 meters since the start of phase four and 50 meters in ten seconds is roughly equal to 20km/h, which was the previously mentioned threshold. If the context of two VRU disappearances overlapped they were merged.



(a) *Lateral Distance to VRU for Left-hand Traffic* (b) *Lateral Distance to VRU for Right-hand Traffic* (c) *Adjusted Lateral Distance to VRU for Right-hand Traffic*

Figure 3.4: *The sign of the lateral distance data collected refers to which side of the vehicle (grey rectangle) the object (grey circle) detected is, however this is not suitable when referring to attributes related to right- and left-hand traffic. In this example the obstacle is located towards the lane edge (notice that the travelling direction is indicated by the vertically-aligned arrows) and as can be observed the sign of lateral distance varies depending on if the data comes from GB or not. Fig 3.4c illustrates how data not collected in GB is adjusted to have reversed signs. This means that a positive lateral distance corresponds to that the obstacle is located towards the lane edge, whereas a negative lateral distance corresponds to that the obstacle is located towards the adjacent lane.*

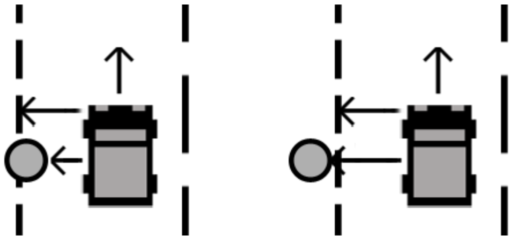
3.5 Finding Measurements of the Closest Valid VRU

Due to sheer amount of data available, that is data corresponding to hours upon hours of driving, it was necessary to reduce it as much as possible. To do this without losing instances where an overtaking-manoeuvre of a VRU was performed a filter-algorithm was implemented.

A data-vector containing Boolean values was created for each of the four objects that the ME-system is able to track. The values in the Boolean vectors was then set to true if the objects which the the vector was assigned to fulfilled certain criteria and false otherwise. These criteria were based on the objects' type and the road type (Boolean vectors which purpose is to identify when certain criteria are fulfilled will from here on be referred to as masks). The masks filter for when the objects were either bicyclist or pedestrian and the road type was either rural or country motorway. These masks were then applied to the corresponding objects' lateral and longitudinal data. However, for data collected in Great Britain (GB) a negative lateral distance meant that the object was located towards the adjacent lane whereas data not collected in GB negative lateral distance instead meant that the object was located towards the lane edge, that is the appearance of some data was related to whether it was collected in right- or left-hand traffic. For data not collected in the GB the sign of the lateral data was reversed in order to make the data consistent. Thus, a positive lateral distance meant that the object was located towards the lane edge, whereas a negative lateral distance meant that the object was located towards the adjacent lane. This is illustrated in figure 3.4. Since overtaking-manoeuvres occur when the VRU is located towards the lane edge negative values of the lateral distance were filtered out. In instances where more than one object was classified to be a VRU the minimum distance to the VRU was examined. The data corresponding to the closest VRU was then kept, namely the lateral distance, the longitudinal distance and the relative velocity to the closest VRU-object.

It is worth noting that the VRU objects were only possible to be considered as the closest object when in front of the ego vehicle, that is in detection range, and thus present in the raw data-signals. This meant that for instance if an overtaking of two VRUs occurred the first VRU was only possible to be closest as long as it was in front of the ego vehicle. When it was out of detection range its existence was forgotten in the sense that no extrapolation of its position took place (The extrapolation occurred later in the data processing, see section 3.9). For instance if a VRU was two meters behind the ego vehicle the closest VRU would still be considered to be a VRU that is ten meters in front of the vehicle.

A mask was created based on where the lateral distance to the VRU data existed. The distance to the lane shoulder data was then filtered with this mask resulting in a vector containing data-points only where data for the VRU was present.



(a) *VRU in lane* (b) *VRU outside lane*

Figure 3.5: *Scenarios in left-hand traffic. In figure 3.5a the distance from the vehicle (grey rectangle) to the road shoulder is lesser than the mean of distance from the vehicle to the VRU (grey circle) and thus considered in lane, whereas in figure 3.5b the distance from the vehicle to the road shoulder is greater than the mean of distance from the vehicle to the VRU and thereby considered to be out of the lane.*

3.6 Data Reduction - Classifying Segments as Overtaking Events

To determine whether the VRU was in the lane or outside the lane, a comparison between the mean of the distance to the road shoulder and the mean of distance to the VRU was made. If the former was greater than the latter the VRU was considered to be in the lane, as illustrated in figure 3.5.

Finally, segments which might be overtaking-manoeuvre events were declared. This was done by iterating through an interval defined by the first and last time-instance where the ME detects a VRU. If there were at least 15 data-points (one and a half seconds worth of data) fulfilling the criteria of lateral distance to lane edge being larger than the lateral distance to VRU and the VRU was considered to be located in the lane. This threshold was present due to that misclassification in the ME-data rarely lasted for longer than a few samples, however if several misclassification happened in a row the accumulated sum of values could sometimes exceed 15-data points and thus another threshold was also introduced. The number of invalid data in the interval had to be below 33% for the segment to still be considered an overtaking-manoeuvre candidate. The number 33% was chosen by trial and error to include as many actual overtaking segments as possible and exclude as many of the non-overtaking segments as possible.

Lastly, by identifying peaks in the lateral acceleration of the vehicle and then at those time-stamps examining the lateral distance to the VRU. More specifically, first the mean of six samples around the peak, excluding the peak itself, was computed for the lateral distance to the VRU. The number six was chosen somewhat arbitrarily, but its purpose was to catch instances where the lateral distance was increasing. Then the mean value of the lateral distance to the VRU for the segment was computed. These two mean values were then compared, if the former was greater than the latter it was considered to be an overtaking manoeuvre.

3.7 Adjusting the Distance to the Lane Signal

The data for the 'lateral distance to the lane edge' (D2Lane) and 'lateral distance to the adjacent lane' (D2ALane) signals were observed to sometimes have a default value of 1.880 m assigned to them for short periods of time or contain NaN values. To circumvent this issue every default value and NaN value were reassigned using spline interpolation [18]. The interpolation was used to mimic what the data might have looked like if it was not missing. However, the behaviours of the interpolation is very unpredictable when large gaps of data is missing. An example of this unpredictable behaviour in effect can be seen in figure 3.6 and 3.7. In these figures the distances to lane edges are peaking at roughly 30 meters due to the interpolation. In order to avoid these kind of instances, data entries which had 25% or more missing data points were removed. Note that some of the interpolations used in these examples are extreme, interpolating over several seconds. However, only the data close to the actual overtaking are used in the analysis. Future studies should impose limits on interpolated time duration to ensure high data quality.

However, it was observed that it was not uncommon for vehicles to enter the adjacent lane during an overtaking-manoeuvre and this disturbed the lane-tracking. For instance, in a right-hand traffic setting with a single lane in both travelling directions the signal for tracking the distance to the lane-edge will go from tracking the lane-shoulder to tracking the central line while the signal for tracking the distance to the adjacent-lane will

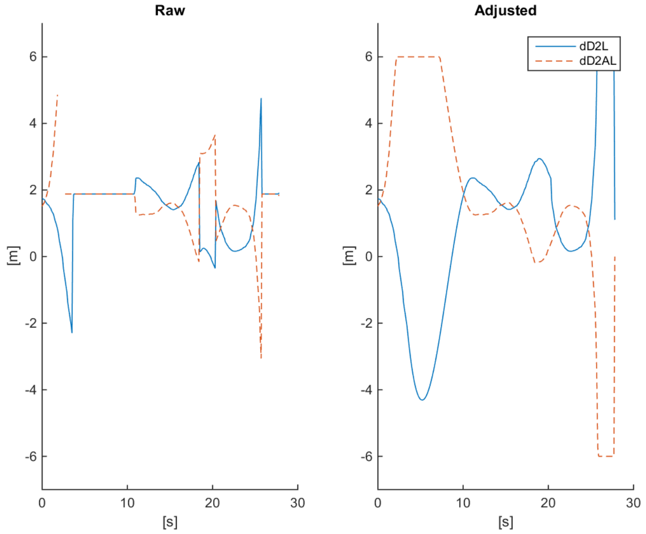


Figure 3.6: *A comparison between the raw data and the final adjusted data. The example uses only mutual peaks of the time derivative to adjust signal values.*

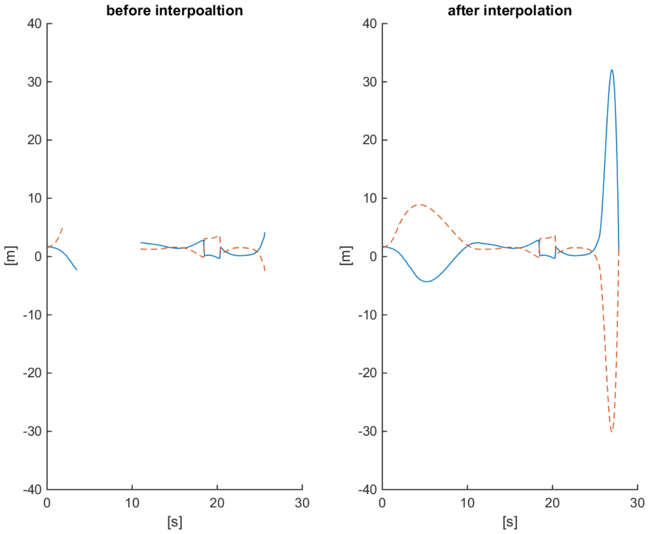


Figure 3.7: *NaN values and default values are interpolated. The example uses only mutual peaks of the time derivative to adjust signal values.*

go from tracking the central line to the adjacent lane's lane-shoulder. In other words, the distance to the road shoulder will increase until the ego vehicle changes lane and then the value gets close to zero (referred to as a "reset"). The reverse happens for the distance to adjacent lane, it will decrease and then when the lane-change occurs increase instantly to approximately its previous value. Thus, there was need to offset the values when a reset happened.

This was solved by finding offset-values. The offset should be distance between the lowest value and the highest value when a "reset" occurs. To locate where the "resets" occurs was done via looking at the time-derivative of the signals D2Lane and D2ALane. The time-stamps for peaks which exceeded a value corresponding to 150% of the lane-width were stored for both the lane edge and adjacent lane. Negative and positives peaks were also kept separate. A negative peak for the time-derivative of D2Lane (dD2L) indicated the start of a lane-change and positive peaks indicated the return to the original lane. The reverse was applied for the time derivative of D2ALane (dD2AL).

Next the start and end time-stamps for lane-changes were compared between dD2L and dD2AL. If a start or end occurred on the same time-stamp in both dD2L and dD2AL they were referred to as mutual, an example of this can be seen in figure 3.9. For a start point maintain its status as mutual there needed to exist a mutual end point which had a greater time-stamp. If the amount of mutual start points did not equal the amount of mutual end points adjustments the list of mutual start and end points was adjusted. If there were more mutual start points than end points only the first X start points were considered mutual, where X is the number of mutual endpoints. If instead there were more mutual end points than start points, only the Y last endpoints were considered mutual, where Y was the number of mutual start points. If a peak was not considered mutual, it was considered to be non-mutual. Next the non-mutual peaks in dD2L and dD2AL were looked at. For each non-mutual start-point the closest non-mutual end-point (with a higher time-stamp) was mapped, as can be seen in the example depicted in figure 3.11. This was done to dD2L- and dD2AL-peaks separately. Then, for each non-mutual start and end point pair the offset was calculated by taking the maximum and minimum value within two time-stamps around both the start- and the end-point. The range of five (X-2 to X+2 where X is the start- or end-point) was chosen due to the lane-change artifact in the signal, occurring in less than five time-stamps. The minimum value was subtracted from the maximum value to create an offset. Thus, for each start and end point pair there were two offset-values. The biggest of these offset-value was then applied to the interval between the start and end point in the corresponding signal. For the D2Lane it was added, whereas D2ALane the offset was subtracted. The process for calculating and applying offsets was then repeated for the mutual start and end points. Examples of these methods can be seen in figure 3.8 and 3.10.

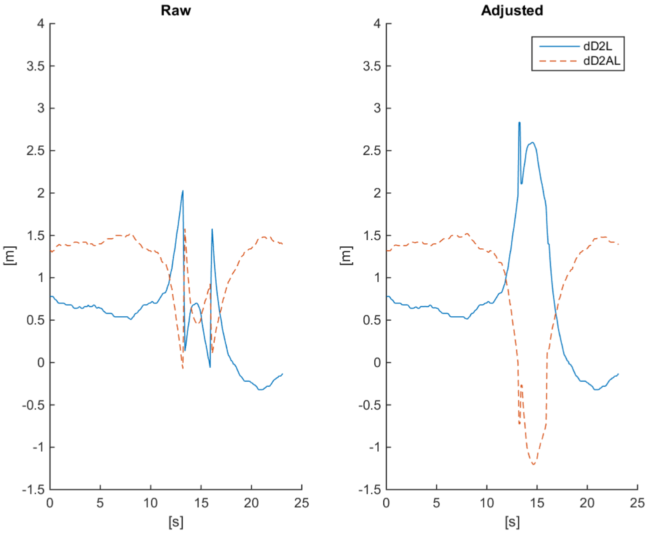


Figure 3.8: *A comparison between the raw data and the adjusted data. An example which only used mutual peaks of the time derivative to adjust signal values. The peaks are visualized in figure 3.9.*

Lastly, the signals were saturated to 6/-6 meters. This was due to the spline interpolation sometimes causing large peaks exceeding six meters, as can be seen in figure 3.7. However, instances where this happened were addressed later, as can be read about in section 5.4. It should be noted that in order for the correction of the

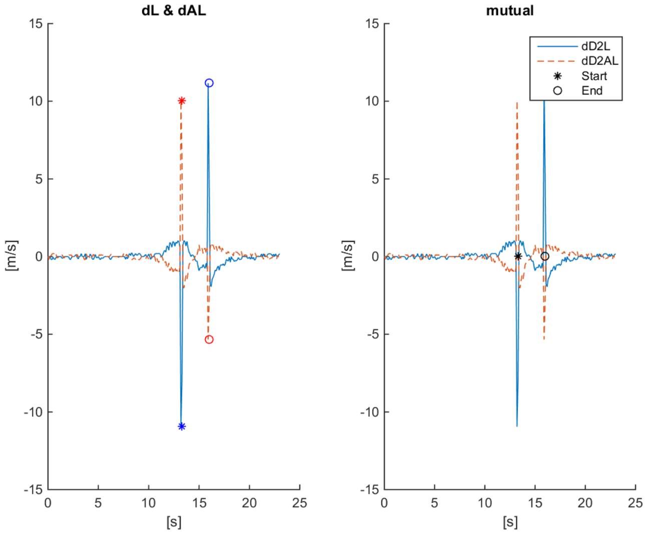


Figure 3.9: *An example of detection of peaks which occurred at the same time-stamp. An example which only used mutual peaks of the time derivative to adjust signal values. The peaks were used to locate an offset and improve data as can be seen in figure 3.8.*

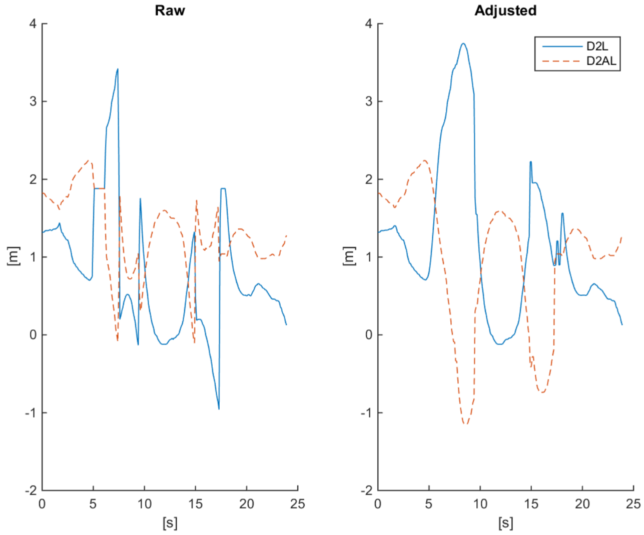


Figure 3.10: *A comparison between the raw data and the adjusted data. An example which only used non-mutual peaks of the time derivative to adjust signal values. The peaks are visualized in figure 3.11.*

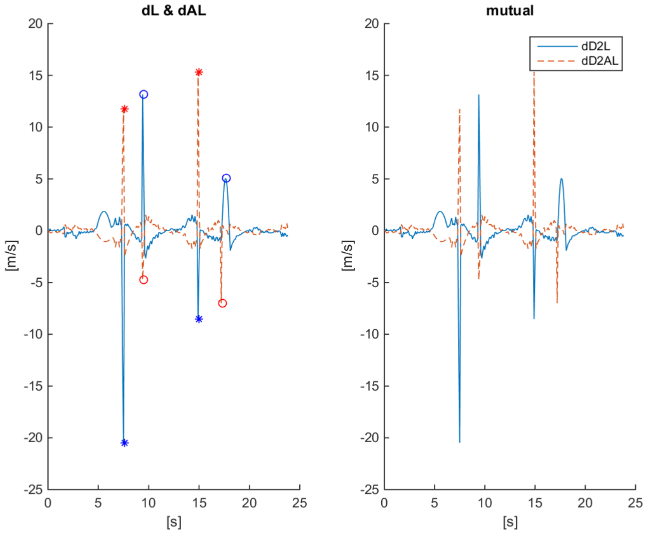


Figure 3.11: *An example of detection of peaks which did not occur at the same time-stamp. An example which only used non-mutual peaks of the time derivative to adjust signal values. The peaks were used to locate an offset and improve data as can be seen in figure 3.10*

resets to function at all both the start point and the end point of the reset needs to be included in the segment.

3.8 Velocity of the VRU

To obtain a measure such as TTC with the VRU, the velocity of both the ego vehicle and the VRU were necessary. A reasonable approach would be to simply calculate an approximate value for the velocity of the VRU via

$$V_{VRU} = RV + V \tag{3.1}$$

where RV is the raw signal from the ME which specifies the relative velocity to the VRU and V is the speed of the ego vehicle. The mean value of the velocity of the VRU, $\overline{V_{VRU}}$, could then be chosen to be the approximate speed of the VRU. The relative velocity is assumed to be negative due to the nature of an overtaking manoeuvre, that is the ego vehicle is travelling at a higher speed than the VRU.

However, RV was deemed to not be consistently accurate; sometimes netting in a $\overline{V_{VRU}}$ exceeding 40 km/h for bicyclists. 40 km/h is an achievable speed on a bicycle when travelling downhill or putting effort into travelling fast, however these speeds were sometimes concluded from segments where video displayed bicyclists that appeared to not be fulfilling either of previously mentioned criteria and thus was RV deemed inaccurate as a signal. The threshold value for high V_{VRU}, 40 km/h, was chosen based on discussion with seasoned bicyclists as well as reading on bicycle-related discussion forums on the internet [37]. In an attempt to work around this inaccuracy V_{VRU} was instead acquired via an estimated linear polynomials of the longitudinal distance to the VRU. The first degree term in this polynomial was considered to be the relative velocity, $RV2$, with unit meters per second. The velocity of the VRU was then derived in kilometers per hour via

$$V_{VRU} = RV2 * 3.6 + \overline{V} \tag{3.2}$$

where $\overline{V}$ was the average velocity of the vehicle over the time frame where the longitudinal distance data was available. The 'longitudinal distance to the VRU'-signal (LongD) was assumed to decrease at an equal or slower pace compared to the ego vehicle velocity when approaching bicyclists travelling in the same direction, but there were instances where LongD had a sharp positive time-derivative. A very clear example of this when the ego vehicle was overtaking several VRUs. As described in section 3.5 only the closest VRU in front of the vehicle were present in the data, thus each time a VRU was going out of detection range and there was another VRU further ahead LongD increased. To counteract this issue, the data was divided into sections. The boundaries of these sections were defined by NaN values in the signal as well as the positive peaks of the time derivative of LongD. Additionally, the first data point and the last data point in LongD were also considered boundary points. Polynomials which corresponded to a relative velocity (neglecting the sign of the relative velocity) exceeding the ego vehicle speed and positive relative velocities were removed. The mean of the first degree term of the remaining polynomials was then considered to be the relative velocity between the vehicle and the VRU.

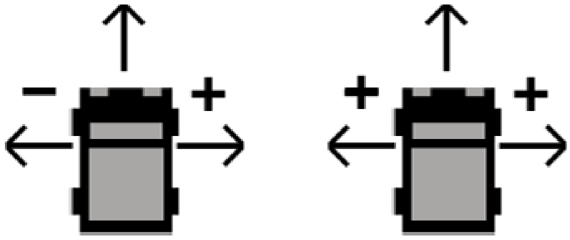
3.9 Holding and Extrapolating the VRU's Lane-Position

Due to the sensors only detecting objects in-front of the ego vehicle it was necessary to extrapolate the data outside of the detection range. The extrapolation of the longitudinal distance to the VRU was performed via subtracting a measure corresponding to the relative velocity between the ego vehicle and the VRU, multiplying it with the sampling time and adding (cumulatively) it to the last known longitudinal distance for each time instance where the longitudinal distance to the VRU was unknown

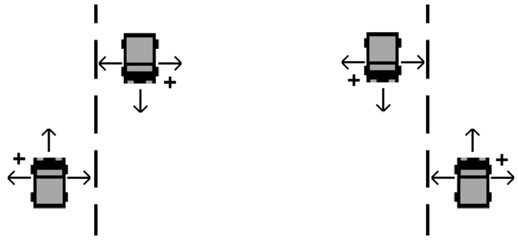
$$LongD(i) = LongD(i-1) - ts * \frac{(V(i) - V_{VRU} + RV2)}{2} \tag{3.3}$$

where LongD is the longitudinal distance to the VRU, ts is the sampling time, i is the time instance, V is the velocity of the ego vehicle, V_{VRU} the velocity of the VRU and $RV2$ the relative velocity between the ego vehicle and the VRU. As can be seen in equation 3.3 there is a mean of two different relative velocities subtracted, namely $V(i) - V_{VRU}(i)$ and $RV2$. This was due to using solely the $RV2$ would not take into account the potential change of V after the VRU had gone unknown (such as during accelerative overtaking manoeuvres).

Continuing with the lateral distance, the first thing that needed to be done was adjusting the lane-edge data based on country in similar fashion to what have been described already in section 3.5; The distance to the left lane edge was negative so its sign was reversed as shown in figure 3.12. Next, the distance to road shoulder was selected based on whether the data was collected in GB or not. If it was collected in the GB it was the distance to the left lane edge, else it was the distance to the right lane edge as can be seen in figure 3.12.



(a) *Measurements to the Lane Edges* (b) *Adjusted Data*



(c) *Left-Hand Traffic* (d) *Right-Hand Traffic*

Figure 3.12: *The distance to the left and right lane edges from the vehicle (grey rectangle). In figure 3.12b the distance to the left lane edge, which previously was negative, has had its sign reversed. If the data was collected in the GB the distance to the left lane edge was selected as the distance to road shoulder, else the distance to the right lane edge was selected as illustrated in figures 3.12c and 3.12d.*

The lateral distance between VRU and the lane edge was calculated by using the mean of the five last known instances of the difference between the vehicle's distance to the lane edge and the distance to the VRU. This value was then used as an offset on the vehicle distance to the lane edge where VRU data was unknown. In other words the lateral position of the VRU was held constant with regards to the lane-edge, but a dynamic value was achieved via

$$LatD = D2Lane - Offset \tag{3.4}$$

where $LatD$ is the lateral distance to the VRU and $D2Lane$ is the distance to the lane edge and Offset is the previously mentioned mean of last known instances of $LatD$. Some implications of this further discussed in section 5.1.1.

3.10 Comfort Zone Boundary - Time Headway and Time to Collision

Time headway (THW) was derived by using the longitudinal distance to the VRU and the speed of the vehicle via

$$\frac{LongD}{V} \tag{3.5}$$

and TTC was derived derived via

$$\frac{LongD}{RV} \tag{3.6}$$

where $LongD$ was the longitudinal distance to the VRU. It is worth noting that the way the TTC was derived did not take actual course of the ego vehicle into account, that is discarding the lateral distance and only using the longitudinal distance between the ego vehicle and the VRU. In other words there were values available in the TTC signal even when the ego vehicle was steering away from the VRU.

3.11 Comfort Zone Boundary - Lateral Clearance

Lateral Clearance (LatC) was derived via

$$LatC = LatD - (\frac{W_{VRU}}{2} + \frac{W_{Veh}}{2}) \tag{3.7}$$

where LatD is the lateral distance between the center of the vehicle and the center of the closest valid VRU object and W_{VRU} is the width of the VRU obtained from the ME-data and W_{Veh} is the width of the ego vehicle.

3.12 Comfort Zone Boundary - Minimum Distance

In order to derive minimum distance the VRU object is seen as a rectangle with a fixed length of the 1.70 meters (corresponding to an average length of a bicycle) and width W_{VRU}. While the vehicle is behind the VRU object the minimum distance is derived via three cases;

$$\sqrt{LongD^2 + (LatD - (W_{VRU} + W_{Veh})/2)^2} \tag{3.8}$$

$$\sqrt{(LatD - (W_{VRU} + W_{Veh})/2)^2} \tag{3.9}$$

$$\sqrt{(LongD + VehL)^2 + (LatD - (W_{VRU} + W_{Veh})/2)^2} \tag{3.10}$$

depending on whether the vehicle is in front of the bicyclist, passing the bicyclist or has past the bicyclist.

3.13 Annotating

As mentioned earlier, the ideal scenario of 100% hit-rate was deemed to not be reachable. Instead the data reduction only resulted in potential candidates, that is there were lots of false positives and presumably also false negatives (a more detailed breakdown can be found in section 4). Thus, filtering was only one of two tools; The second tool being annotations. Several attributes of the overtaking-manoeuvres were annotated, mainly due to limitations of the performance of the signal-processing and the data not being fully trustworthy. In other words the annotation acted as a verification of the filtered data. When annotations were performed the annotators had access to a graphs of signals such as the ego vehicle speed and the video-feed. If an overtaking occurred the attributes annotated were:

VRU type - What type of VRU was overtaken, that is bicyclist, pedestrian, motorbicyclist or other.

Overtaking Type - What type of overtaking-manoeuvre was performed, that is flying or accelerative.

Multiple VRU Overtaken - Was more than one VRU overtaken.

VRU Same Direction - Was the VRU travelling the same direction as the ego vehicle.

Visible Lane-Markings - What type of lane-markings were visible at the site of the overtaking, that is a central marking dividing the two adjacent lanes, markings separating the lane from the lane-shoulder, both or none.

Leading Vehicle Present - Was there a vehicle in front of the ego vehicle travelling in the same direction as the ego vehicle during the overtaking manoeuvre with a THW less than three seconds.

Oncoming Traffic Present - Was there oncoming traffic present in the adjacent lane during any of the phases of the overtaking-manoeuvre.

Synchronization Error - A synchronization error between signals and video-feed.

To distinguish between a flying and an accelerative manoeuvre, accelerative manoeuvres were defined as overtaking manoeuvres where the speed increased with ten km/h during the overtaking. This difference was found by looking at the CAN-data signal which contained the vehicle velocity and look at how the vehicle velocity changed throughout the duration of the overtaking manoeuvre. In addition to the previously listed attributes, a time-series was edited to indicate the start of each phase. Phase one started when the VRU was visible in the video-feed. Phase two began when the vehicles starts to alter its direction due to the VRU. As mentioned in section 3.3 phase three started when the longitudinal distance to the VRU first reached below three meters and phase four started when the longitudinal distance to the VRU first reached below minus three meters. Phase four ended when the distance driven since the start of phase four reached 50 meters. In other words, annotating the first two phases relied solely on the video-feed whereas the last two phases relied on a derived data-signal. Thus, only phase one and phase two were annotated whereas phase three and phase four were derived automatically. A complete list of what was annotated can be found in table A.2. The synchronization error was a known artifact of the preprocessing in the UDRIVE database which affected the synchronization between the video feed and the signal-data and had to be addressed. The synchronization error was individual for each record but was also observed to change trough-out the duration of the record. However, in this study the synchronization error was assumed to be static as it would be very hard to track how and when the error changed for each record. The synchronization error affected the definition of phase one and phase two, as they were derived using the video-feed. Additionally, it affected the subjective interpretation of ME-signals (e.g. lateral position of VRU) as it became harder to relate ME-data to the video-feed. The synchronization error was found by looking at the signal for ego vehicle velocity and the video feed at instances where the ego vehicle stopped and then compare the time-instances for the video and the signal. Finally, it is worth mentioning that the annotations were performed by four different people (including the author) and that they mainly focused on annotating segments from one country. The goal was to annotate 20 cases of flying overtaking manoeuvre and 20 cases of accelerative overtaking manoeuvres for each country. The number 20 was found as a compromise between a big samplesize and the timesink that annotation is.

4 Results

4.1 Quality Check of Data

From a dataset of 300’000 car records available for analysis in the database roughly 5000 segments were ”nominated” as potential overtaking manoeuvres, that is 5000 times a VRU disappeared from the ME-data under the ”right” conditions (e.g. on rural roads etc.). It is worth noting that one car record could be responsible for more than one of the 5000 segments, as per definition a segment was only a section of a record and a record could be decomposed into multiple segments. Continuing, of these 5000 ”nominated” segments approximately 500 were annotated. Of these 500 segments 211 were confirmed as overtaking manoeuvres and of these 211 entries 101 had no missing or ”bad” data for signals such as LatD, LongD, V, D2Lane and D2ALane and attributes such as I1, I2, I3 and I4. More specifically

1. 180 had all phases defined with time-stamps.
2. 171 had all phase durations being greater than zero seconds.
3. 34 had missing signal values in LatD, LongD or V.
4. Four had negative V_{VRU}.
5. Eight had too high V_{VRU} (exceeding 40 km/h).
6. 7 had at least 25% of the data being the default value (1.88) for D2Lane or D2ALane.
7. 17 were manually removed due to the processing of D2Lane and d2Adjacent signals failing on those segments.

After the brief quality check of the data was done there were 101 data entries (segments) left. 22 involved several VRUs and was finally removed resulting in the final number of 79 data entries. Before the quality check was performed only one country (Germany) did not fulfill the goal of 20 cases for both accelerative- and flying overtaking manoeuvres. The number of segments that were nominated to be potential overtaking manoeuvres were also much lower for Germany compared to the rest of the countries, that is a number in the hundreds versus numbers in the thousands. After the quality check no country fulfilled the criterion of 20 cases for each type of overtaking. This can be seen in table 4.1 which shows the distribution of data entries between countries. In figure 4.1 the velocity of the ego vehicle is shown, sorted by country and type of overtaking. The anchor is the time-instance where the longitudinal distance between the ego vehicle and VRU is zero. The main thing to take away from figure 4.1 is that the speed should be near constant for flying overtaking manoeuvres, but this is not the case for multiple data entries.

Table 4.1: Distribution of data entries between countries. The columns ’% Acc’ and ’% Fly’ refers to the percentage given by dividing the number of accelerative and number of flying overtaking manoeuvre entries with the total number of entries for the specified country respectively, whereas the column ’% Total’ refers to the percentage given by dividing the total number of data entries for arbitrary country with the total amount of data entries, that is 79.

Country	**No. Acc**	**No. Fly**	**No. Total**	**% Acc**	**% Fly**	**% Total**
France	9	18	27	33%	67%	34%
Germany	4	9	13	31%	69%	16%
Great Britain	8	14	22	36%	64%	28%
Poland	11	6	17	65%	35%	22%
	32	47	79	40%	60%	

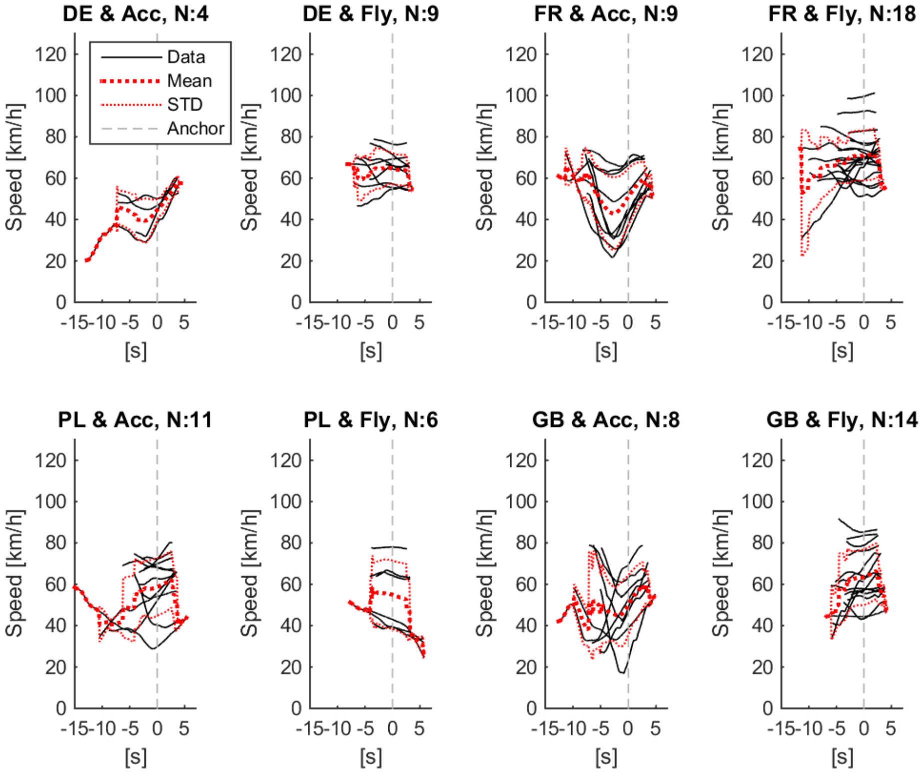


Figure 4.1: *Plots of ego vehicle speed, sorted by country.*

4.2 A Brief Comparative Analysis

Table 4.2 and 4.4 contain minimum distance (CZB) data gathered in this study, whereas table 4.3 and 4.5 contain minimum distance (CZB) data gathered in a previously conducted simulator study by Moretto [21]. Table 4.6-4.9 contain the p-values acquired via two-sample t-test on the data gathered in this study. The null hypothesis of the t-test is that the mean of the respective measures have the same mean and equal but unknown variances for flying and accelerative overtaking manoeuvres, for each respective phase. The p-values in the tables presented in this chapter and in the appendices is at which significance level the rejection of the null hypothesis still holds. The significance level (alpha) was set to 5%. It was set to 5% based on what previous studies had used, and when needed it was adjusted to handle multiple comparisons using Bonferroni correction. Significant p-values and the minimum values are bold in the tables.

Table 4.2: Minimum Distance (CZB) [m]. Data divided by flying and accelerative overtaking manoeuvre strategy.

		Phase 1	Phase 2	Phase 3	Phase 4
Accelerative	Mean	**57.21**	**15.07**	**2.77**	**46.83**
	STD	±54.05	±11.61	±2.05	±28.78
Flying	Mean	123.50	56.61	3.24	64.48
	STD	±112.81	±68.86	±2.51	±37.68
T-Test	Scalar	**p<0.001**	**p<0.001**	0.2	**p<0.001**

Table 4.3: Minimum Distance (CZB) [m], Moretto [21]. Data sorted by overtaking manoeuvre strategy.

		Phase 1	Phase 2	Phase 3	Phase 4
Accelerative	Mean	**60.32**	2.77	1.71	2.99
	STD	±72.45	±0.36	±0.59	±0.39
Flying	Mean	92.22	**2.39**	**0.93**	**2.50**
	STD	±57.88	±0.15	±0.27	±0.11
T-Test	Scalar	0.16	**0.002**	**p<0.001**	**p<0.001**

A notable thing about the data is the huge standard deviation when compared to the data in the Moretto [21] study, which partly might be explained by it being naturalistic data and not simulator data, that is presence of disturbance and noise in the signals. The values of the minimum distance (CZB) were heavily influenced by when the phases start. Both phase one and phase two were annotated. Whereas the minimum distance (CZB) is lower for flying overtaking manoeuvres in the simulator data the opposite is apparent in the naturalistic data.

Table 4.4: Minimum Distance (CZB) [m]. Data sorted by overtaking manoeuvre strategy and presence of oncoming vehicle.

			Phase 1	Phase 2	Phase 3	Phase 4
Accelerative	Oncoming Traffic Present	Mean	**39.91**	16.36	**2.61**	**53.05**
		STD	±33.65	±15.68	±2.02	±35.21
	Oncoming Traffic Not Present	Mean	64.65	29.79	3.07	54.24
		STD	±56.71	±42.91	±1.77	±34.01
Flying	Oncoming Traffic Present	Mean	41.96	**15.44**	5.09	59.79
		STD	±36.09	±9.74	±6.20	±39.84
	Oncoming Traffic Not Present	Mean	117.81	44.41	3.52	67.90
		STD	±104.43	±57.73	±2.50	±39.60

Table 4.5: Minimum Distance (CZB) [m], Moretto [21]. Data sorted based on presence of oncoming vehicle.

		Phase 1	Phase 2	Phase 3	Phase 4
Oncoming Traffic Present	Mean	81.40	**2.55**	**1.34**	**2.66**
	STD	±16.72	±0.07	±0.07	±0.13
Oncoming Traffic Not Present	Mean	**74.56**	2.62	1.73	2.73
	STD	±17.72	±0.34	±0.47	±0.30
T-Test	Scalar	0.35	0.22	**p<0.001**	0.62

The data between the two studies agrees when examined based on the presence of oncoming traffic. Both simulator data and naturalistic data suggests that the minimum distance (CZB) is lower when there is oncoming traffic present, which is in line with what studies such as Dozza et al. [9] and Llorca et al. [15] previously have concluded as well. This stands true both for flying and accelerative overtaking manoeuvres, as can be seen in table 4.4.

Table 4.6: p-values for minimum distance (CZB) of Phase 1 when data was sorted by overtaking manoeuvre strategy and presence of oncoming vehicle.

		Accelerative		Flying	
		Onc. Present	Onc. Not Present	Onc. Present	Onc. Not Present
Accelerative	Onc. Present	-	**p<0.001**	0.40	**p<0.001**
	Onc. Not Present	-	-	**p<0.001**	**p<0.001**
Flying	Onc. Present	-	-	-	**p<0.001**
	Onc. Not Present	-	-	-	-

Table 4.7: p-values for minimum distance (CZB) of Phase 2 when data was sorted by overtaking manoeuvre strategy and presence of oncoming vehicle.

		Accelerative		Flying	
		Onc. Present	Onc. Not Present	Onc. Present	Onc. Not Present
Accelerative	Onc. Present	-	**p<0.001**	0.57	**p<0.001**
	Onc. Not Present	-	-	**p<0.001**	**p<0.001**
Flying	Onc. Present	-	-	-	**p<0.001**
	Onc. Not Present	-	-	-	-

Table 4.8: p-values for minimum distance (CZB) of Phase 3 when data was sorted by overtaking manoeuvre strategy and presence of oncoming vehicle.

		Accelerative		Flying	
		Onc. Present	Onc. Not Present	Onc. Present	Onc. Not Present
Accelerative	Onc. Present	-	0.10	**$p<0.001$**	**0.002**
	Onc. Not Present	-	-	**0.007**	0.16
Flying	Onc. Present	-	-	-	0.02
	Onc. Not Present	-	-	-	-

Table 4.9: p-values for minimum distance (CZB) of Phase 4 when data was sorted by overtaking manoeuvre strategy and presence of oncoming vehicle.

		Accelerative		Flying	
		Onc. Present	Onc. Not Present	Onc. Present	Onc. Not Present
Accelerative	Onc. Present	-	0.58	0.01	**$p<0.001$**
	Onc. Not Present	-	-	0.05	**$p<0.001$**
Flying	Onc. Present	-	-	-	**0.003**
	Onc. Not Present	-	-	-	-

4.3 Various Data Visualizations

In this section various data is presented. How the data is presented varies; some figures present the data with eighth initial categories, defined by whether oncoming vehicles (Onc) and leading vehicles (Lead) were present during the overtaking manoeuvre ('T' and 'F' represents 'Present' and 'Not Present' respectively) as well separation depending on if it was a flying or accelerative overtaking manoeuvre. Other plots only present the data based on whether it was flying or accelerative. 'N' is the number of data entries. Unless stated otherwise, the data in the plots are anchored to where LongD first is equal to or less than zero, that is roughly in the middle of phase three. Some boxplots contain eight different phases on the x-axis, labeled '0-1', '1', '1-2', '2', '2-3', '3', '3-4' and '4'. The labels consisting of two digits separated by a dash indicates the transition between two phases. For these transition-values the average of three adjacent data points was used, that is values at a given time-stamp ±1, whereas the mean value of the phase was used for the single digit phase labels. For a more complete visualization of the results, please see to appendices A-E.

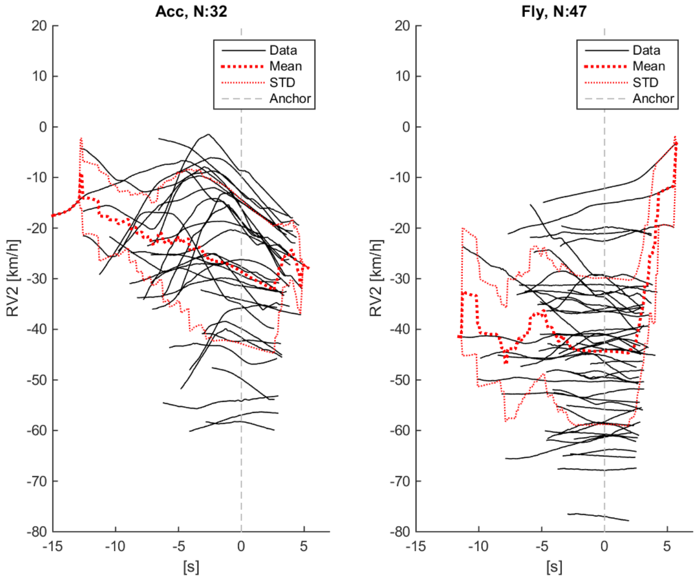


Figure 4.2: *Plots of relative velocity.*

Figure 4.2 displays the relative velocity which was derived using the vehicle velocity from the CAN-data and the V_{VRU} which in turn was derived from ME-data.

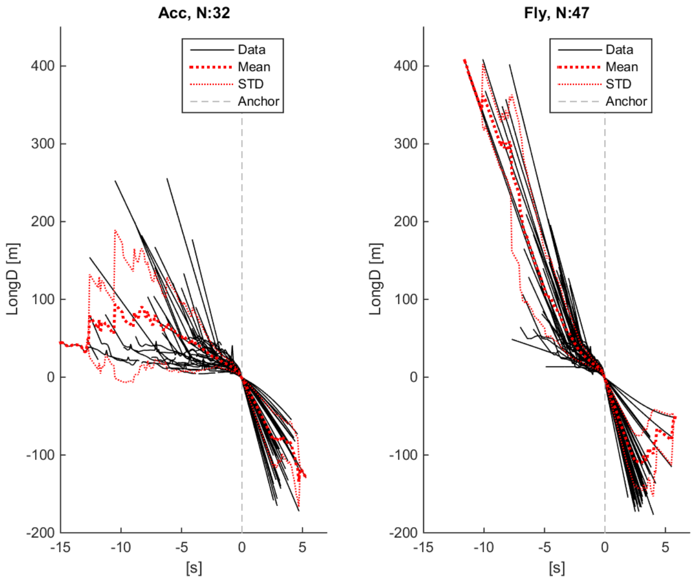


Figure 4.3: *Plots of longitudinal distance to VRU.*

In figure 4.3 and 4.4 the longitudinal distance to the VRU is presented in two different ways, namely how the data is anchored.

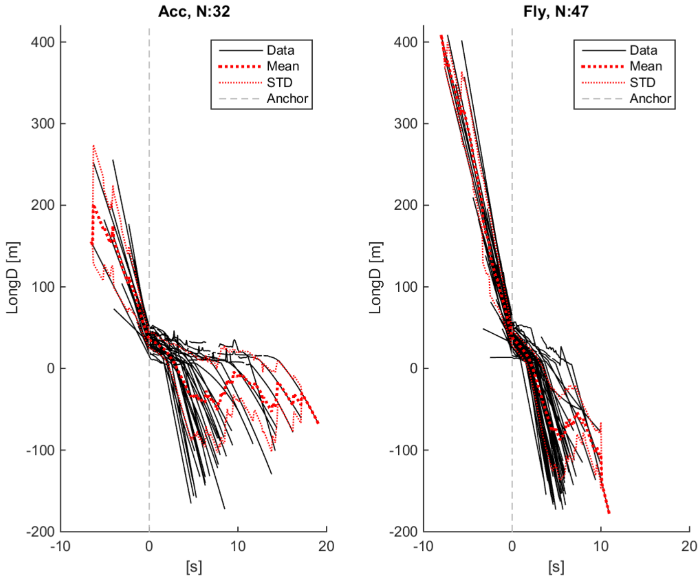


Figure 4.4: *Plots of longitudinal distance to VRU anchored to where ME first detected the VRU.*

In figure 4.5 the phase duration is presented. As can be seen, phase three is very short compared to the other phases. This further supports the decision to increase the definition of phase three, compared to how itwas defined in Dozza et al. [9]. The ego vehicle velocity is shown in figure 4.6.

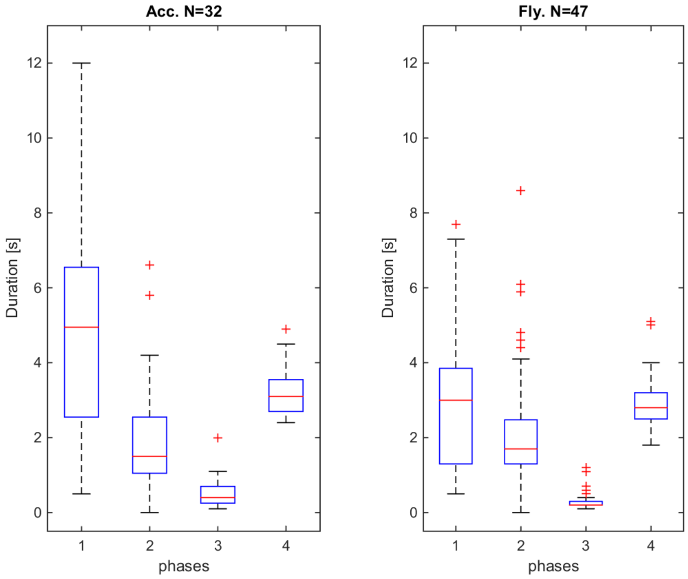


Figure 4.5: *Boxplot of phase duration for each phase.*

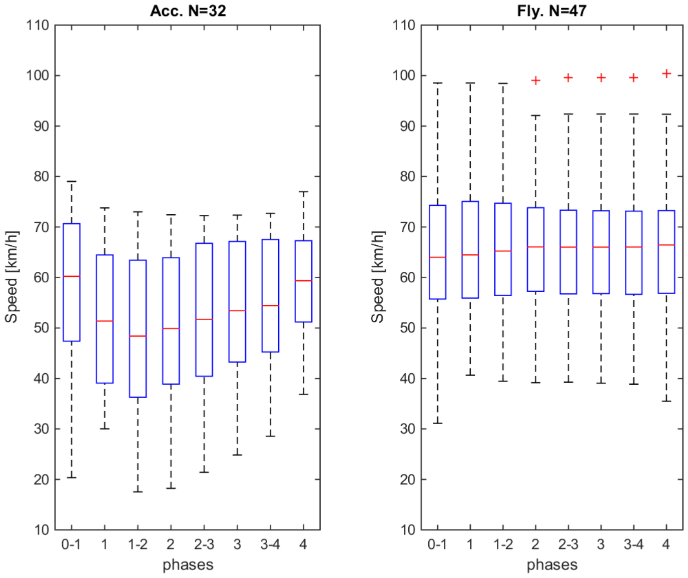


Figure 4.6: *Boxplots of mean velocity for each phase. The entries on the x-axis which consist of two digits separated with a dash indicates the transition between the two phases.*

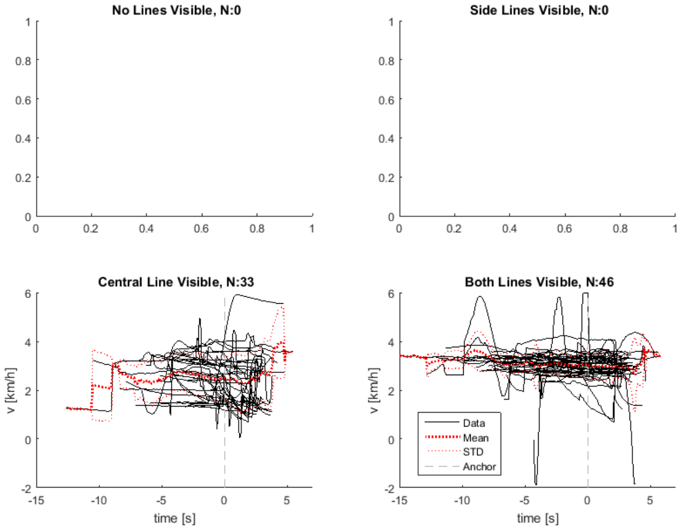
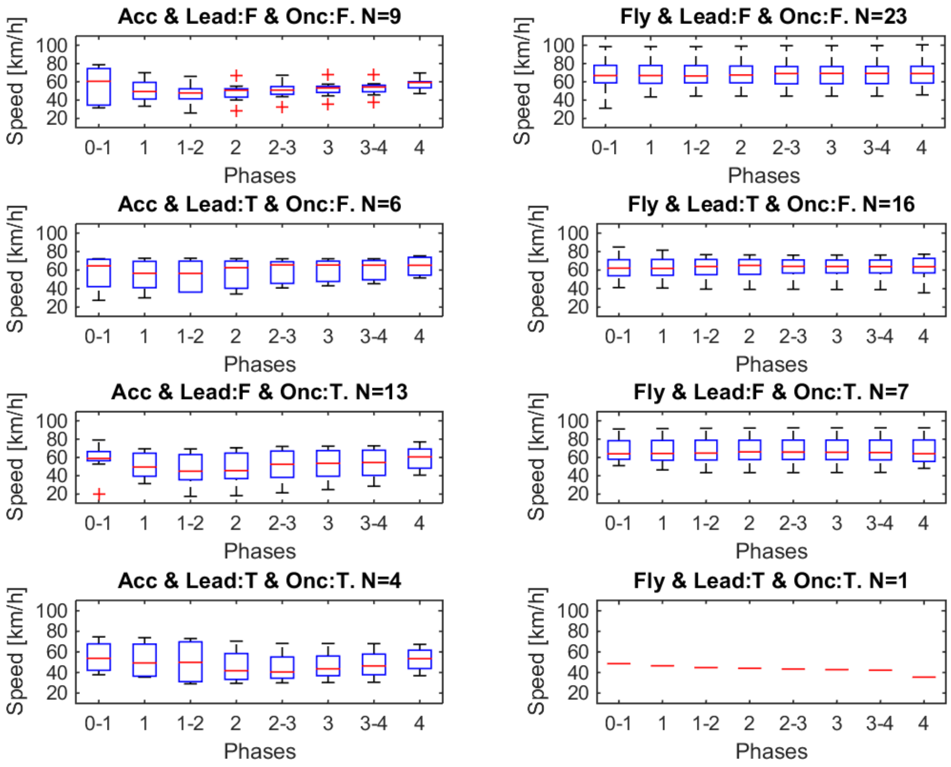


Figure 4.7: *Plots of lanewidth. As can be seen only entries which had the central line and/or both lines present made it trough the brief qualitative check of the data.*

In figures 4.7 it can be seen that no data entries which had only side line present or no lane-marking present at all made the final cut in the quality check. This vaguely implies that the presence of a central line marking was essential for the implemented methodology to function properly. The lane width (LW) also appears to be less deviating when both side and central lane markings are present, although the data contains quite a few peaks.

As can be seen in figure 4.8 the speed of the ego vehicle (V) is higher during flying overtaking manoeuvres than accelerating whereas figure 4.9 shows that V_{VRU} was around 20 km/h for all instances.


Acc & Lead:F & Onc:F. N=9

Fly & Lead:F & Onc:F. N=23

Acc & Lead:T & Onc:F. N=6

Fly & Lead:T & Onc:F. N=16

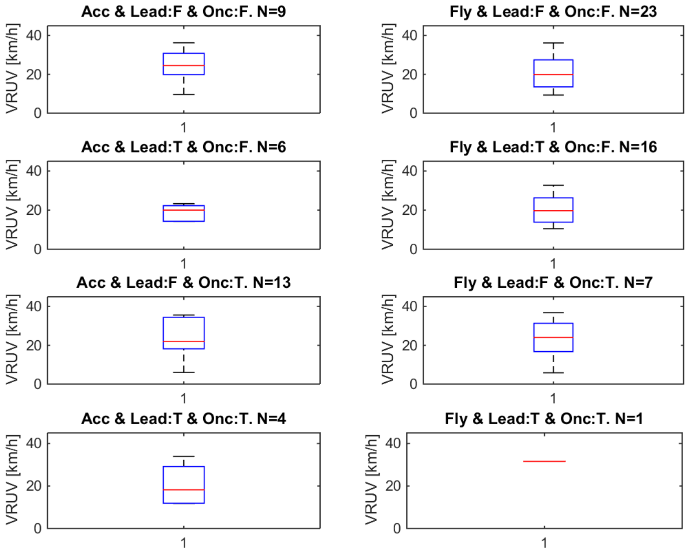
Acc & Lead:F & Onc:T. N=13

Fly & Lead:F & Onc:T. N=7

Acc & Lead:T & Onc:T. N=4

Fly & Lead:T & Onc:T. N=1

Figure 4.8: *Boxplot of average ego vehicle speed for each phase. The data is divided both by presence of oncoming vehicle and presence of leading vehicle.*


Acc & Lead:F & Onc:F. N=9

Fly & Lead:F & Onc:F. N=23

Acc & Lead:T & Onc:F. N=6

Fly & Lead:T & Onc:F. N=16

Acc & Lead:F & Onc:T. N=13

Fly & Lead:F & Onc:T. N=7

Acc & Lead:T & Onc:T. N=4

Fly & Lead:T & Onc:T. N=1

Figure 4.9: *Boxplot of VRU speed.*

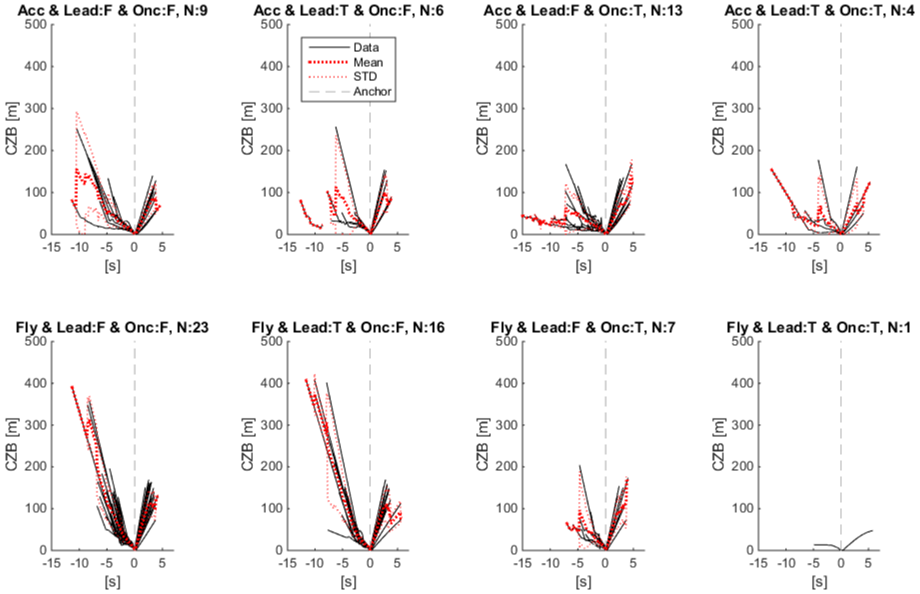


Figure 4.10: *Plots of minimum distance (CZB) . The data is divided both by presence of oncoming vehicle and presence of leading vehicle.*

In figure 4.10 continuous data is shown for minimum distance (CZB), but a closer look is required to understand what is going on during phase three due to the huge range of this CZB measure. This is provided in figure 4.11. The function of these two figures is to show the overall behaviour of the minimum distance (CZB) (more specifically during phase three), that is not to provide insight about details of the minimum distance (CZB). To be enlightened about details regarding the minimum distance (CZB), please see to the appendices.

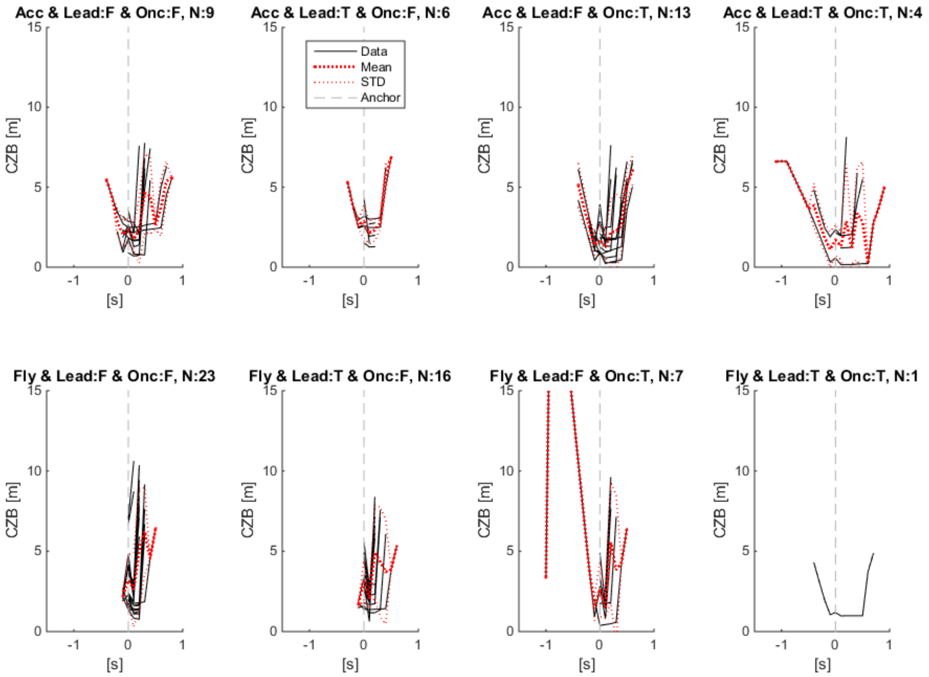


Figure 4.11: *Plots of minimum distance (CZB) during phase three. The data is divided both by presence of oncoming vehicle and presence of leading vehicle.*

For a more complete visualization of the results, please see to appendices A-E.

4.4 Miscellaneous Results

In this section plots that are more related to the methodology are presented. Figure 4.12 provides examples of how data entries were excluded due to containing the default value of 1.88 in D2Lane or D2ALane for more than 25% of the segment. Figure 4.13 illustrates how the correction of the D2ALane failed for a data entry. Figure 4.14 displays how using polynomial fit on several sections affected a data entry which only contained one VRU, in contrary to what was assumed in section 3.8. This particular case resulted in a negative V_{VRU} as can be seen by doing a calculation using the values provided in the legend, that is mean of VehV + RV2 which roughly equals $34 + (-46) = -12$ km/h. This would mean that VRU was traveling in the opposite direction of the ego vehicle, which was not the case. It should also be noted that 'Polyfit over all data' refers to the polynomial fit using all the data in LongD whereas 'Polyfit over "sections" of data' divided the data into smaller sections, as described in the last paragraph of section 3.8. 'Polyfit when -RV2<VehV & RV2<0' implies that the initial 'polyfit over "sections" of data' was filtered with those conditions. 'Used polyfit (mean)' refers to the average of those filtered polynomials.

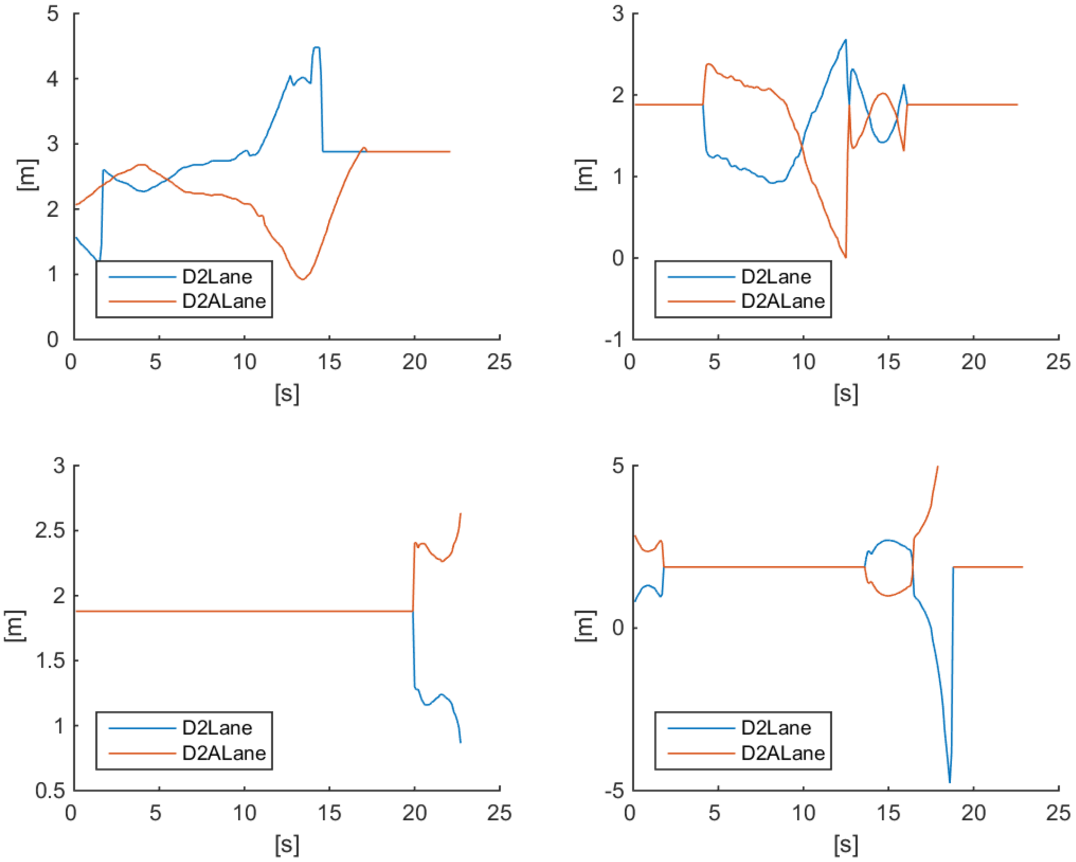


Figure 4.12: *Data entries where the distance to the lane edge was not missing, but contained large sections of default values instead.*

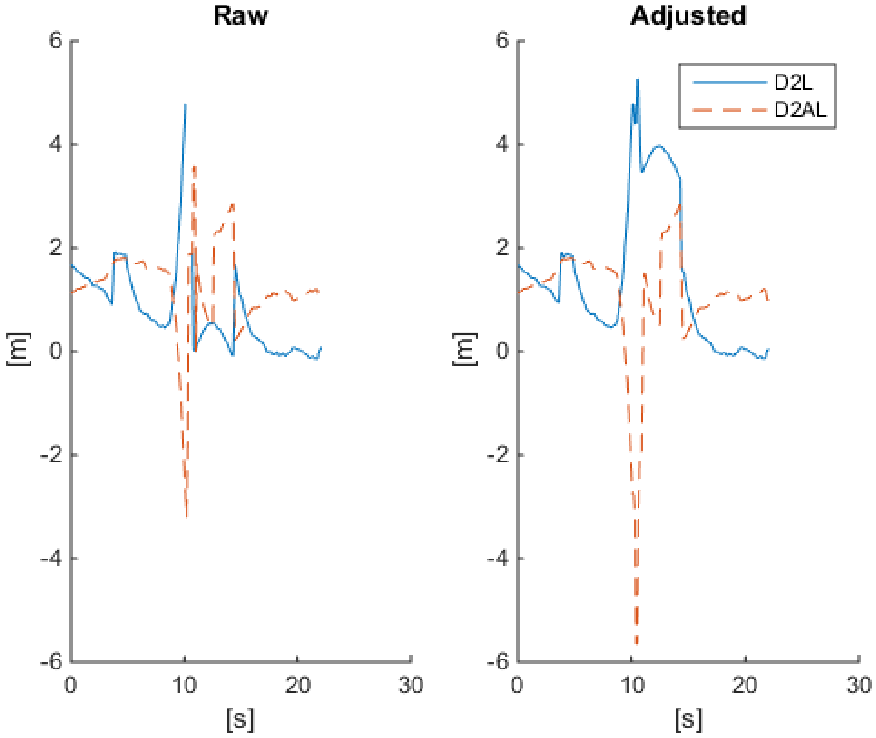


Figure 4.13: *Example of data entry where the correction of D2Alane failed.*

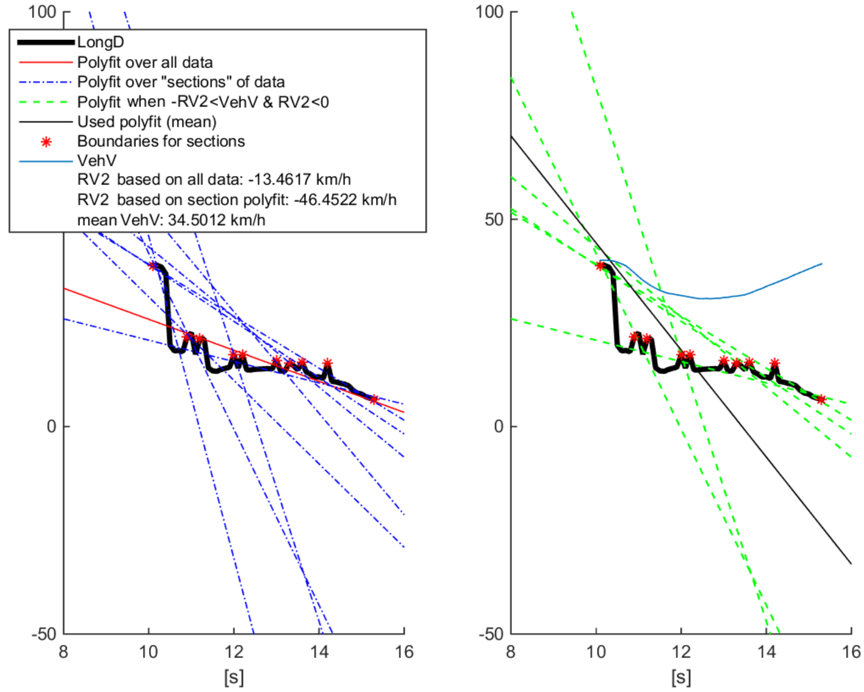


Figure 4.14: *Example of data entry where the calculation of the RV2 was miscalculated resulting in a negative* V_{VRU}*. The unit of the y-axis is [km/h] for VehV but [m] for every other line.*

5 Discussion

5.1 Data-gathering

Where the previous studies utilizing naturalistic data, for instance the Chuang et al. [8] paper, had a very limited amount of data to work with due to the excessive process it is to collect it, this work had access to a huge database containing naturalistic driving data from several European countries such as Great Britain, Poland, Germany and France. So for this study the time sink was instead to extract relevant events from the database. Furthermore, the data was collected from the vehicle and not the VRU, so it made for another perspective of the overtaking-manoeuvre.

5.1.1 Data Reduction

Initially, this study aimed at looking at overtaking-manoeuvres of either one pedestrian or one bicyclist, however halfway through the allotted time-frame it was found that overtaking-manoeuvres pedestrians were a rare occurrence. The focus was then changed to only look at overtaking manoeuvres involving bicyclists. Segments with any amount of bicyclists were included. However, after realizing that this would dilute the data for example via a longer phase three for overtaking manoeuvres involving multiple VRUs, the study was then shifted to only focus on overtaking manoeuvres involving a single bicyclist. This had a negative impact on the data-reduction.

To begin with, segments which only contained VRU data from pedestrians were considered valid overtaking manoeuvres. It goes without saying that if the interest lies in overtaking manoeuvres of bicyclists then overtaking manoeuvres involving pedestrians is not of interest. However, it was observed that bicyclists could if only for the slightest amount of time be misclassified as pedestrians. Including pedestrian data prevented this potential data loss. It was also observed that road-signs had a tendency to be misclassified as pedestrians, resulting in complete false positive entries, that is where no VRU was present at all. This type of false positive could most likely have been avoided if there was a requirement that the segment must contain at least some VRU data originating from bicyclists since road-signs were not miss-classified as bicyclists to the same extent. The code was kept as it was mainly due to the time-constraints. Adjustments had most likely been found to worsen the performance in the short term, due to the changes cascading trough algorithms and making thresholds found by trial-and-error obsolete etc. It was concluded that it was not necessary to change something that functioned somewhat properly.

5.1.2 Annotations

Only overtaking-manoeuvres involving a single bicyclist travelling in the same direction were regarded in this work. That meant that overtaking-manoeuvres of multiple bicyclists or any amount of pedestrians, horse-riders, motorbike riders, etc were discarded for this work. This was due to the annotation being a time-sink and the fact that there were very few (less than ten instances) overtaking-manoeuvres of pedestrians in the 500 annotated segments). Annotation of segments was a necessary step, in this case for several reasons. The most prominent cause was to verify that the filtering functioned somewhat properly (i.e. there were false positives and they should be removed). The second reason was that phase one was considered to start when the VRU first appeared in the video feed and not when it first was detected by the ME-system. The third reason was that the algorithms developed did not supply a good estimation of where the vehicles starts to swerve around the VRU during the overtaking, that is start of phase two. It should be possible to implement an algorithm to detect the latter and it can be assumed that appropriate signals to look at could include but not be limited to the lateral acceleration, steering-wheel angle and yaw-rate.

However, in the Parkin and Meyers [24] paper it was concluded that a marked bicycle lane might lead to reduced lateral clearance. This behaviour, where the motorists do not increase the lateral clearance and thus not alter the yaw of the vehicle, might result in it being harder to distinguish the phases mentioned in section 2.1. More specifically distinguishing the beginning of the pull out phase.

Additionally, the yaw-rate had a tendency to be unusable in the sense that signal consisted of an unreasonably high and constant value as can be seen in table A.3. The steering-wheel data could also be tricky to use, for instance if the overtaking occurs on a non-straight road there is an offset to the data due to the steering-wheel being turned well before the overtaking occurs. Furthermore, those offsets could be hard to mitigate due to the limited time-frame for each segment (approximately 20 seconds worth of data). Even if that issue would have been addressed properly the steering wheel data also suffered from being distorted or missing for several entries. Another signal worth mentioning is the road-curvature-radius, R. It could potentially have been used to mitigate the offset in some way. However, due to the time constraint of the project it was not examined further than that it existed and included in the saved data. In other words, a decision was made that it was not worth the hassle to try eliminate human interference in the data-gathering with the consequences being that

- Some of the data was subjectively categorized by the annotator (e.g. is the road straight? has phase two began yet?)
- The sample size for the analysis was reduced due to the slower processing-procedure where annotations were a necessity
- The algorithms were developed faster due to not being required work flawlessly since annotations functioned as a fail-safe

Furthermore, the attributes in table A.2 are not the only attributes that could have been annotated. For instance, the weather conditions are completely neglected and it is easy to imagine that fog would affect the start of phase one. Other measures that could have been annotated include some further attributes such as gender of the bicyclist, whether the clothing was of highly visible materials etc. However, due to the low resolution of the videos, motion blur and censorship blur, details required to accurately determine the gender etc. would be vague if not completely indistinguishable. Thus, it would probably result in the annotations being even more subjective. As mentioned in section 3.13 annotators mainly focused on annotating segments from one country each, this might have introduced some additional bias into data on a country versus country basis.

In figure 4.1 it can be seen that some entries of flying overtaking manoeuvres in France and Great Britain are mislabeled as flying while they in fact are accelerative, at least according to vehicle velocity. To see this look for lines which has a positive slope for a majority of its existence or upside-down bell-shaped curves. Both indicate accelerative manoeuvres; A continuous accelerating motion or a decelerative motion that is followed by an accelerative motion respectively. This implies that the annotations were done in an inconsistent manner. As mentioned in section 3.13, when annotations were done the annotators had access to a graph of the ego vehicle speed and the video-feed, however there was no clear indication of whether the condition for an accelerative overtaking was met. Even though the accelerative overtaking manoeuvre was defined by a difference in speed by ten km/h it still left room for subjective interpretation. For instance, if the speed difference was seven km/h and this seven km/h difference occurred in a short time-span it would probably more resemble a accelerative overtaking than a flying overtaking manoeuvre. The appearance of the overtaking in the video feed most likely also affected the annotators in some fashion, for example if an annotator had a pretty stern view on how an accelerative overtaking should look like (may be based on own driving experience) he or she probably had a tendency to discard the definition for accelerative overtaking manoeuvre based on vehicle velocity. A possible amendment in order to reduce subjectivity would be to somehow incorporate the time-frame of which the difference in speed took place or at the very least display to the annotator that the condition is met for the accelerative overtaking. Another small improvement would be to change certain things in the interface that was used to annotate, namely instead of having check-boxes representing true and false statements (where a checked box means true), instead implement multiple choice lists. For example, the check-box for the condition "This segment is an overtaking manoeuvre" which by default has the check-box empty, that is false, maintains its value unchanged even after being checked by the annotator. If the options for that statement were, "not determined", "true" and "false" it would be easier to keep track of if the segment had been annotated or not.

5.2 Inaccurate Speed of VRU

In figure 4.5 it can be seen that phase three is by far the shortest phase, with a median value (red line) of 0.4 seconds for accelerative and 0.2 seconds for flying overtakings. Phase three was defined as beginning when the distance between the rear of the bicycle and the front of the vehicle was less than three meters and ending when the distance between the front of the bicycle and the rear of the vehicle was greater than three meters. Discarding the distance covered by bicycle during phase three the vehicle traveled at least six meters in 0.2 during a flying overtaking. That translates to

$$6/0.2 = 30m/s \tag{5.1}$$

$$30 * 3.6 = 108km/h \tag{5.2}$$

a vehicle speed of 108 km/h for flying overtaking manoeuvres. However, looking at figure 4.6 the mean speed of the ego vehicle is nowhere near 108 km/h. Note, this is assuming the bicyclist is stationary, that is neglecting some the extra meters travelled. In figure 4.3 it should be expected that the longitudinal distance is decreasing linearly for flying overtaking manoeuvres since the velocity of the ego vehicle should be somewhat constant and assuming the velocity of the VRU is constant. However this is not the case. It is even clearer that something is wrong when looking at the same data but the data is anchored to when the VRU is detected by the ME-system, as shown in figure 4.4. Even without the anchor point marked in the figure it would be fairly easy to point out where the extrapolation ends and the actual ME data begins (and where the actual ME-data ends and the extrapolation begins again) due to the sudden change of the slope (as previously mentioned the slope corresponds to the relative velocity). As described in section 3.9 the extrapolation of the bicyclists position is done via using both using a static and a dynamic relative velocity as described in section 3.8. The dynamic relative velocity took the vehicle velocity into account for each time-stamp whereas the static relative velocity only depended on the rate of change of the longitudinal distance. The bicyclist speed was calculated via taking the average vehicle speed for the limited time-frame defined by when the ME system had the bicyclist in detection range. The dynamic relative velocity then utilized this bicyclist speed. It could be argued that the velocity of the bicyclist was uncertain and that this uncertainty stems from the usage of "sections" when executing polynomial fits on longitudinal distance to the VRU, an extreme case of this is depicted in figure 4.14. A high relative velocity leads to a low bicycle speed which leads to a greater rate of change of longitudinal distance in the extrapolation. This inaccurately extrapolated longitudinal distance was then used to automatically define the phase three and four. Due to the unintended behaviour shown in figure 4.14 being discovered very late in the project, there was no time left to rewrite the code so the polynomial fit always was done on the full segment, as this seemed to during a very brief (read three to five cases) qualitative check give more reasonable values for V_{VRU}. This mistake resulted in that the value of the V_{VRU} may or may not be a reasonable estimation. Furthermore, the most critical phase, phase three - the passing phase, is defined using signals which rely on the V_{VRU}.

If it was desired to still maintain overtaking manoeuvres involving multiple VRUs in the dataset, a solution would be to keep the upto four objects from the ME separate and not combine them into one object, that is the closest VRU as described in section 3.5. However, if this was not desired one potential amendment is to always use a 'polynomial fit' on the whole data sequence. Another approach, which involves excluding extrapolation of the VRUs longitudinal position is to annotate the time instance when the vehicle is parallel to the VRU and then use the vehicle velocity to calculate the distance offsets which defines phase three, under the assumption that the VRU is stationary. This removes the constraint of relying on extrapolated signals when defining phase three. However, when the ego vehicle is parallel with the VRU the VRU is not visible in the front middle video-feed/camera and other camera angles have to be used. This may or may not be an issue; When annotating the synchronization error (instance where the ego vehicle stops) some annotators mentioned that when they looked at the side view camera and that the side view camera was not synchronized with the front middle camera. However, this might be purely subjective. In other words, this potential synchronization issue between cameras angles would have to be further investigated before the previously proposed approach of annotating when the vehicle is parallel to the VRU even could be considered as viable.

5.3 Lateral Extrapolation

As mentioned in section 3.5 the lateral position of the VRU was held constant with regards to the lane edge. This approach neglects any change of lateral position the VRU may do, such as turning. However, extrapolating the lateral speed of the VRU might be tricky due to the wobbling motion that is created when bicycling [35]. A small change in lateral position might cause the extrapolation to place the bicyclist in the middle of the lane when in fact it only was a wobbling motion and the bicyclist pretty much remained at the same lateral position. However, the overtaking manoeuvre is executed quickly once the ME loses track of the bicyclists position. This means that the VRU has a very limited time-frame to adjust its lateral position, in other words it would require a very sudden (and relatively big) change of the VRUs lateral position. Extrapolating such changes would risk to greatly distort the actual lane position of the VRU. Returning to the previous scenario where a wobbling motion could misplace the VRU to the middle of the road, the extrapolation of a great and sudden motion could instead very well place the VRU on the other side of the road. Thus, extrapolating the lateral position was seen unnecessary. The "holding" of the lateral position was deemed to provide a better overall representation for both of the previously mentioned scenarios.

5.4 Qualitative Examination

As mentioned in section 3.7 adjusting the distance to the lane-edge when crossing a lane had some issues. The artifacts in the signals arose due to several reasons including specific road conditions such as lanes merging or the central lane marking disappearing etc. The sources for these artifacts which made the D2Lane and D2ALane signals behave unusually were not thoroughly examined. However, even though no thorough examination has been performed it was noted that the lane-tracking-signals failed to address sharp turns properly. Such changes could cause "false" peaks in the data which resulted in the data being altered inaccurately. Entries containing these malicious alteration of data were manually removed. There were also instances were the signal partially (or entirely) had the default value of 1.88, as mentioned briefly in section 3.7. As can be seen in figure 4.12 the curves are flat-lining at certain points and this is where the signal is taking the default value. Data entries which had the default value for at least 25% of the raw data in d2Lane and d2ALane were removed from the dataset. As previously mentioned the manual removal of data entries was performed to remove entries where the processing had failed. Note that the removal only took one factor into consideration, namely the shape of the signals D2Lane and D2ALane. To make the qualitative check better the other signals should also be looked at and all the signals should be compared to what the video feed shows. D2Lane and D2ALane were the signals which contained the information of the distance to the lane shoulder and the distance to the adjacent lane and they were used to derive the lateral position of the vehicle as well as the holding of the lane position of the VRU and thus also the minimum distance and LatC CZB measures, in other words they were used in a lot of derived measurements. Towards the very end of the time-frame allotted for this study it was also noted that the raw ME signal for lateral distance to VRU seemed to function a bit unreliable, namely that the lateral distance to VRU was greater than the lateral distance to the lane edge, effectively placing the VRU outside of the lane for instances despite the video-feed showing the VRU being located in the middle of the lane. These type of mismatches need to be investigated further as well, not only for the lateral distance but also for other raw data measures. In other words, there was a need for a more thorough and meticulous quality check of the data.

Another thing worth pointing out is the apparent difference of behaviour when it comes to lanewidth, that is the sum of D2Lane and D2ALane for each time-instance, regarding the presence of lane-markings. As can be seen in figure 4.7, it seems like when both type of lane-markings are present that the signals are more reliable, in the sense that they have a lower range of values and thus a lower standard deviation (although there are some outlier values present, i.e. the peaks of 6 meters and -2 meters). In table A.3 it can be seen that the lanewidth varies between the minimum value of -2.6 meters and the maximum value of 8.4 meters with an average of 2.8807 meters and a standard deviation of 0.7409 meters. Whether this behaviour originates from the raw data or the algorithm which adjusts lane-crossings, as described in section 3.7, is unexamined in this work. It would however, be useful to investigate how much the signal behaviour differ based on the presence of lane-markings. Notable, no data entries which did not have the central line marking did pass the qualitative check. This might be due to the ME data being less stable when there is no central line marking.

5.5 Comparative Analysis with Previous Studies

In figure 4.11, which depicts minimum distance between the VRU and the ego vehicle (CZB) for phase three, it could be expected that the curves should resemble the shape of a 'U', 'V' or an upside down bell curve. This stands true for some data entries, but far from every entry. For many of the flying overtaking cases there the start of phase three begins when the longitudinal distance is zero, this might have to do with the relative velocity that had a tendency to be (negatively) greater during flying overtaking manoeuvres, as shown in figure 4.2. This relates to the boundary for the start of phase three, that is first instance of where LongD is less or equal to three meters. This boundary might have been set insufficiently small for the very same reason two meters was deemed too small, as described in section 3.3. Although the relative velocity (and other measures which relied on it, e.g. LongD and minimum distance) were passed trough a qualitative check, additional quality issues arose very late into the project making it impossible to adress in this thesis. This also affects the rest of the data, in the sense that phase three and phase four are not certain to start where they are said to start since the definitions of phase three and four depend on LongD. Thus, comparing the LatC (CZB) during the passing phase (phase three), which previously have been used as one of the most prominent safety measures

[35] [9] [15], loses its meaning when the time-frame of phase three might be shifted/displaced. In other words, minimum distance (CZB) which partly is derived using the extrapolated LongD, will potentially not only be examined at the wrong time-frame but also with the wrong values. How much of an impact these issues causes is hard to say as it boils down to how inaccurate the derivation of the VRU speed was. Instead of trying to justify an analysis based on uncertain data, this analysis instead serves as an demonstration of how such a comparison can be performed.

When comparing data gathered in this study with data from the driving simulator study conducted by Moretto [21] it can be seen that the standard deviations for the minimum distance (CZB) is much higher in this study, as seen in table 4.2 and 4.3. This might be explained by how the phases were defined. If we look at the phase-duration depicted in 4.5 we can see that both phase one and phase two varies quite a bit in duration. Assuming the vehicle was not following the VRU for a majority of the duration it can be said the minimum distance between the VRU and ego vehicle (CZB) will vary quite a lot, as the ego vehicle has plenty of time to close in longitudinally on the VRU. This can cause the minimum distance (CZB) to have a wide variety of values for phase one and phase two. The data from the Moretto [21] study also shows that accelerative overtaking manoeuvres have a lower minimum distance (CZB), while the data of this study claims otherwise, that is that flying overtaking manoeuvres have a lower minimum distance (CZB).

When looking at the data divided by the road/traffic condition of oncoming traffic we can see that the data in this study suggests that oncoming traffic reduces the minimum distance (CZB), both for flying and accelerative overtakings. This is in line with what can be seen in the data from the work of Moretto [21] as well.

5.6 Choice of Overtaking Strategy

Although the data can be argued to be unreliable, if only looking at the number of each type of overtaking while keeping in mind that the annotated labels such as 'Fly', 'Acc', 'Lead' and 'Onc' might be slightly inconsistently assigned, there are a few things that can be observed. Looking at figure 4.8, it can be seen that the flying overtaking strategy is lesser used when there is oncoming traffic present than when there is not (eight and 39 instances respectively). For the accelerative overtaking manoeuvres this is more evenly distributed (17 and 15 instances respectively). Although there might be some undocumented bias in the data-reduction algorithm affecting the number of segments extracted of each instance and some subjective classification introduced by the annotators, this still supports the idea that oncoming traffic affects the decision making regarding which overtaking strategy to execute.

6 Conclusions & Future Work

This chapter gives a concluding summary of the author's thoughts on the obtained results and how the initial problems were solved. As well as some suggestions for future work. The thoughts and suggestions vary from general to very specific.

6.1 Conclusions

This thesis demonstrated how to identify overtaking manoeuvres of VRUs on rural roads from naturalistic driving data collected in the UDRIVE project. This was accomplished by identifying instances and extracting segments providing context of where and when VRUs disappeared from the ME-data, that is when the VRU went out of field of view/detection range of the ME-system and then determining the VRUs position in relation to the lane and the ego vehicle. This included an approach on how to keep track of the road shoulder when the ego vehicle was leaving the outermost lane. Continuing, from these extracted segments four time-frames, so called phases, were extracted; two phases were manually annotated and two were automatically derived via extrapolating the longitudinal position of the VRU. Additionally, comfort zone measures for each of the four phases were extracted. Furthermore, this thesis exemplified how to analyze overtaking manoeuvres by dividing the data-set into several categories based on various road/traffic conditions present during the overtaking manoeuvres. Additionally, this study showed that annotation can be a double-edged sword; It is a great way to verify and enrich a subset of data, but if used without caution it can also introduce bias into the subset of data. Quality control in naturalistic driving data is difficult [1]. In this study a set of quality validation procedures were in place. However, late in the work a few fundamental quality issues were identified making interpretation and comparisons of overtaking manoeuvres of bicyclists based on the quantitative results from this thesis problematic.

The main signals that were used in the reduction of the data in this study were the lateral distance to the lane edge and the lateral distance to the VRU. Both were signals from the ME-system. Based on the segments that were annotated, the data-reduction to isolate overtaking manoeuvres had a success-rate of roughly 40% (211 of circa 500). A qualitative validation reduced the number of entries down to 79. Examples of quality issues identified were: signals partly or entirely consisting of NaN values, speed of the VRU exceeding 40 km/h, and distance to lane edge being negative (i.e. implying that the ego vehicle is located on the road shoulder). Only entries which had the central line marking present made it pass the quality check suggesting that either the distance to lane-edge ME-data is more reliable for entries where the road is divided by a central line or that the algorithms implemented in this thesis favored such entries. However, the main problem identified stemmed from the speed of the VRU. It was not derived as accurate as initially expected due to neglecting the eventual presence of noise in the signal for longitudinal distance to the VRU. This uncertainty then imbued the CZB measures (minimum distance, lateral clearance and time to collision) and thus also any eventual comparative analysis done with previously conducted studies.

However, this work shows the potential of using naturalistic driving data for analysis of vehicles' interaction with vulnerable road users. Even if data quality issues were identified, the methodological aspects of the work and the preliminary demonstration of quantification of comfort zone boundary metrics using naturalistic driving data have shown to be sound and progressive.

6.2 Future Work

This thesis is a step in the right direction and has provided knowledge of for example how to extract overtaking segments from a naturalistic driving data-base. But this thesis is only one step of many yet to be made. In other words much work is yet awaiting to done. To narrow down future work a list is presented consisting of potential problems to solve, ideas on how to approach the problems and eventual future use cases of the results acquired in this study.

1. To help determining the ecological validity of driving simulator experiments for the measure of comfort zone boundaries during overtaking maneuvers, future studies may compare how comfort zone measures from this study relate to measures from simulator studies.

2. The UDRIVE data base could support a comparative analysis of data collected in different countries to identify potential national driver culture. This would require more data to be verified.

3. The measures from this study could also aid in the development of driver behaviour models for overtaking manoeuvres.

4. The data set extracted in this study could take on the role as training and/or verification data set for future studies using a machine learning approach (e.g. support vector machines), to classify overtaking manoeuvres.

5. For future work using UDRIVE or similar data it is suggested that a more thorough qualitative check is done on the data, for example by comparing the signal-behaviour with the video feed and see if the message given by signals matches the message given by the video feed.

6. Comparative analysis could be done with more/other "types" of studies, for instance studies using naturalistic driving data collected via instrumented bicycles. Analysis could also be performed with models such as linear regression.

7. Another suggestion is to keep all road objects in the ME-data separate when deriving comfort zone measures. That is instead of only deriving the CZB for the closest VRU the CZB could be derived to oncoming traffic etc. Some issues that comes with such an approach and thus has to be resolved are:

   (a) The ME system sometimes mislabels objects (e.g. bicyclists are labeled as pedestrian) for a few time instances/samples before the correct label is present and thus a decision has to be made on whether to merge or discard data for such events.

   (b) When road objects disappear ME "reassigns" road objects. The ME data keep track of up to four road objects (i.e. RO1, RO2, RO3 and RO4). In order for RO2 to be assigned data, RO1 must already be assigned data, and for RO3 to be assigned data RO2 must already be assigned data etc. In a scenario where road user X is tracked in RO1 and then road user Y suddenly appears it is tracked in RO2. If road user X suddenly disappears, road user Y is tracked in RO1 instead of RO2 for future time instances. Fortunately, this issue should be feasible to solve as the ME-data contains an identifier value for each object tracked.

8. If a derivation of the velocity of the VRU is not successfully implemented to allow an accurate extrapolation of positional data, forthcoming work might need to consider reformulating the definition of the phases and/or how to derive them. Furthermore, if phase one and phase two are not defined manually via annotation (i.e. defined automatic instead), the synchronization error would pose much less of a problem, that is the error would supposedly only be a slight inconvenience when comparing signals to video in a validation process of data. One suggestion is to redefine the start of phase one as when the VRU first is appearing in the ME-data.

References

[1] J. Bärgman, K. Smith, and J. Werneke. Quantifying drivers' comfort-zone and dread-zone boundaries in left turn across path/opposite direction (LTAP/OD) scenarios. *Transportation Research Part F: Traffic Psychology and Behaviour* **35** (2015), 170–184.

[2] M. Bíl, M. Bílova, and I. Müller. Critical factors in fatal collisions of adult cyclists with automobiles. *Accident Analysis & Prevention* **42** (2010), 1632–1636.

[3] S. Boufous et al. Risk factors for severe injury in cyclists involved in traffic crashes in Victoria, Australia. *Accident Analysis & Prevention* **49** (2012), 404–409.

[4] *Cambridge Dictionary.* `http://dictionary.cambridge.org/dictionary/english/overtake` 2017-02-08.

[5] J. R. Chapman and D. A. Noyce. Influence of roadway geometric elements on driver behavior when overtaking bicycles on rural roads. *Accident Analysis & Prevention* **1** (2014), 28–38.

[6] E. S. Chataway et al. Safety perceptions and reported behavior related to cycling in mixed traffic: A comparison between Brisbane and Copenhagen. *Accident Analysis & Prevention* **23** (2014), 32–43.

[7] N. Chaurand and P. Delhomme. Cyclists and drivers in road interactions: A comparison of perceived crash risk. *Accident Analysis & Prevention* **56** (2013), 1176–1184.

[8] K.-H. Chuang et al. The use of quasi-naturalistic riding method to investigate bicyclists' behaviors when motorists pass. *Accident Analysis & Prevention* **56** (2013), 32–41.

[9] M. Dozza et al. How do drivers overtake cyclists? *Accident Analysis & Prevention* **88** (2016), 29–36.

[10] S. Fatahtooei Nejad. "Drivers' comfort zone boundaries during overtaking of bicycles in Japan without oncoming traffic". MA thesis. Chalmers University of Technology, 2017.

[11] G. Hegeman, K. Brookhuis, and S. Hoogendoorn. Opportunities of Advanced Driver Assistance Systems towards Overtaking. *European Journal of Transport and Infrastructure Research* **5** (2005), 281–296.

[12] P. E. Hemeren et al. "The Use of Visual Cues to Determine the Intent of Cyclists in Traffic". *2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA).* 2014.

[13] P. E. Hemeren et al. *Uppmärksamhet hos Bilförare i Möte med Oskyddade Trafikanter del II: Synbarhet och Sociala Signaler.* Tech. rep. Institutionen för Informationteknologi, 2014.

[14] J.-K. Kim et al. Bicyclist injury severities in bicycle–motor vehicle accidents. *Accident Analysis & Prevention* **39** (2007), 238–251.

[15] C. Llorca et al. Motor vehicles overtaking cyclists on two-lane rural roads: Analysis on speed and lateral clearance. *Safety Science* **92** (2017), 302–310.

[16] D. C. Love et al. Is the three-foot bicycle passing law working in Baltimore, Maryland? *Accident Analysis & Prevention* **48** (2012), 451–456.

[17] *MATLAB.* `https://se.mathworks.com/products/matlab.html` 2017-05-31.

[18] *MATLAB Spline Interpolation.* `https://se.mathworks.com/help/matlab/ref/interp1.html?#bvkp4l9` 2017-06-02.

[19] T. M. Matson and T. W. Forbes. *Overtaking and Passing Requirements as Determined from a Moving Vehicle.* Proceedings of the Highway Research Board, 1938, pp. 100–112.

[20] L. B. Meuleners, A. H. Lee, and C. Haworth. Road environment, crash type and hospitalisation of bicyclists and motorcyclists presented to emergency departments in Western Australia. *Accident Analysis & Prevention* **39** (2007), 1222–1225.

[21] C. Moretto. "Drivers' comfort zone boundaries during overtaking of bicycles in Japan". MA thesis. Chalmers University of Technology, 2017.

[22] J. Olivier and S. R. Walter. Bicycle Helmet Wearing Is Not Associated with Close Motor Vehicle Passing: A Re-Analysis of Walker, 2007. *Public Library of Science* **8** (2013).

[23] *Oxford Dictionaries.* `https://en.oxforddictionaries.com/definition/overtake` 2017-02-08.

[24] J. Parkin and C. Meyers. The Effect of Cycle Lanes on the proximity between motor traffic and cycle traffic. *Accident Analysis & Prevention* **42** (2010), 159–165.

[25] C. C. Reynolds et al. The Impact of Transportation Infrastructure on Bicycling Injuries and Crashes: a Review of the Literature. *Environmental Health* **8** (2009), 47.

[26] T. Richter and S. Ruhl. *Untersuchung von Maßnahmen zur Prävention von Überholunfällen auf einbahnigen Landstraßen.* Tech. rep. 24. Gesamtverband der Deutschen Versicherungswirtschaft, 2014.

[27] P. Savolainen et al. Lateral Placement of Motorists when Passing Bicyclists. *Transportation Research Record Journal of the Transportation Research Board* **2314** (2012), 14–21.

[28] R. Schindler and V. Bast. “Drivers’ comfort boundaries when overtaking a cyclist”. MA thesis. Chalmers University of Technology, 2015.

[29] K. Schleinitz et al. The influence of speed, cyclists’ age, pedaling frequency, and observer age on observers’ time to arrival judgments of approaching bicycles and e-bikes. *Accident Analysis & Prevention* **92** (2016), 113–121.

[30] T. Shamir. How should an autonomous vehicle overtake a slower moving vehicle: design and analysis of an optimal trajectory. *IEEE Transactions on Automatic Control* **49** (2004), 302–310.

[31] *Svenska Akademiens Ordbok.* `http://www.saob.se/artikel/?seek=omk%C3%B6ra&pz=1#U_O569_261824` 2017-02-08.

[32] Trafikverket. *Säkrare cykling – gemensam strategi för år 2014-2020.* Tech. rep. Trafikverket, 2014.

[33] *UDrive Deliverable D24.1.* 2017.

[34] *Vulnerable Road Users.* `https://www.swov.nl/sites/default/files/publicaties/gearchiveerde-factsheet/uk/fs_vulnerable_road_users_archived.pdf`. 2012.

[35] I. Walker. Drivers overtaking bicyclists: Objective data on the effects of riding position, helmet use, vehicle type and apparent gender. *Accident Analysis & Prevention* **39** (2007), 417–425.

[36] I. Walker, I. Garrard, and F. Jowitt. The Influence of a Bicycle Commuter’s appearance on drivers’ overtaking proximities: An on-road test of bicycle stereotypes, high-visibility clothing and safety aids in the United Kingdom. *Accident Analysis & Prevention* **64** (2014), 69–77.

[37] *What is a reasonable speed for long distances on a bike?* `https://bicycles.stackexchange.com/questions/1103/what-is-a-reasonable-speed-for-long-distances-on-a-bike` 2017-05-31.

[38] J. M. Wood et al. Bicyclists overestimate their own night-time conspicuity and underestimate the benefits of retroreflective markers on the moveable joints. *Accident Analysis & Prevention* **55** (2013), 48–53.

A Tables of Variables

In this appendix available signals and attributes are listed with a brief description, abbreviation etc. Not all signals listed in the tables were used but were considered at one point or another during the execution of this study. Also, not all signals available in the database are listed here, however all signals used in this study are. The main purpose of including these tables is to provide an overview of the range of signals via minimum, maximum, standard deviation and mean values.

Table A.1: All the signals extracted from every segment accompanied with the abbreviation and a brief description of the variable

Variables	**Abbreviation**	**Description of the Variable**
Time	T	The time resolution is 0.1 s
LateralDistanceToVRU	LatD	The lateral distance to the closest VRU from the ego vehicle (discarding dimension of objects*).
Lateral Clearance	LatC	Lateral distance between closest VRU and ego Vehicle taking into account the width of ego vehicle and the VR
Minimum Distance	MD	Minimal pythagorian distance between ego vehicle and VRU, taking into account the dimensions of the objects*
LongituidnalDistanceToVRU	LongD	The longitudinal distance to the closest VRU from the ego vehicle (discarding dimension of objects*).
Speed	V or VehV	Speed of the ego vehicle
VRUVelocity	V_{VRU}	Speed of the VRU
RelativeVelocityToVRU	RV	Relative velocity between the closest VRU and the ego vehicle. (Extrapolated ME-data)
RelativeVelocityToVRU	RV2	Relative velocity between the closest VRU and the ego vehicle. (Based on ME longitudinal distance)
DistanceToLaneEdge	D2Lane or D2L	Distance between the ego vehicle and the lane edge (laneshoulder if travelling in the outmost lane)
DistanceToAdjacentLane	D2ALane or D2AL	Distance between the ego vehicle and the adjacent edge
LateralAcceleration	LatA	Lateral acceleration of the ego vehicle
LongitudinalAcceleration	LongA	Longitudinal acceleration of the ego vehicle
DistanceBetweenLaneEdgeAndVRU	DBLE_VRU	Distance between the VRU closest to the vehicle and the lane edge
THW	THW	Time headway to the closest VRU based on the longitudinal distance to the closest VRU
TTC	TTC	Time to collision to the closest VRU based on the longitudinal distance to the closest VRU
VRUwidth	VRUW	Width of the VRU
YawRate	YawR	Yawrate
SteeringWheel	SteWl	Steeringwheel angle
THW_to_Vehicle	THWc	Time headway to the closest vehicle
Radius	R	Road curvature
SpeedLimit	SpLim	Road speedlimit
Inclination	Inc	Road inclination
TurnIndicator	TurnI	Turn indicator lights status, either on or off.
Time of Day	TimeOfDay	The time of day expressed as a MATLAB date-string
GPS_Lat	GLat	Latitudinal position of the ego vehicle
GPS_Long	GLong	Longitudinal position of the ego vehicle
Driver ID	DriverID	Individual identification number for drivers within the UDRIVE database
Vehicle ID	VehID	Identification number for vehicle model
CountryInfo	OS	Operation site; In which country the trip was recorded
DayType	DayType	If the trip was recorded during dawn, day, dusk or night
LaneWidth	LaneW	The dynamic width of the lane in which the ego vehicle is located
BeginPhase1Index	ME_C	Index where the ME-system detects the VRU-object
BeginPhase3Index	I3	Index where the Long D-signal first reaches below 3 meters
BeginPhase4Index	I4	Index where the Long D-signal first reaches below -3 meters
EndPhase4Index	I4End	Index where the distance travelled since phase 4 began reaches above 50 meters.

*Dimensions of objects refer to the length and width of the ego vehicle and VRU

Table A.2: All the attributes annotated for every segment accompanied with the abbreviation and a brief description of the variable.

Variables	**Abbrevation**	**Description of the Variable**
IsOvertaking	isOT	Boolean verifying that the ego vehicle performed an overtaking of VRU during the segment
MultipleVRU	MultipleVRU	Boolean indicating that the ego vehicle overtook more than one VRU during the segment
TypeOfVRU	VRUtp	Categorical: bicyclist, pedestrian, motorbike or other VRU
SameDir	VRUSameDir	Boolean indicating whether the VRU was travelling in the same direction as the ego vehicle.
ManoueverType	OT	Categorical: flying or accelerative overtaking
StraigthRoad	StraRo	Boolean indicating whether the ego vehicle performed the overtaking on a straigth road
LeadingVehiclePresent	Lead or L	Boolean indicating whether the ego vehicle performed the overtaking while piggybacking (THW ¡ 3s)
OncomingTrafficPresent	Onc or O	Boolean indicating whether there was traffic in the adjacent lane while performing the overtaking
AnnotatedPhase1Index	I1	Index where the to-be-overtaken VRU is distinguishable in the video-feed
AnnotatedPhase2Index	I2	Index where the ego vehicle can be seen to adjust its trajectory due to the VRU in the video-feed
SyncError	SErr	How much time the signal comes before the video-feed.

Table A.3: All the variables available for each segment accompanied with the origin (i.e. raw data from CAN or ME, or derived from the raw data), whether it is continuous or categorical, the unit and the range (min/max/standard deviation/mean) of the variables.

Abbrevation	**Origin**	**Unit**	Min	Max	Std	Mean
T	Raw	Continuous [s]	-	-	-	-
LatD	Derived	Continuous [m]	-4.058	8.667	1.016	2.033
LatC	Derived	Continuous [m]	-5.347	7.331	1.009	0.7271
CZB	Derived	Continuous [m]	0.0369	791.1	157.0	190.7
LongD	Derived	Continuous [m]	-774.8	791.1	246.4	30.62
V	Raw	Continuous [km/h]	17.04	106.4	15.48	63.71
V_{VRU}	Derived	Continuous [km/h]	0.4748	39.83	8.759	21.73
RV	Derived	Continuous [km/h]	-31.52	10	5.755	-8.525
RV2	Derived	Continuous [km/h]	-78.15	0.0144	15.832	-39.63
D2Lane	Derived	Continuous [m]	-4.415	6	0.8616	1.6456
D2ALane	Derived	Continuous [m]	-3.902	5.337	0.7134	1.235
LatA	Raw	Continuous [m/s^2]	-0.4015	0.4491	0.0743	-0.0111
LongA	Raw	Continuous [m/s^2]	-0.4750	0.3377	0.0842	0.0059
DBLE_VRU	Derived	Continuous [m]	-6.8978	4.9125	0.8613	-0.6796
THW	Derived	Continuous [s]	-16.42	15.23	3.75	0.4911
TTC	Derived	Continuous [s]	-498.2	36.31	7.757	0.8224
VRUW	Derived	Continuous [m]	0.3905	0.6709	0.0624	0.5921
YawR	Raw	Continuous [$degrees/s$]	-17.00	6554	3258	3273
SteWl	Raw	Continuous [$degrees$]	-3277	72.40	958.6	-310.2
THWc	Raw	Continuous [s]	0.3118	14.6840	1.8859	2.1422
R	Raw	Continuous [m?]	-983	975	271.2	4.879
SpLim	Raw	Continuous [km/h]	48	100	12.65	87.58
Inc	Raw	Continuous [$degrees$]	-2504	7.047	143.4	-67.00
TurnI	Raw	Continuous [N/A]	-	-	-	-
TimeOfDay	Raw	Continuous [$dd-mm-yy$]	-	-	-	-
GLat	Raw	Continuous [$degrees$]	-	-	-	-
GLong	Raw	Continuous [$degrees$]	-	-	-	-
DriverID	Raw	Categorical [N/A]	-	-	-	-
VehID	Raw	Categorical [N/A]	-	-	-	-
OS	Raw	Categorical [N/A]	-	-	-	-
Dtp	Raw	Categorical [$dawn/day/dusk/night$]	1	4	0.2802	2.0207
LaneW	Derived	Continuous [m]	-2.6471	8.3586	0.7409	2.8807
ME_C	Derived	[N/A]	-	-	-	-
I3	Derived	[N/A]	-	-	-	-
I4	Derived	[N/A]	-	-	-	-
I4e	Derived	[N/A]	-	-	-	-
I1	Annotated	[N/A]	-	-	-	-
I2	Annotated	[N/A]	-	-	-	-
isOT	Annotated	Boolean	-	-	-	-
MultipleVRU	Annotated	Boolean	-	-	-	-
StraRo	Annotated	Boolean	-	-	-	-
VRUtp	Annotated	Categorical [N/A]	-	-	-	-
OT	Annotated	Categorical [N/A]	-	-	-	-
Lead	Annotated	Boolean	-	-	-	-
Onc	Annotated	Boolean	-	-	-	-
SErr	Annotated	[s]	-	-	-	-

B Data Sorted by the Overtaking Strategy

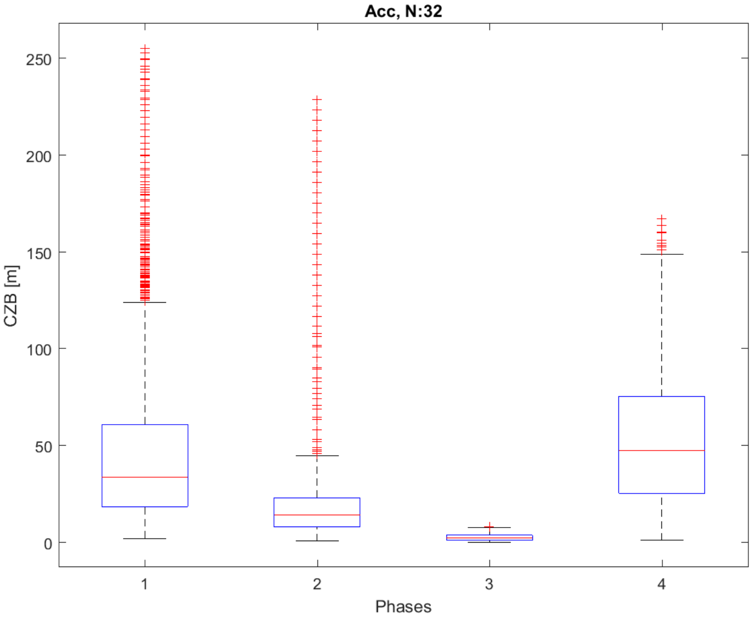


Figure B.1: *Boxplot of minimum distance (CZB) for accelerative overtaking strategy.*

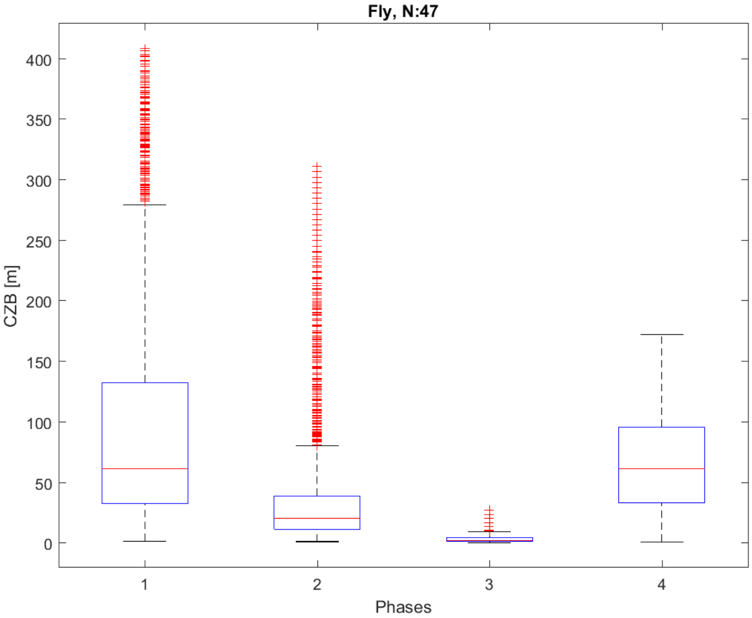


Figure B.2: *Boxplot of minimum distance (CZB) for flying overtaking strategy.*

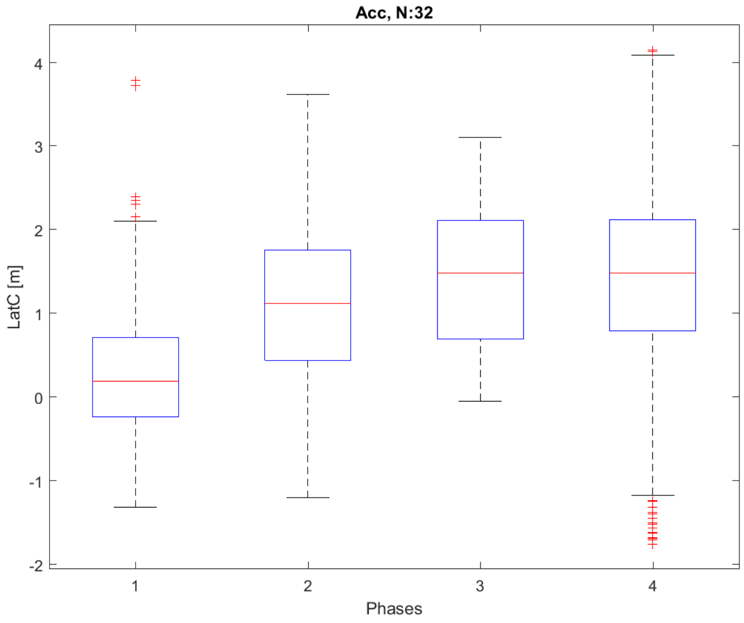


Figure B.3: *Boxplot of lateral clearance for accelerative overtaking strategy.*

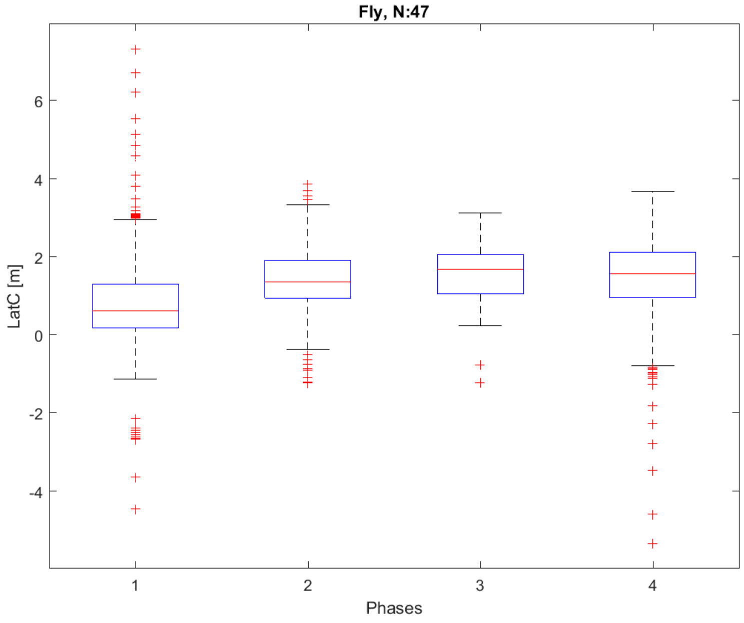


Figure B.4: *Boxplot of lateral clearance for flying overtaking strategy.*

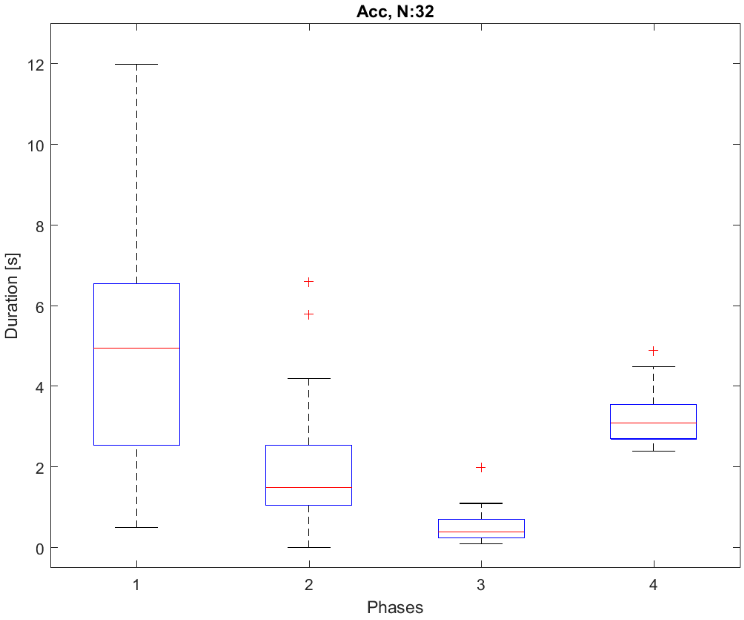


Figure B.5: *Boxplot of phase duration for accelerative overtaking strategy.*

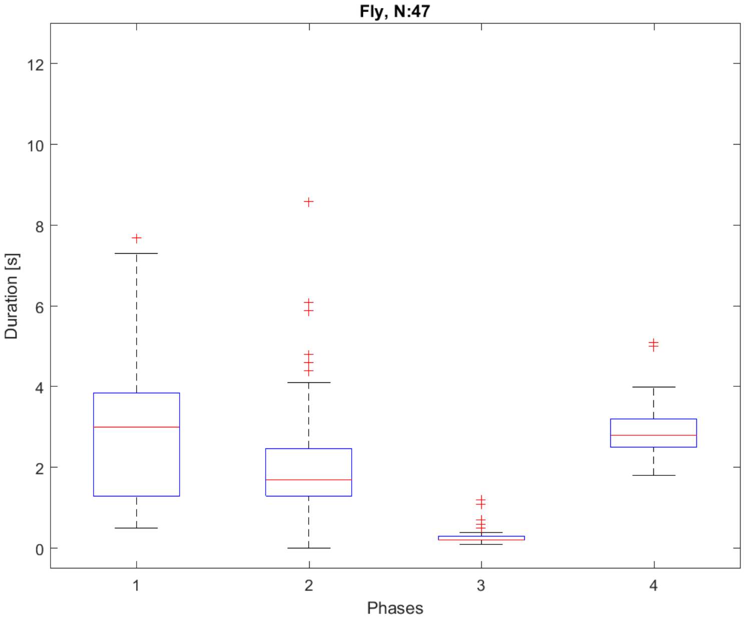


Figure B.6: *Boxplot of phase duration for flying overtaking strategy.*

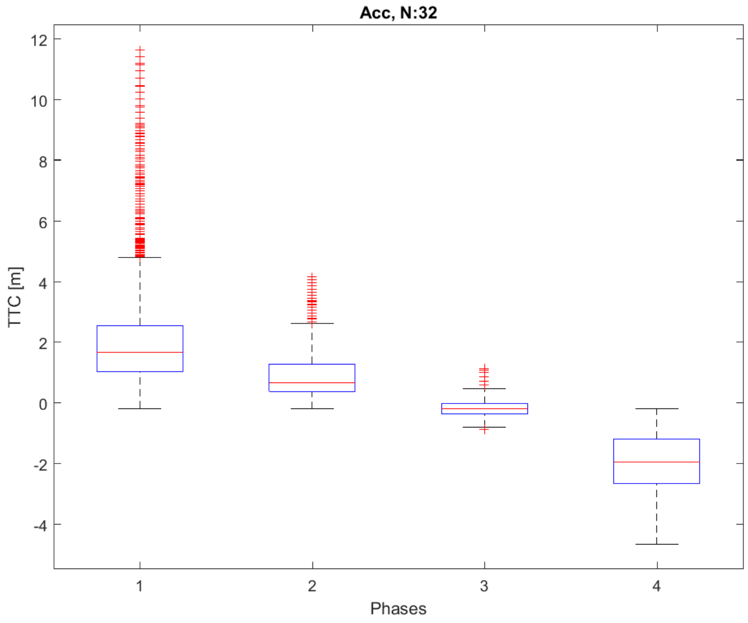


Figure B.7: *Boxplot of time to collision for accelerative overtaking strategy.*

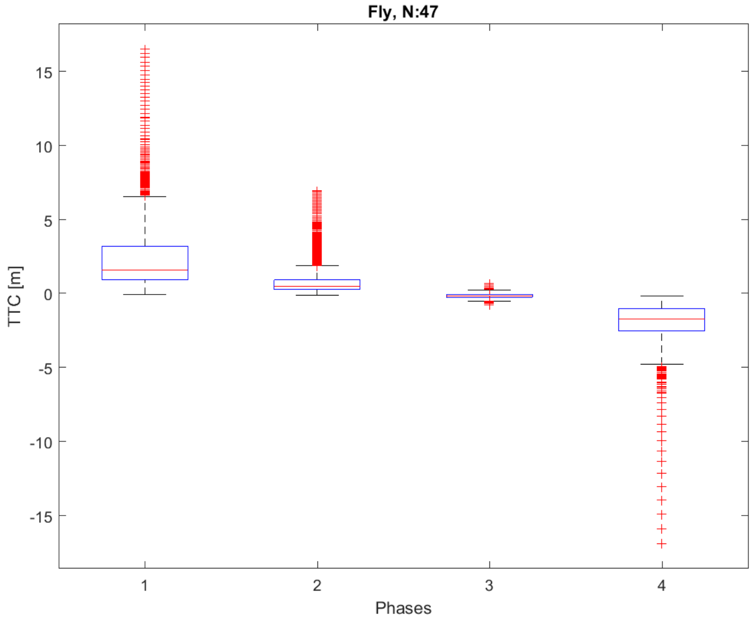


Figure B.8: *Boxplot of time to collision for flying overtaking strategy.*

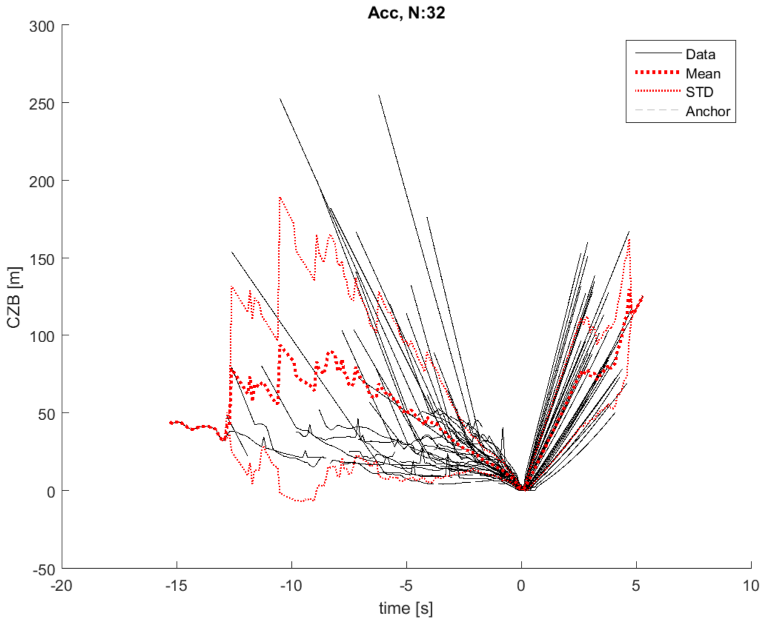


Figure B.9: *Minimum distance (CZB) for accelerative overtaking strategy.*

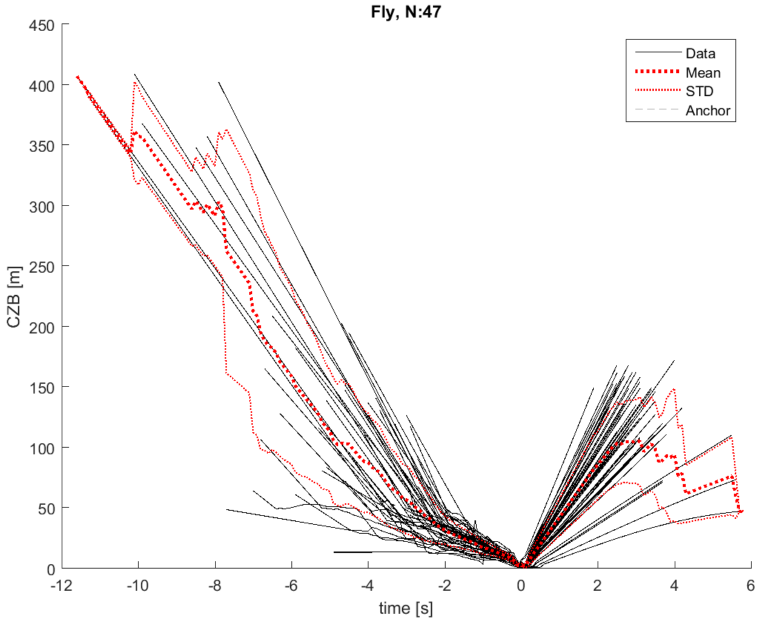


Figure B.10: *Minimum distance (CZB) for flying overtaking strategy.*

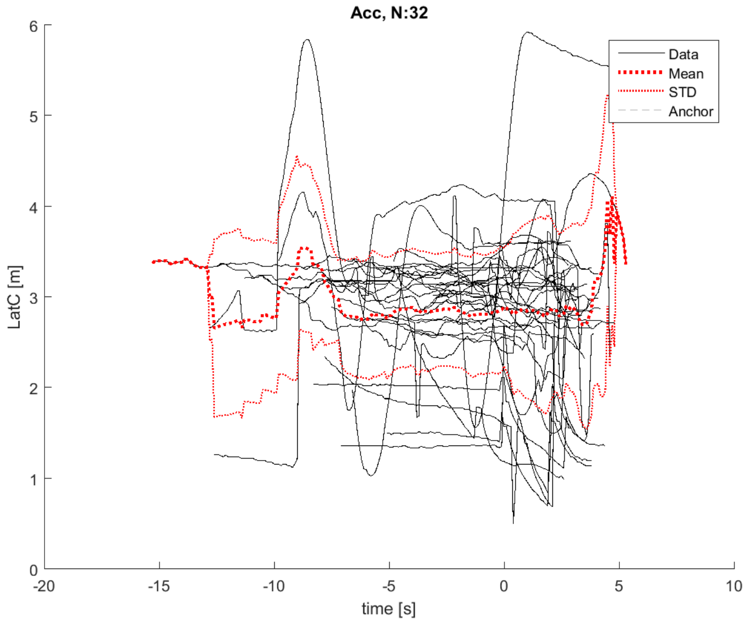


Figure B.11: *Lane width for accelerative overtaking strategy.*

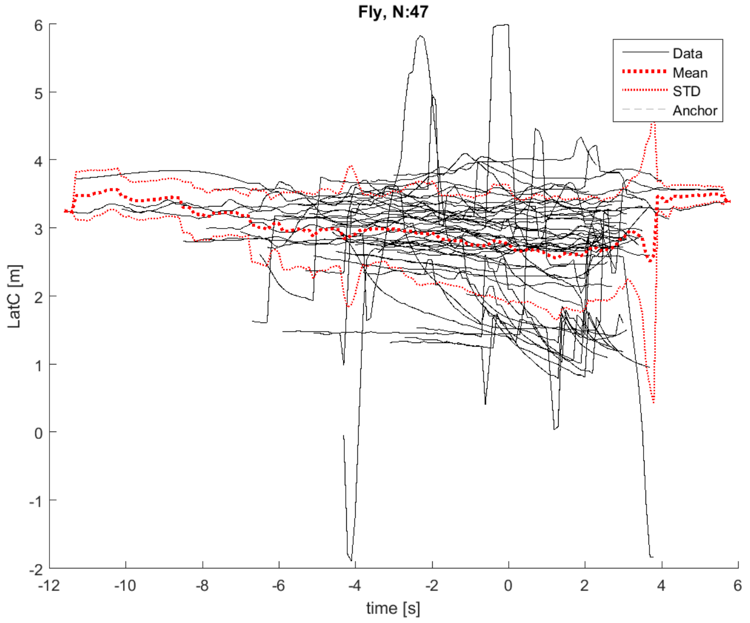


Figure B.12: *Lane width for flying overtaking strategy.*

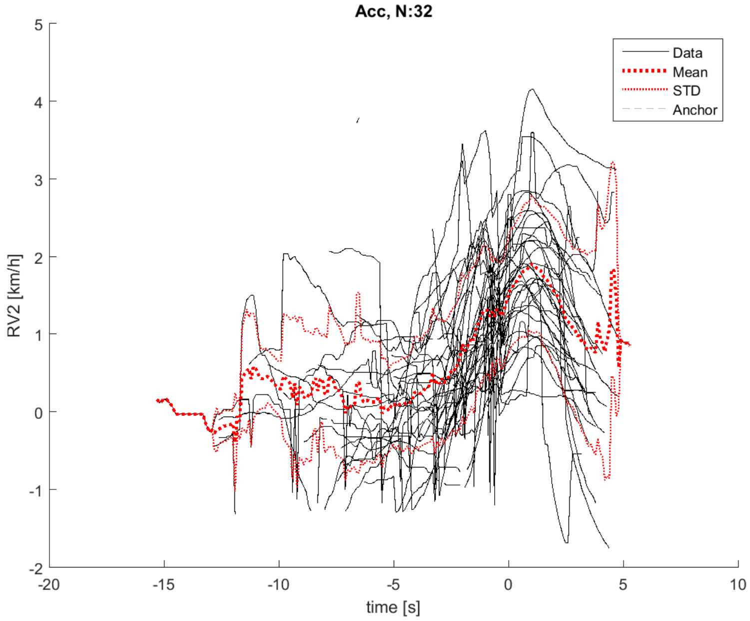


Figure B.13: *Lateral clearance for accelerative overtaking strategy.*

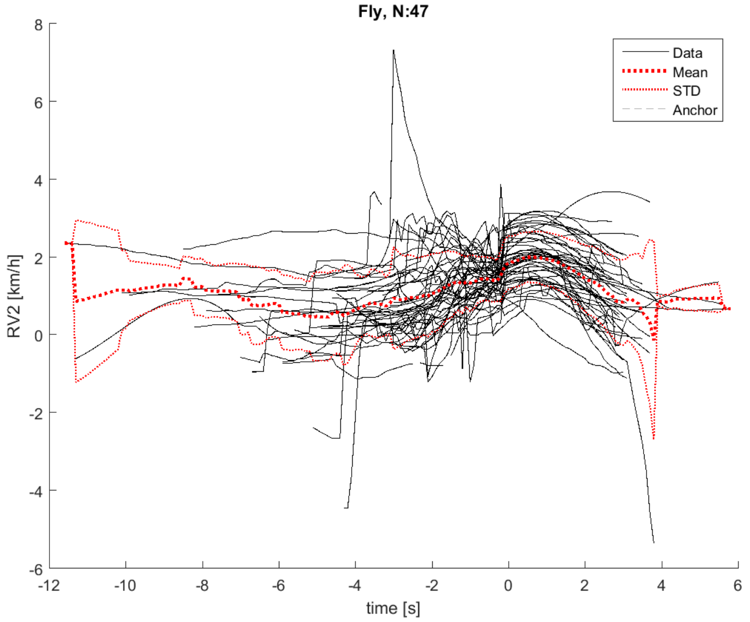


Figure B.14: *Lateral clearance for flying overtaking strategy.*

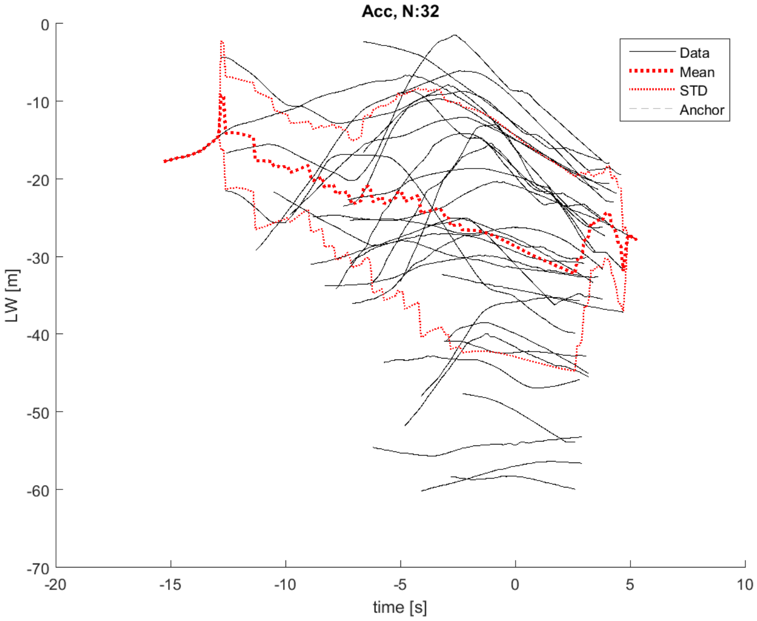


Figure B.15: *Relative velocity to VRU for accelerative overtaking strategy.*

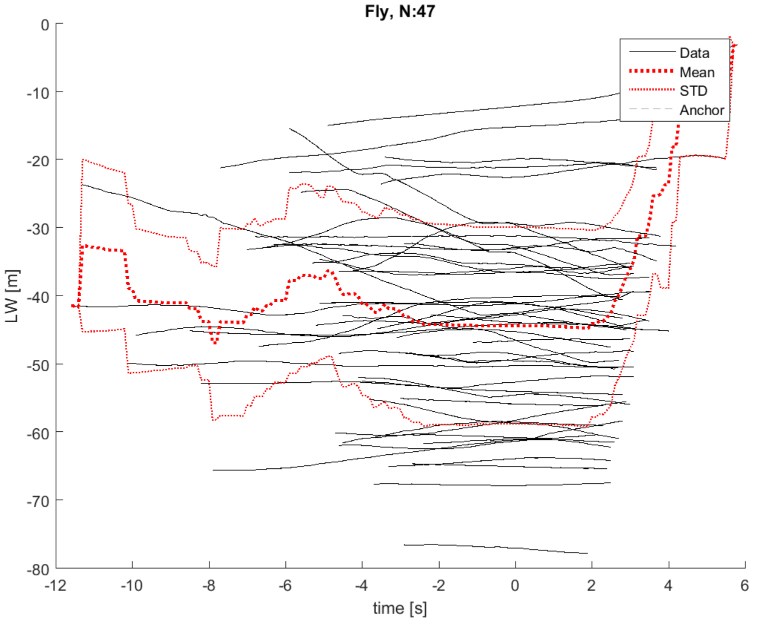


Figure B.16: *Relative velocity to VRU for flying overtaking strategy.*

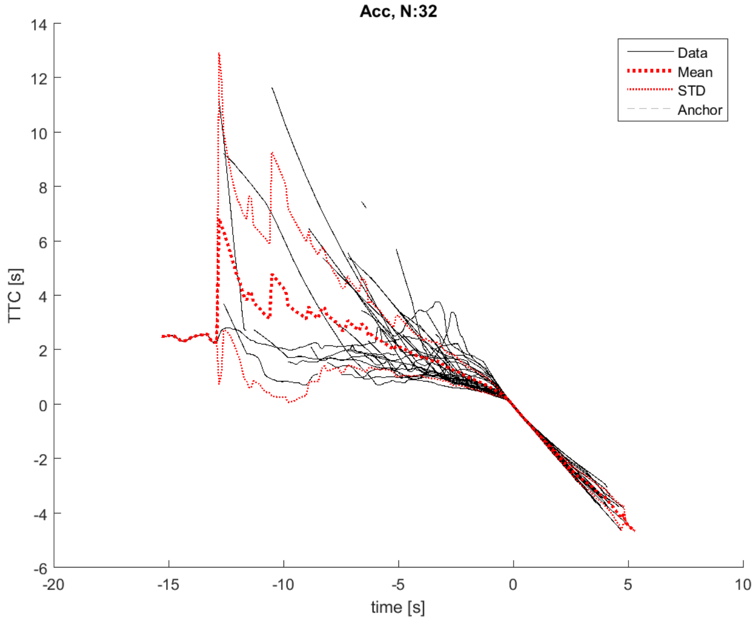


Figure B.17: *Time to collision to VRU for accelerative overtaking strategy.*

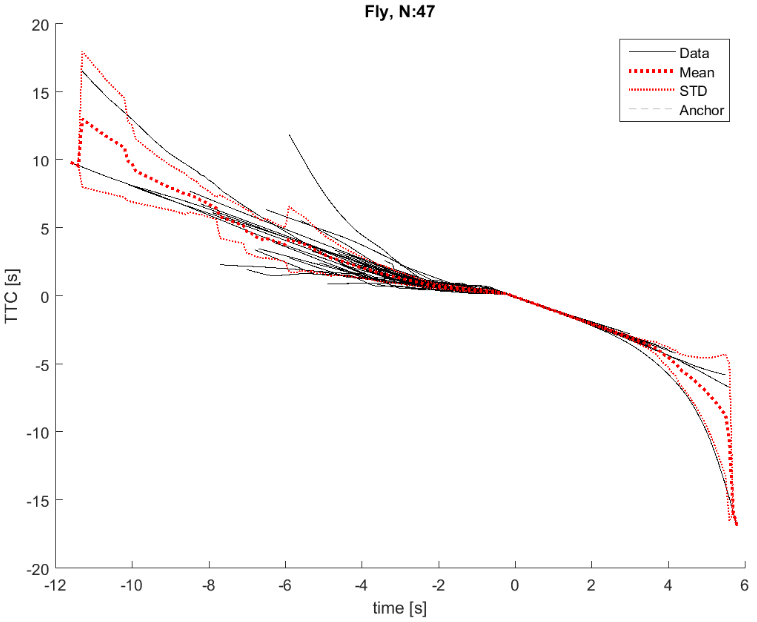


Figure B.18: *Time to collision to VRU for flying overtaking strategy.*

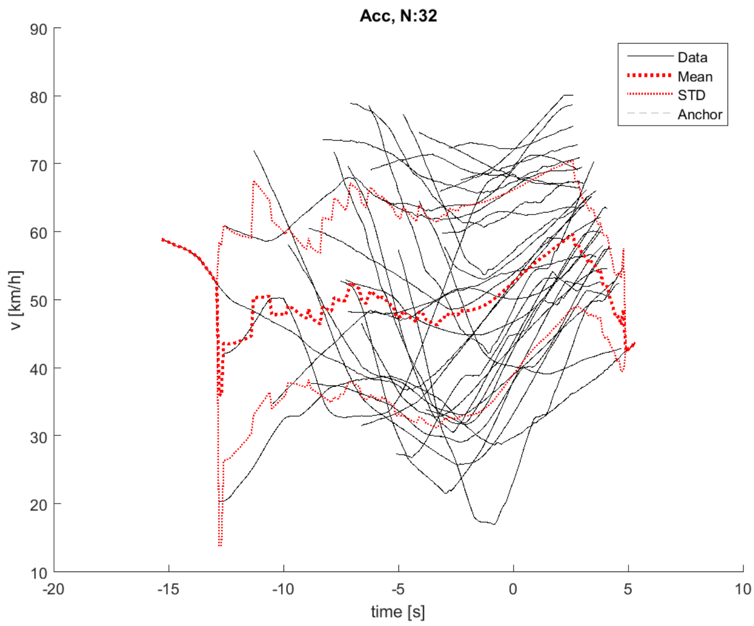


Figure B.19: *Ego vehicle speed for accelerative overtaking strategy.*

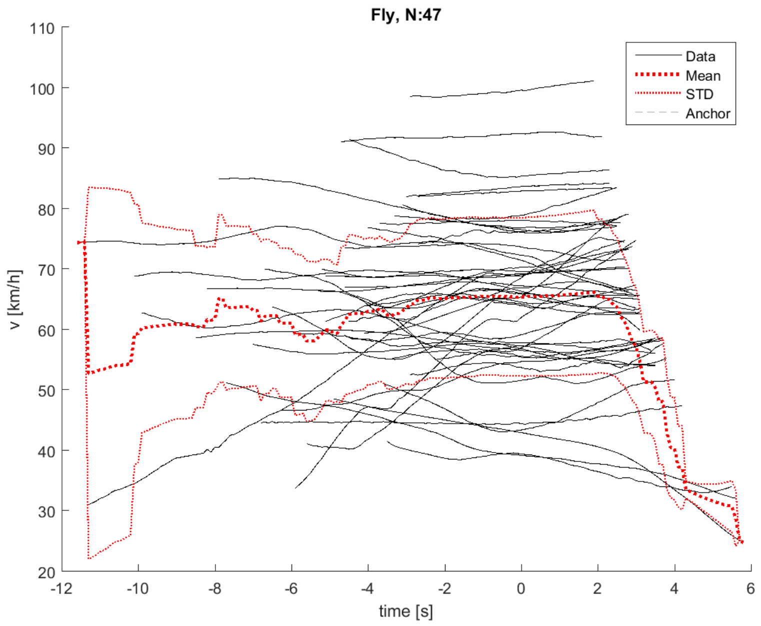


Figure B.20: *Ego vehicle speed for flying overtaking strategy.*

C Data Sorted by Presence of Oncoming Traffic

In the following tables 'Onc. Present' implies that a oncoming traffic was present and 'Onc. Not Present' implies that oncoming traffic was not present during the execution of the overtaking manoeuvre.

C.1 Tables

Table C.1: TTC [s]. Data sorted by overtaking manoeuvre strategy and presence of oncoming traffic.

			Phase 1	Phase 2	Phase 3	Phase 4
Accelerative	Oncoming Traffic Present	Mean	1.98	0.90	-0.13	-2.00
		STD	±1.62	±0.69	±0.36	±0.98
	Oncoming Traffic Not Present	Mean	2.52	0.91	-0.18	-1.87
		STD	±1.98	±0.80	±0.23	±0.89
Flying	Oncoming Traffic Present	Mean	1.16	0.39	-0.06	-2.51
		STD	±0.62	±0.23	±0.33	±2.54
	Oncoming Traffic Not Present	Mean	2.96	0.99	-0.16	-1.80
		STD	±2.79	±1.26	±0.13	±1.08

Table C.2: p-values for TTC. Phase 1. Data Sorted by Presence of Oncoming Traffic.

		Accelerative		Flying	
		Onc. Present	Onc. Not Present	Onc. Present	Onc. Not Present
Accelerative	Onc. Present	-	**$p<0.001$**	**$p<0.001$**	**$p<0.001$**
	Onc. Not Present	-	-	**$p<0.001$**	**$p<0.001$**
Flying	Onc. Present	-	-	-	**$p<0.001$**
	Onc. Not Present	-	-	-	-

Table C.3: p-values for TTC. Phase 2. Data Sorted by Presence of Oncoming Traffic.

		Accelerative		Flying	
		Onc. Present	Onc. Not Present	Onc. Present	Onc. Not Present
Accelerative	Onc. Present	-	0.75	**$p<0.001$**	0.19
	Onc. Not Present	-	-	**$p<0.001$**	0.29
Flying	Onc. Present	-	-	-	**$p<0.001$**
	Onc. Not Present	-	-	-	-

Table C.4: p-values for TTC. Phase 3. Data Sorted by Presence of Oncoming Traffic.

		Accelerative		Flying	
		Onc. Present	Onc. Not Present	Onc. Present	Onc. Not Present
Accelerative	Onc. Present	-	0.33	0.27	0.45
	Onc. Not Present	-	-	0.02	0.43
Flying	Onc. Present	-	-	-	**0.006**
	Onc. Not Present	-	-	-	-

Table C.5: p-values for TTC. Phase 4. Data Sorted by Presence of Oncoming Traffic.

		Accelerative		Flying	
		Onc. Present	Onc. Not Present	Onc. Present	Onc. Not Present
Accelerative	Onc. Present	-	0.03	**p<0.001**	**p<0.001**
	Onc. Not Present	-	-	**p<0.001**	0.21
Flying	Onc. Present	-	-	-	**p<0.001**
	Onc. Not Present	-	-	-	-

Table C.6: LatC [m]. Data Sorted by Presence of Oncoming Traffic.

			Phase 1	Phase 2	Phase 3	Phase 4
Accelerative	Oncoming Traffic Present	Mean	0.14	0.95	1.04	1.35
		STD	±0.63	±0.70	±0.73	±1.00
	Oncoming Traffic Not Present	Mean	0.44	1.23	2.00	1.57
		STD	±0.76	±0.89	±0.71	±1.10
Flying	Oncoming Traffic Present	Mean	0.25	1.13	0.97	1.10
		STD	±0.84	±0.73	±0.69	±0.67
	Oncoming Traffic Not Present	Mean	0.85	1.45	1.88	1.64
		STD	±0.96	±0.77	±0.64	±0.96

Table C.7: p-values for LatC. Phase 1. Data Sorted by Presence of Oncoming Traffic.

		Accelerative		Flying	
		Onc. Present	Onc. Not Present	Onc. Present	Onc. Not Present
Accelerative	Onc. Present	-	**p<0.001**	0.02	**p<0.001**
	Onc. Not Present	-	-	**0.001**	**p<0.001**
Flying	Onc. Present	-	-	-	**p<0.001**
	Onc. Not Present	-	-	-	-

Table C.8: p-values for LatC. Phase 2. Data Sorted by Presence of Oncoming Traffic.

		Accelerative		Flying	
		Onc. Present	Onc. Not Present	Onc. Present	Onc. Not Present
Accelerative	Onc. Present	-	**p<0.001**	0.03	**p<0.001**
	Onc. Not Present	-	-	0.10	**0.002**
Flying	Onc. Present	-	-	-	**p<0.001**
	Onc. Not Present	-	-	-	-

Table C.9: p-values for LatC. Phase 3. Data Sorted by Presence of Oncoming Traffic.

		Accelerative		Flying	
		Onc. Present	Onc. Not Present	Onc. Present	Onc. Not Present
Accelerative	Onc. Present	-	**p<0.001**	0.54	**p<0.001**
	Onc. Not Present	-	-	**p<0.001**	0.21
Flying	Onc. Present	-	-	-	**p<0.001**
	Onc. Not Present	-	-	-	-

Table C.10: p-values for LatC. Phase 4. Data Sorted by Presence of Oncoming Traffic.

		Accelerative		Flying	
		Onc. Present	Onc. Not Present	Onc. Present	Onc. Not Present
Accelerative	Onc. Present	-	**0.001**	**p<0.001**	**p<0.001**
	Onc. Not Present	-	-	**p<0.001**	0.18
Flying	Onc. Present	-	-	-	**p<0.001**
	Onc. Not Present	-	-	-	-

C.2 Figures

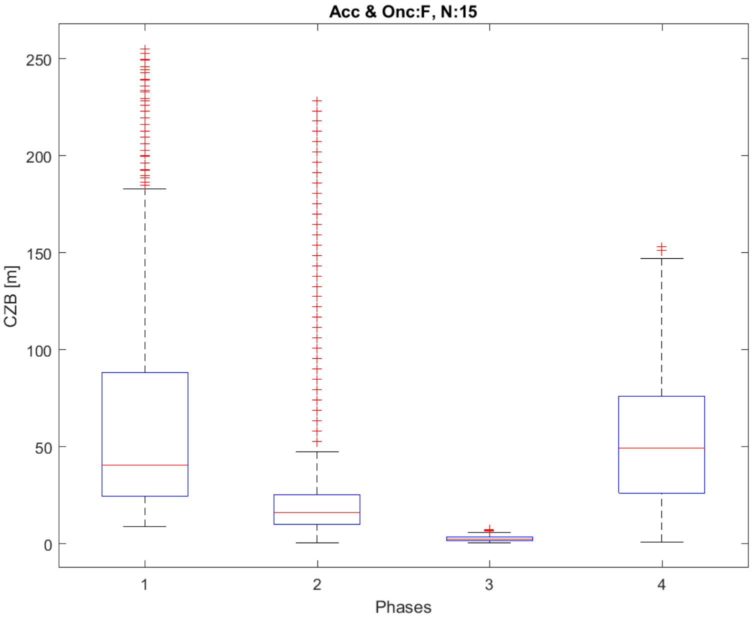


Figure C.1: *Boxplot of minimum distance (CZB) for accelerative overtaking strategy and oncoming traffic not present.*

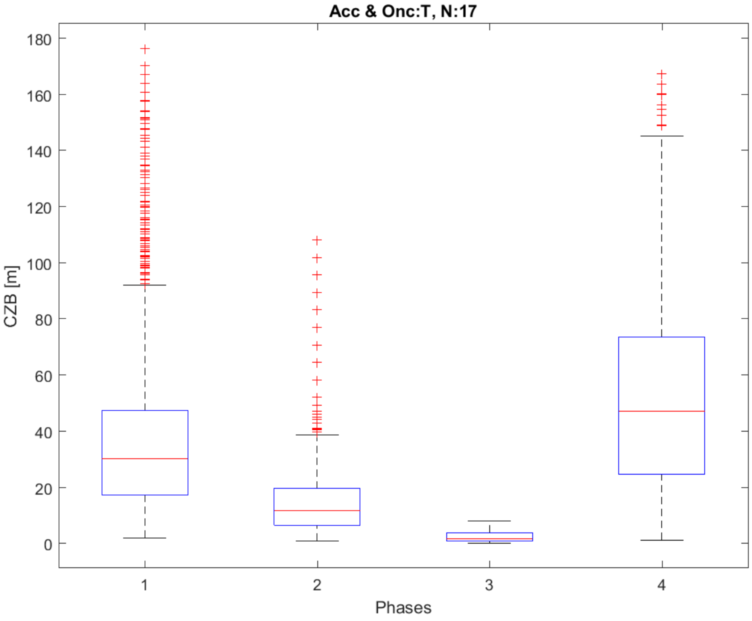


Figure C.2: *Boxplot of minimum distance (CZB) for accelerative overtaking strategy and oncoming traffic present.*

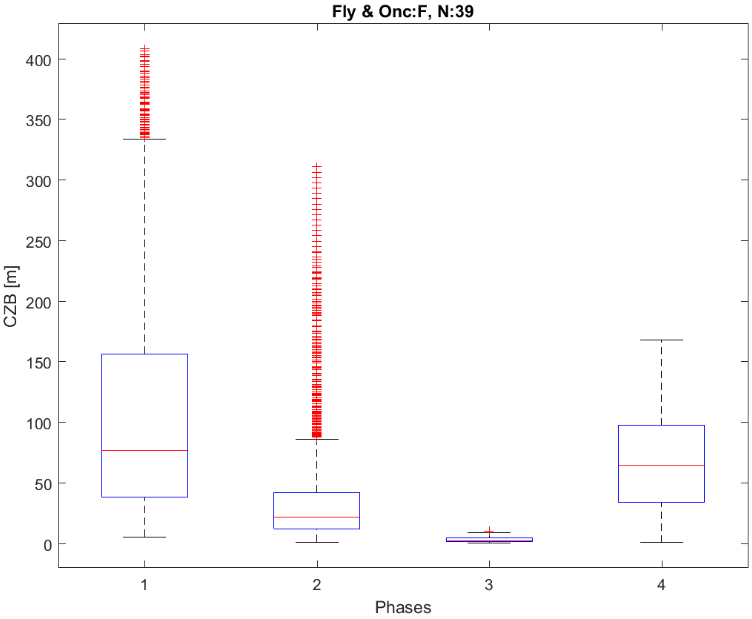


Figure C.3: *Boxplot of minimum distance (CZB) for flying overtaking strategy and oncoming traffic not present.*

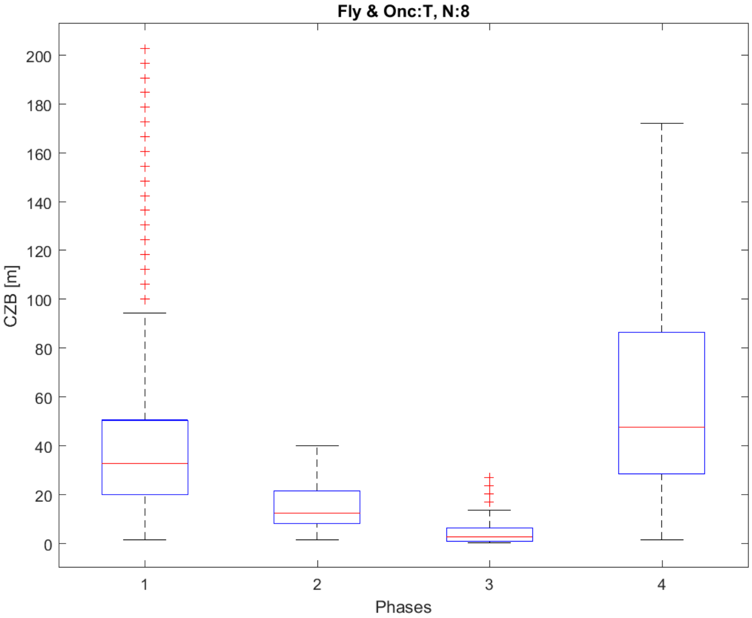


Figure C.4: *Boxplot of minimum distance (CZB) for flying overtaking strategy and oncoming traffic present.*

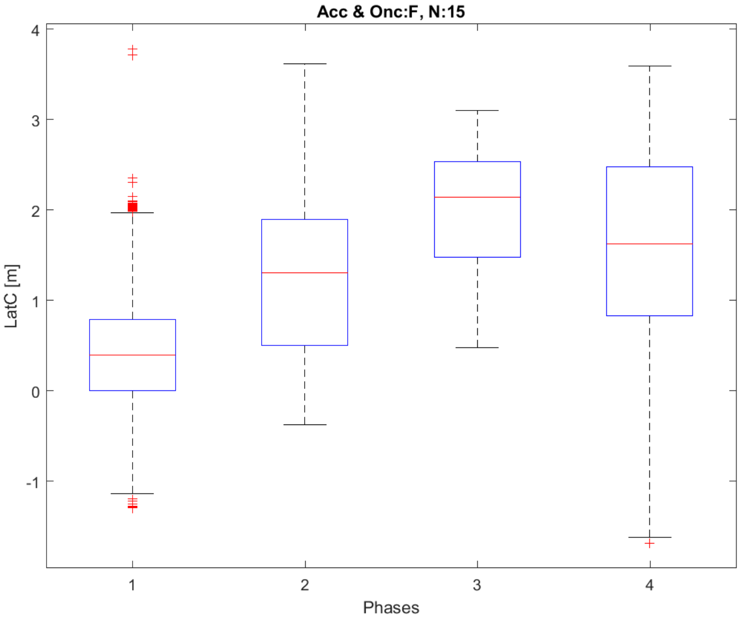


Figure C.5: *Boxplot of lateral clearance for accelerative overtaking strategy and oncoming traffic not present.*

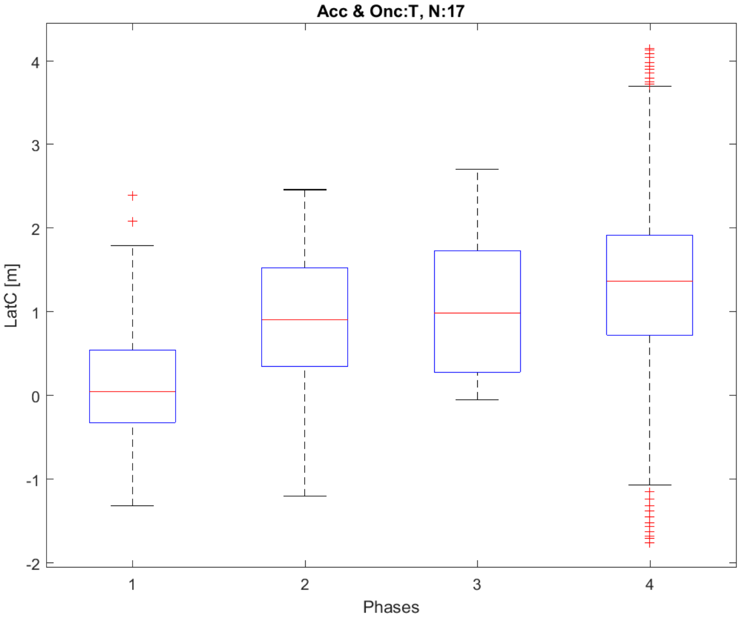


Figure C.6: *Boxplot of lateral clearance for accelerative overtaking strategy and oncoming traffic present.*

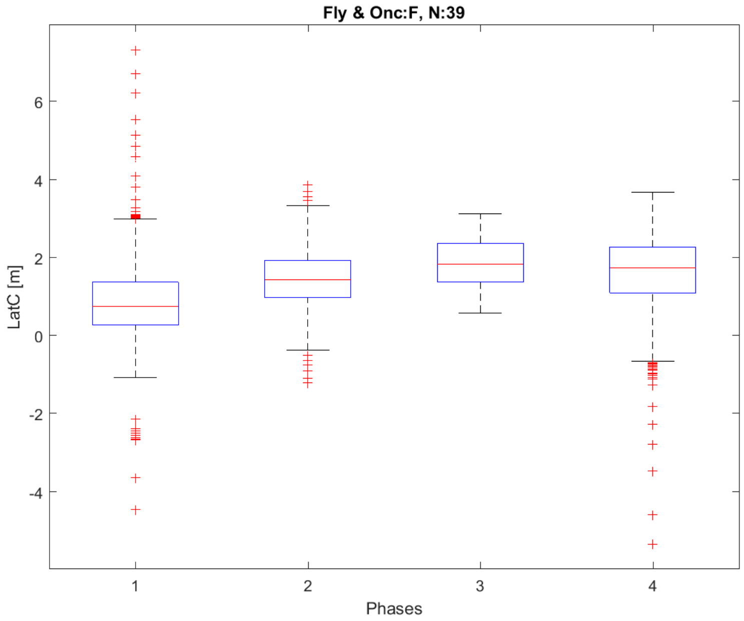


Figure C.7: *Boxplot of lateral clearance for flying overtaking strategy and oncoming traffic not present.*

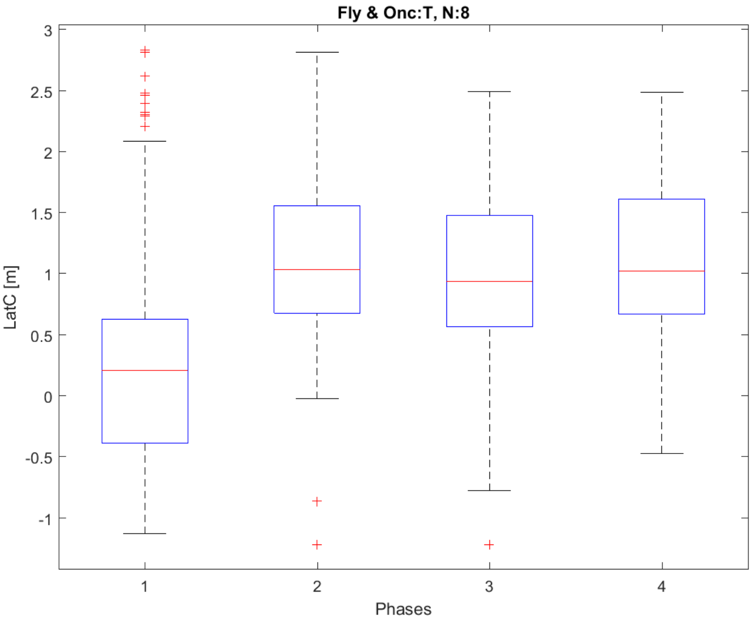


Figure C.8: *Boxplot of lateral clearance for flying overtaking strategy and oncoming traffic present.*

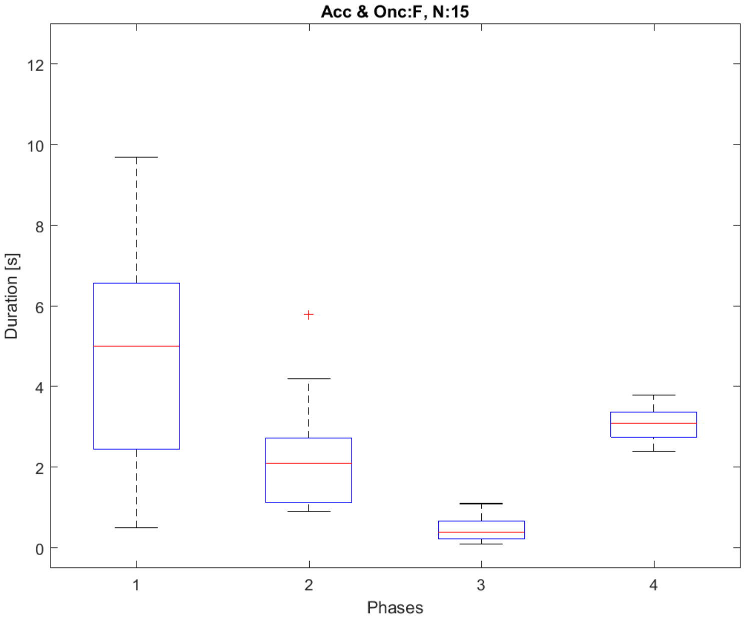


Figure C.9: *Boxplot of phase durations for accelerative overtaking strategy and oncoming traffic not present.*

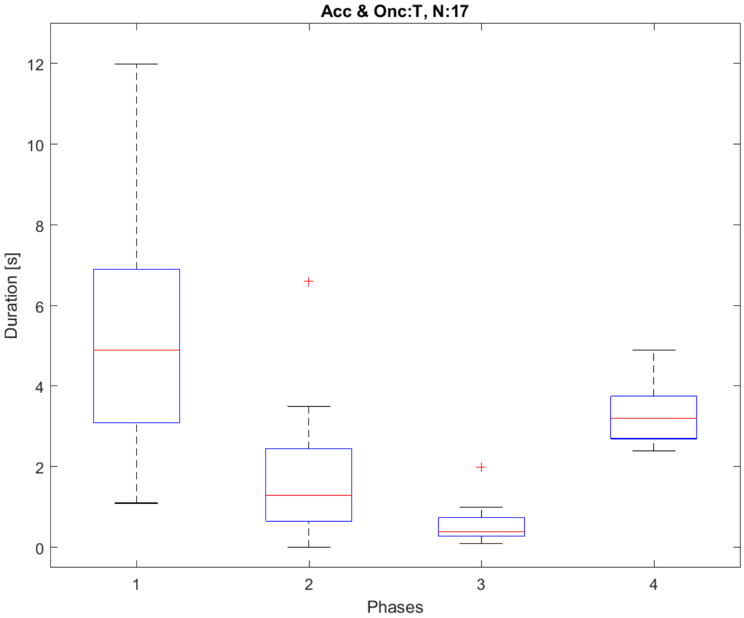


Figure C.10: *Boxplot of phase durations for accelerative overtaking strategy and oncoming traffic present.*

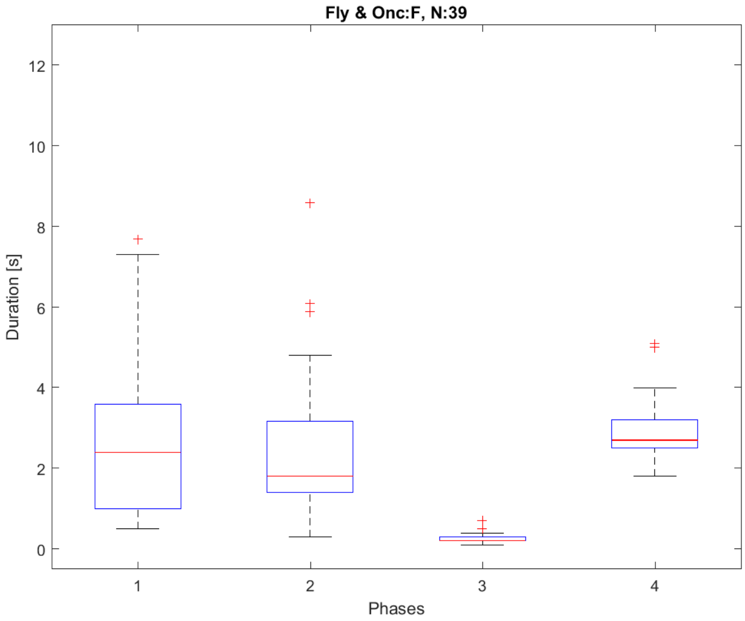


Figure C.11: *Boxplot of phase durations for flying overtaking strategy and oncoming traffic not present.*

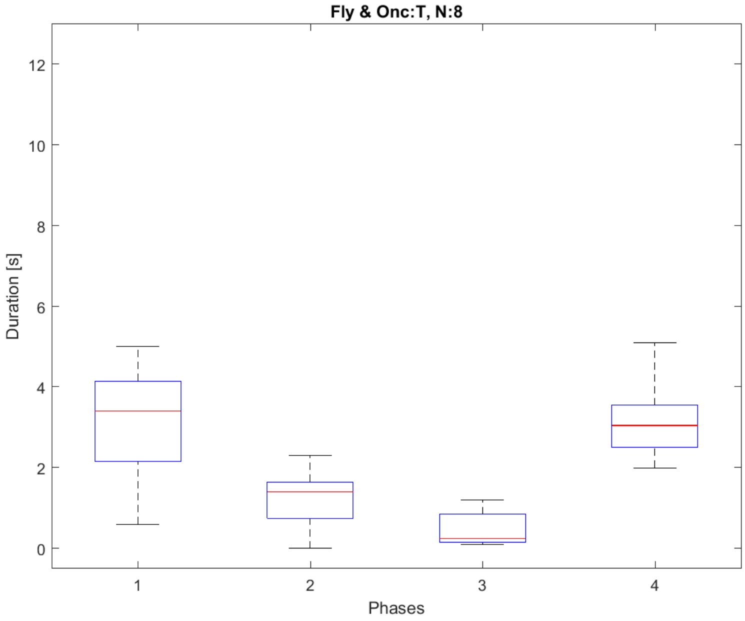


Figure C.12: *Boxplot of phase durations for flying overtaking strategy and oncoming traffic present.*

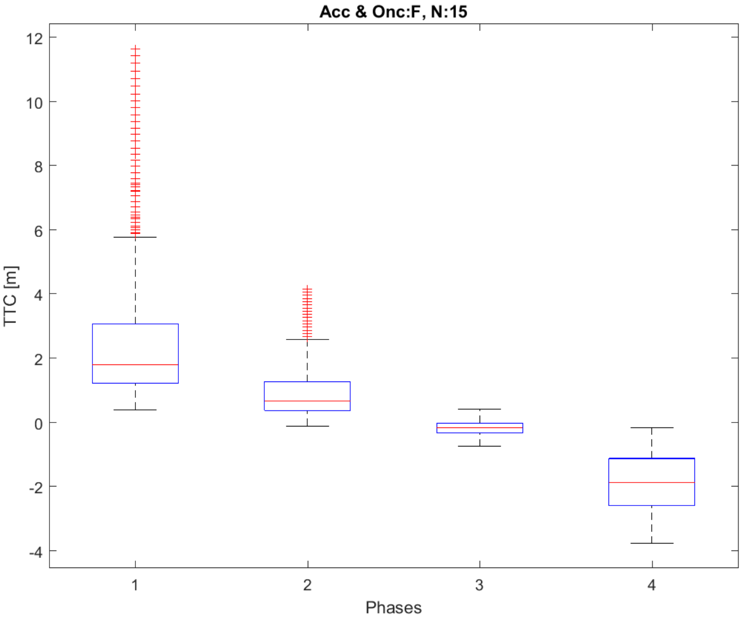


Figure C.13: *Boxplot of time to collision for accelerative overtaking strategy and oncoming traffic not present.*

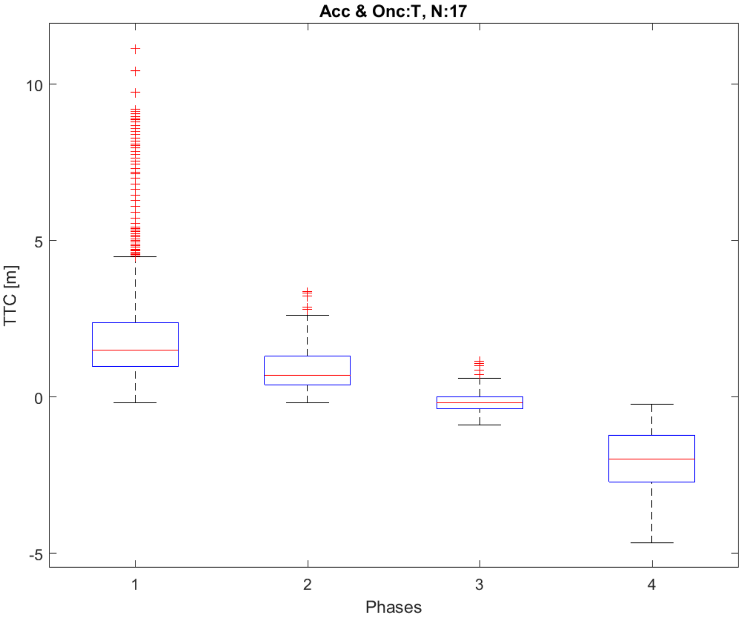


Figure C.14: *Boxplot of time to collision for accelerative overtaking strategy and oncoming traffic present.*

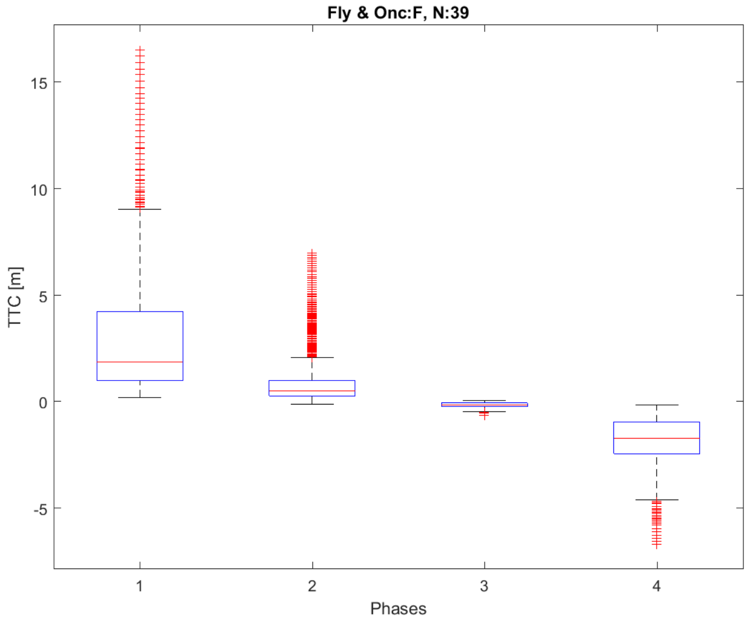


Figure C.15: *Boxplot of time to collision for flying overtaking strategy and oncoming traffic not present.*

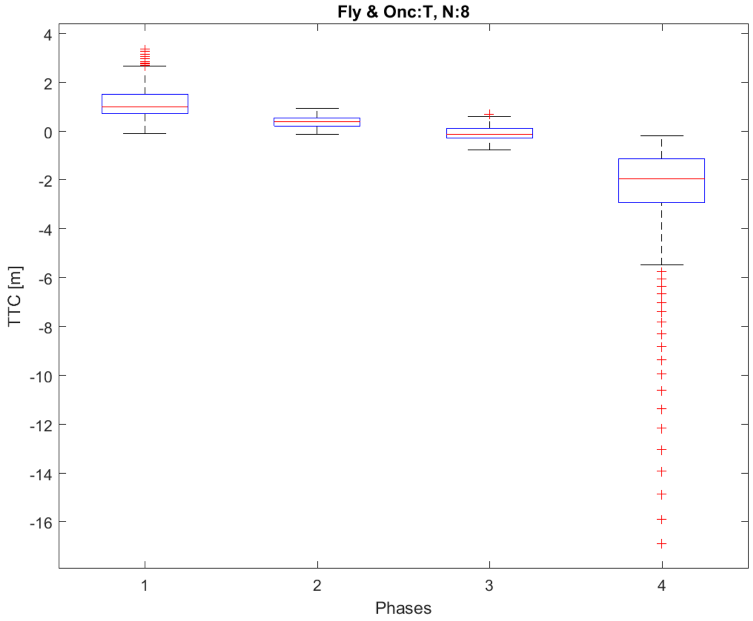


Figure C.16: *Boxplot of time to collision for flying overtaking strategy and oncoming traffic present.*

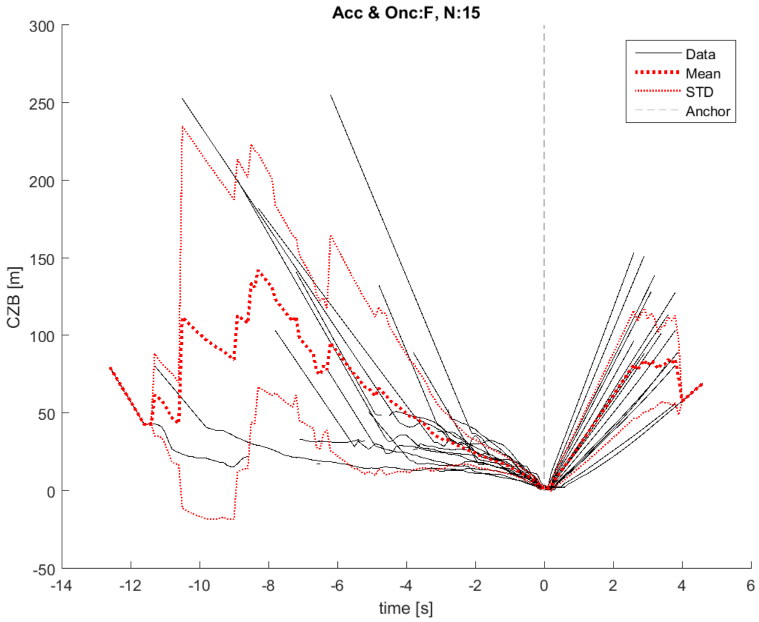


Figure C.17: *Minimum distance (CZB) for accelerative overtaking strategy and oncoming traffic not present.*

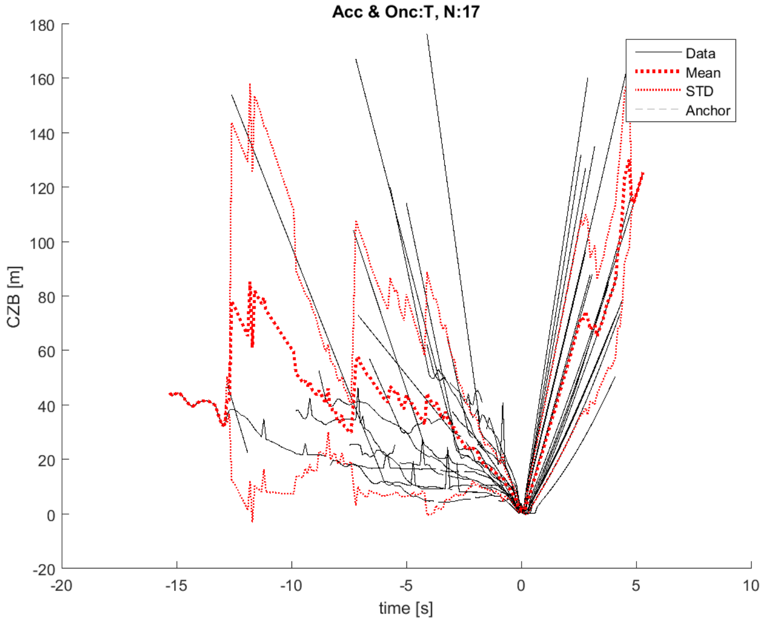


Figure C.18: *Minimum distance (CZB) for accelerative overtaking strategy and oncoming traffic present.*

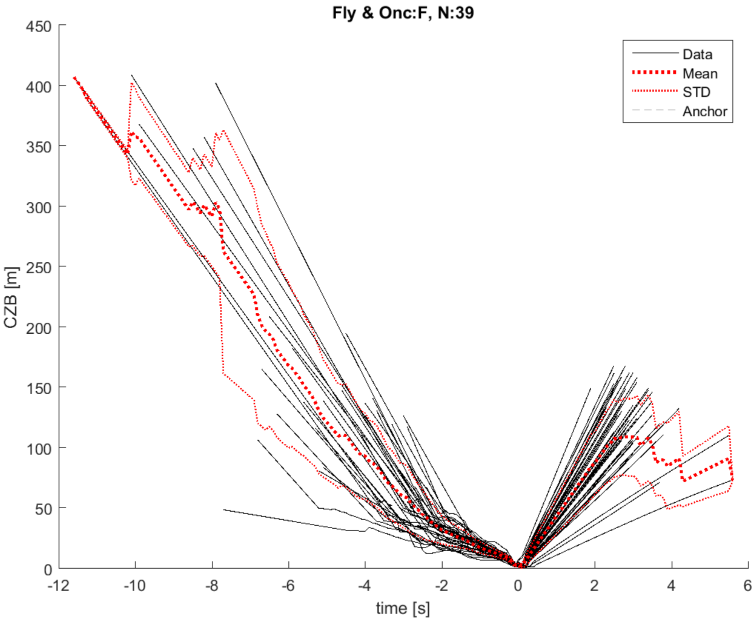


Figure C.19: *Minimum distance (CZB) for flying overtaking strategy and oncoming traffic not present.*

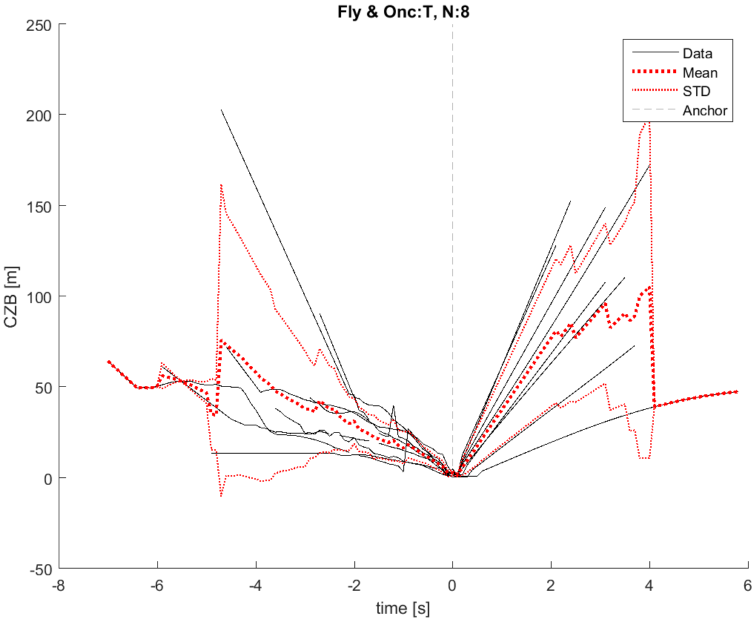


Figure C.20: *Minimum distance (CZB) for flying overtaking strategy and oncoming traffic present.*

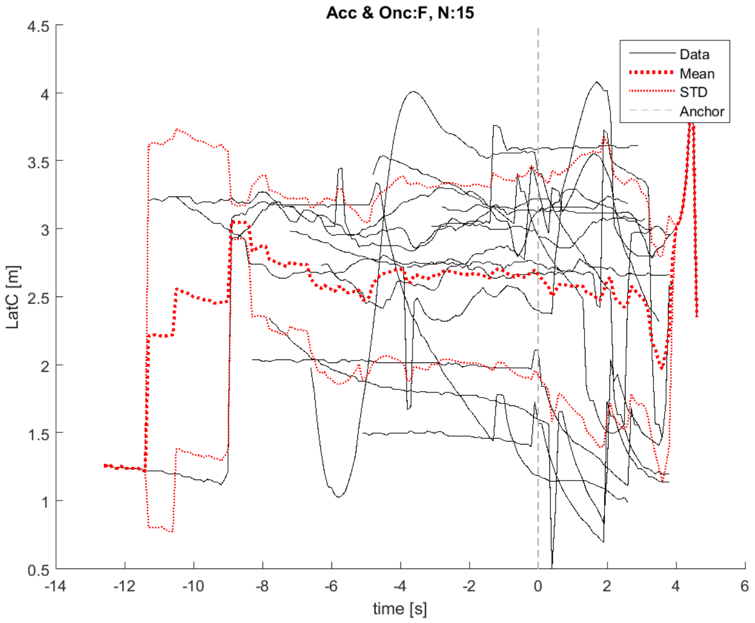


Figure C.21: *Lane width for accelerative overtaking strategy and oncoming traffic not present.*

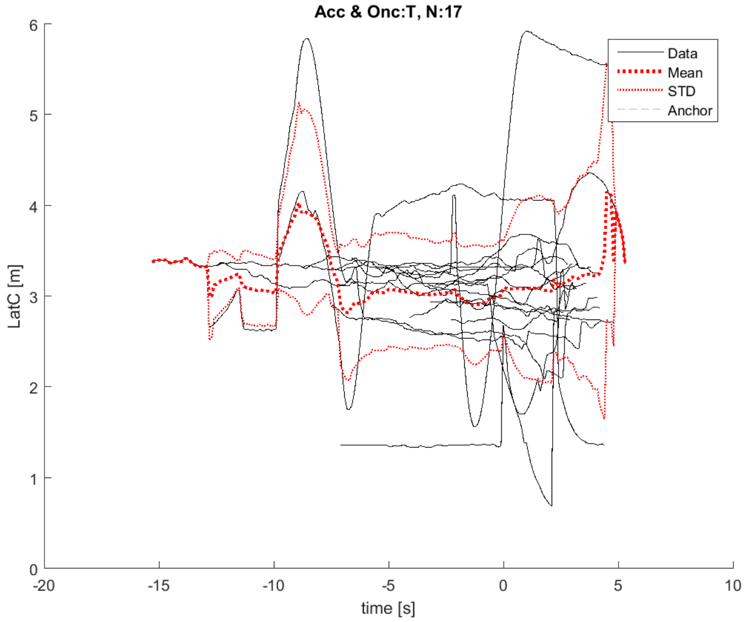


Figure C.22: *Lane width for accelerative overtaking strategy and oncoming traffic present.*

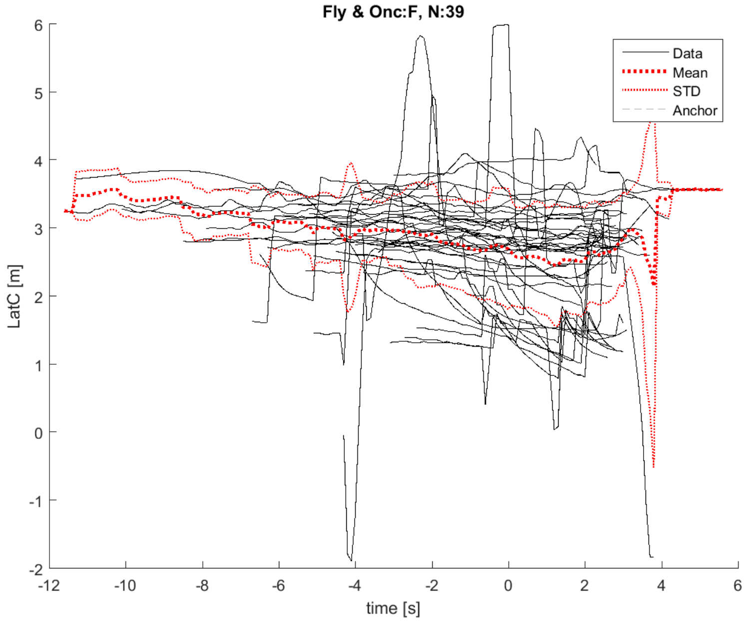


Figure C.23: *Lane width for flying overtaking strategy and oncoming traffic not present.*

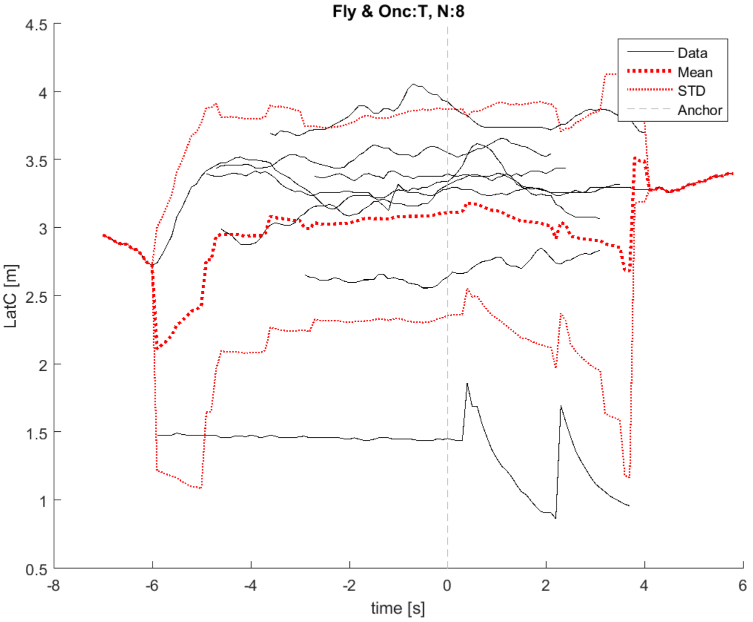


Figure C.24: *Lane width for flying overtaking strategy and oncoming traffic present.*

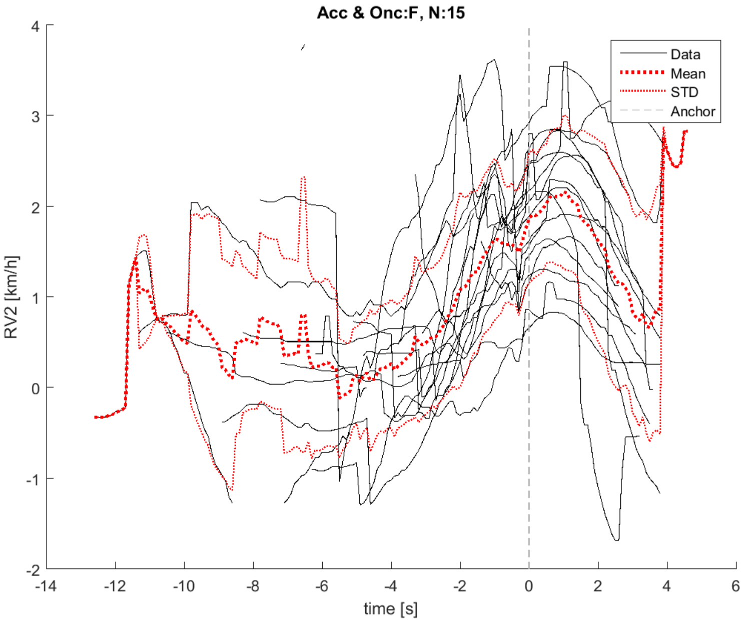


Figure C.25: *Lateral clearance for accelerative overtaking strategy and oncoming traffic not present.*

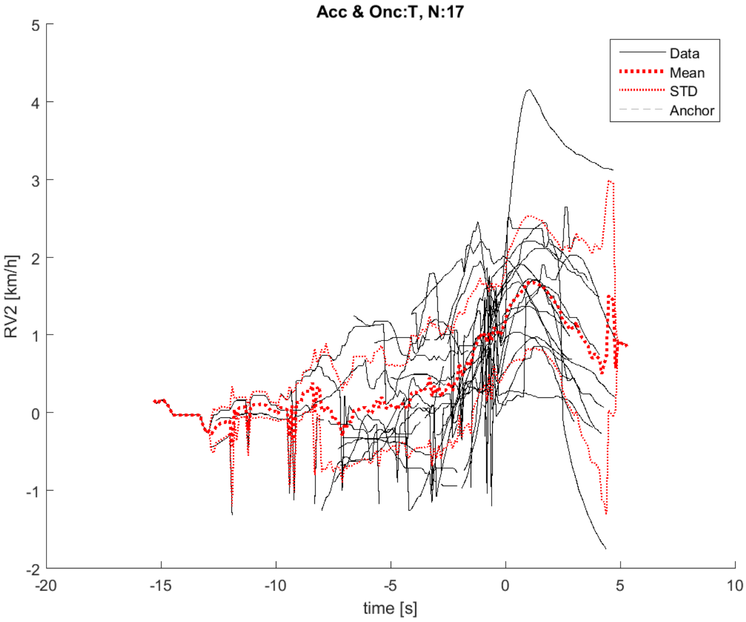


Figure C.26: *Lateral clearance for accelerative overtaking strategy and oncoming traffic present.*

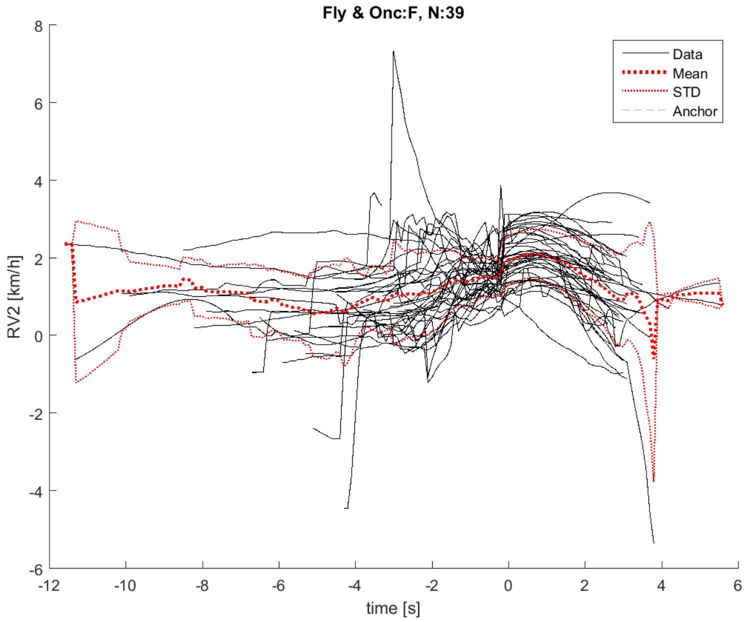


Figure C.27: *Lateral clearance for flying overtaking strategy and oncoming traffic not present.*

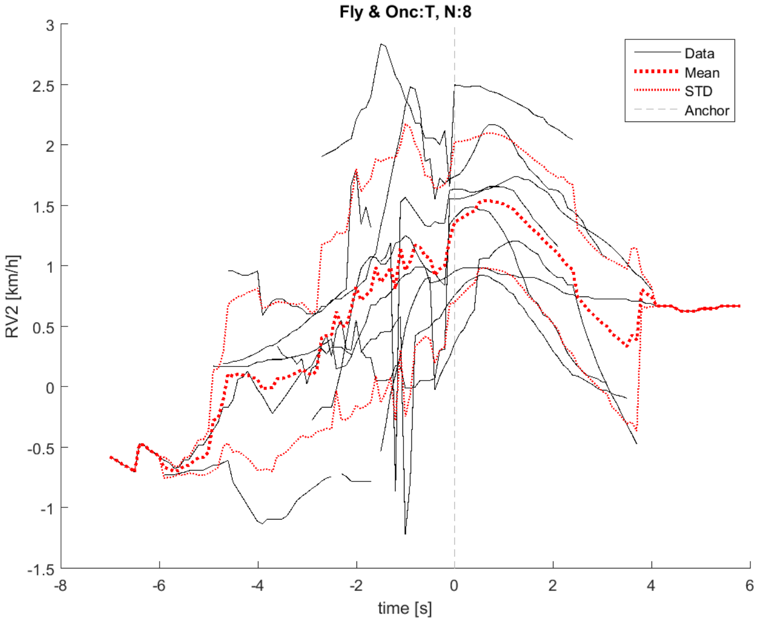


Figure C.28: *Lateral clearance for flying overtaking strategy and oncoming traffic present.*

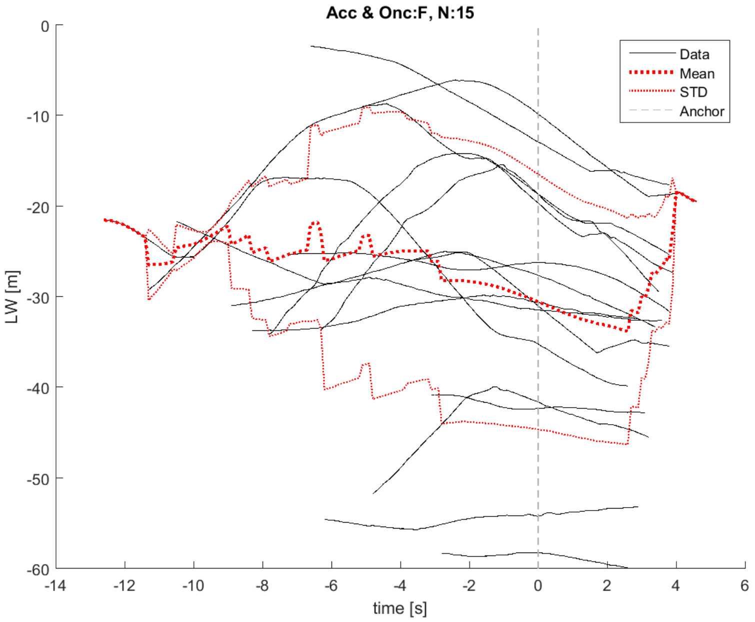


Figure C.29: *Relative velocity to VRU for accelerative overtaking strategy and oncoming traffic not present.*

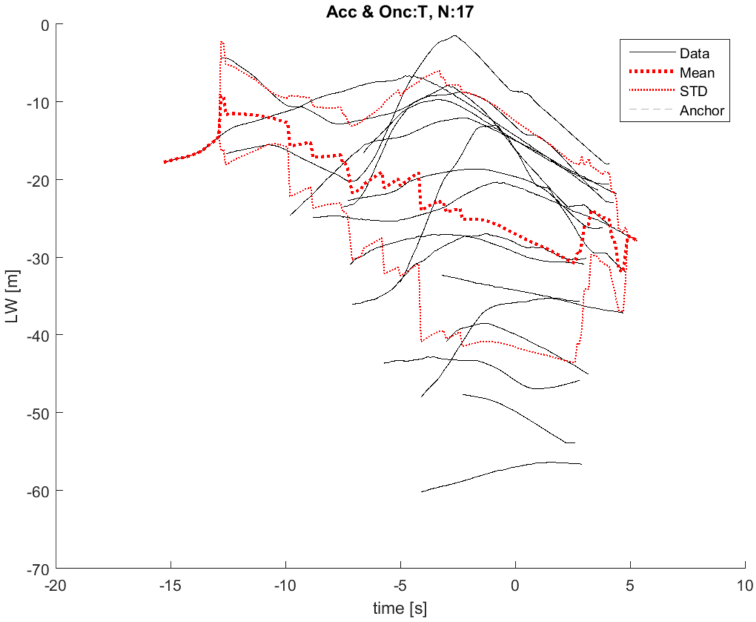


Figure C.30: *Relative velocity to VRU for accelerative overtaking strategy and oncoming traffic present.*

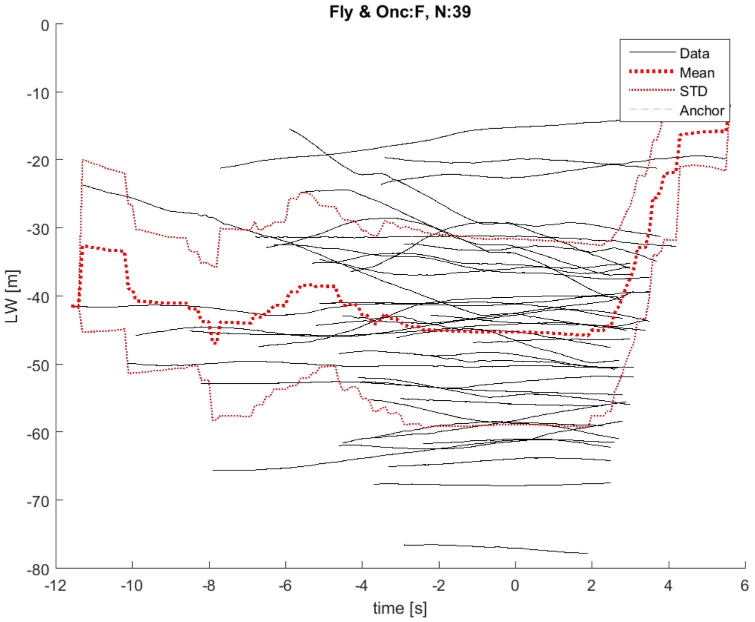


Figure C.31: *Relative velocity to VRU for flying overtaking strategy and oncoming traffic not present.*

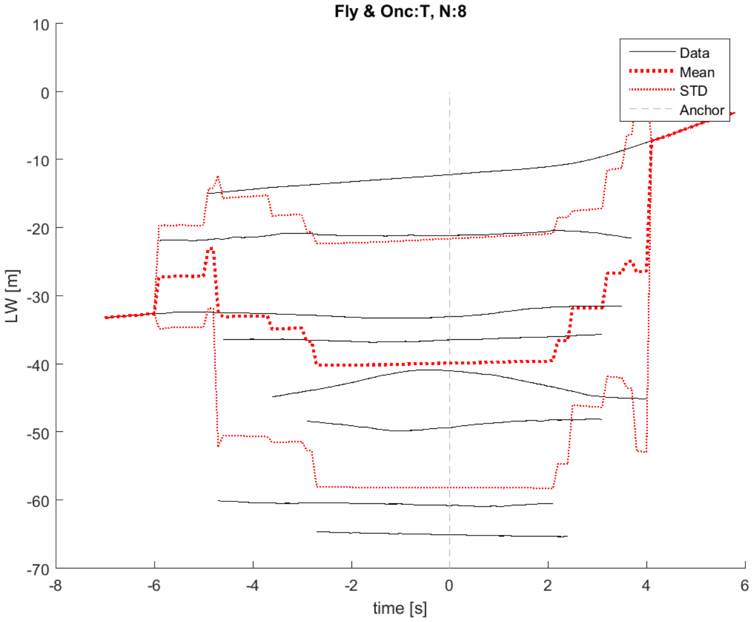


Figure C.32: *Relative velocity to VRU for flying overtaking strategy and oncoming traffic present.*

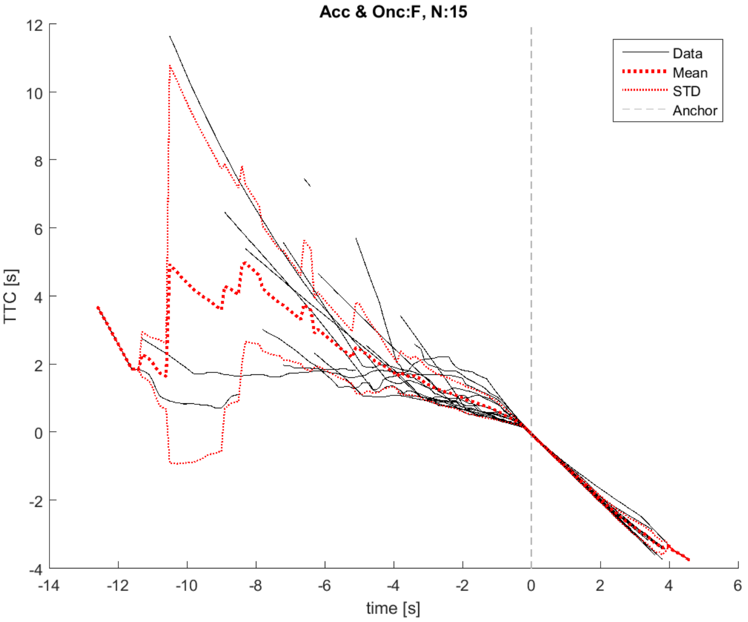


Figure C.33: *Time to collision for accelerative overtaking strategy and oncoming traffic not present.*

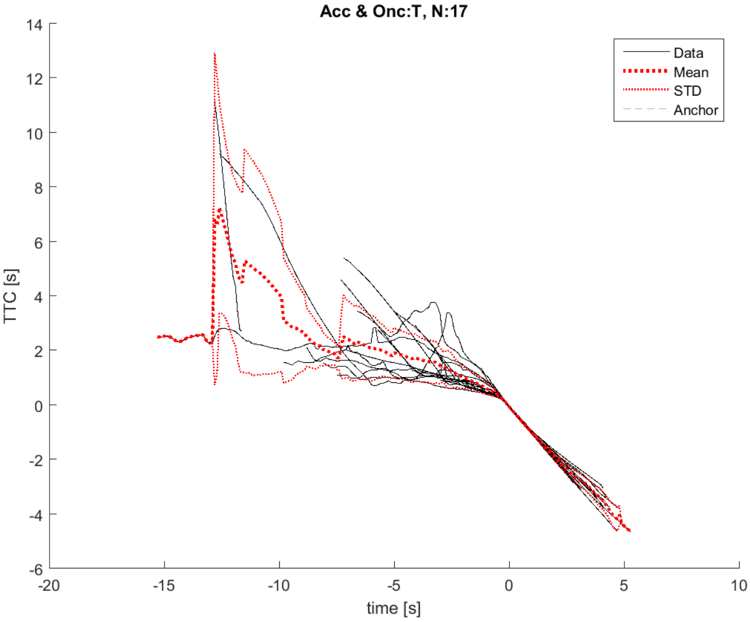


Figure C.34: *Time to collision for accelerative overtaking strategy and oncoming traffic present.*

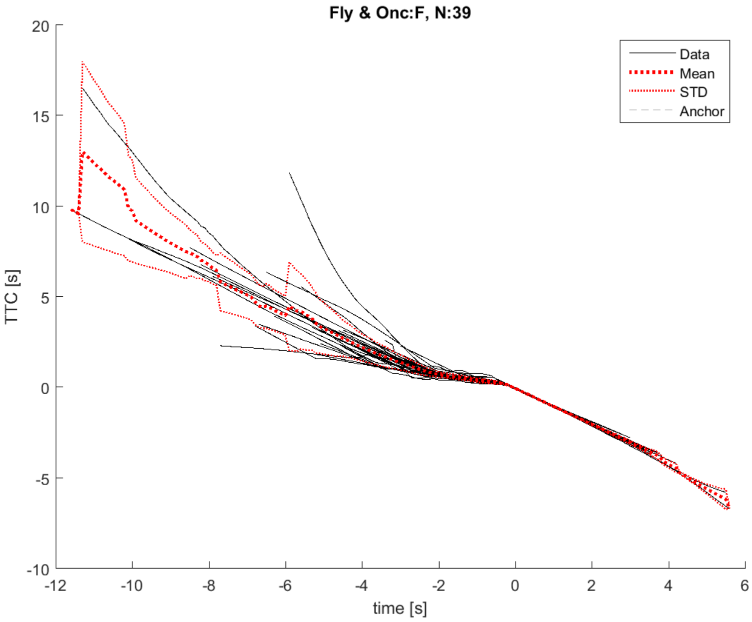


Figure C.35: *Time to collision for flying overtaking strategy and oncoming traffic not present.*

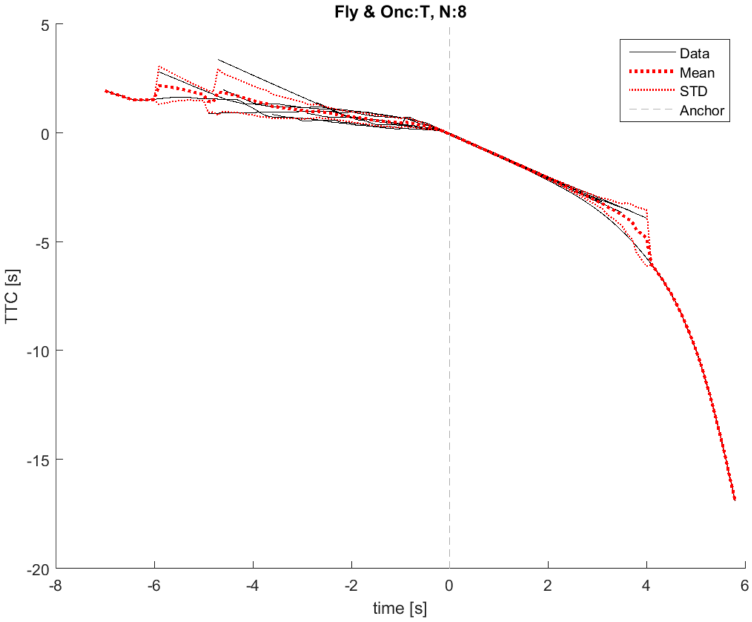


Figure C.36: *Time to collision for flying overtaking strategy and oncoming traffic present.*

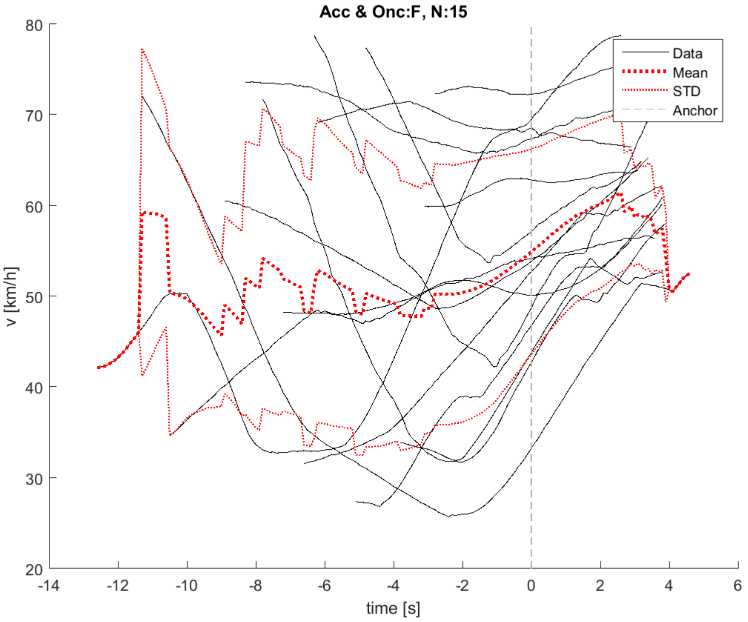


Figure C.37: *Ego vehicle speed for accelerative overtaking strategy and oncoming traffic not present.*

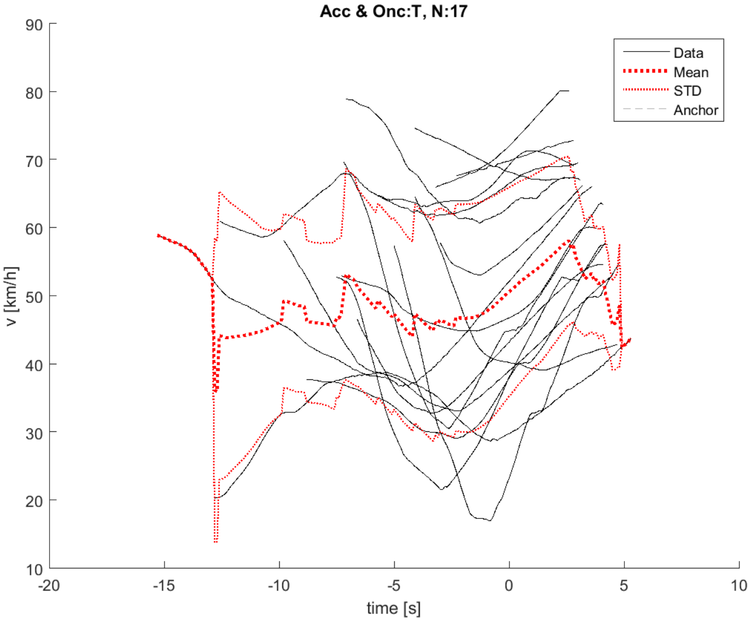


Figure C.38: *Ego vehicle speed for accelerative overtaking strategy and oncoming traffic present.*

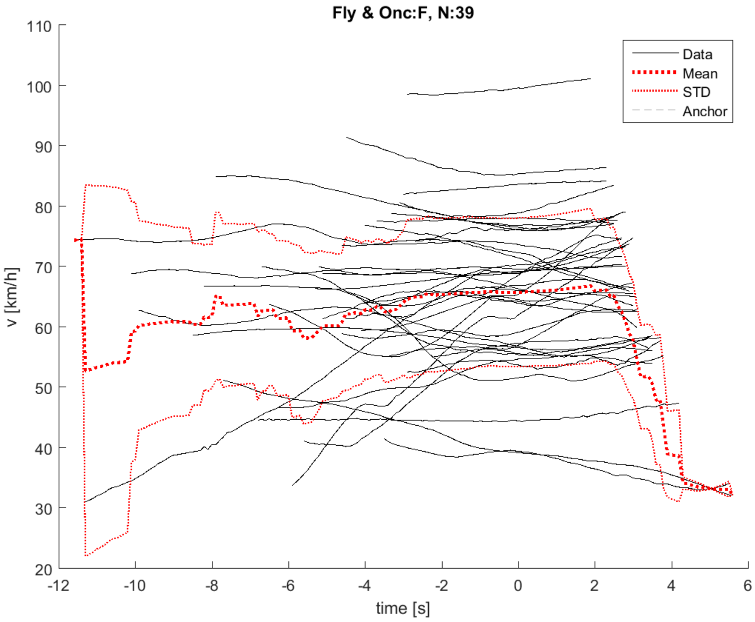


Figure C.39: *Ego vehicle speed for flying overtaking strategy and oncoming traffic not present.*

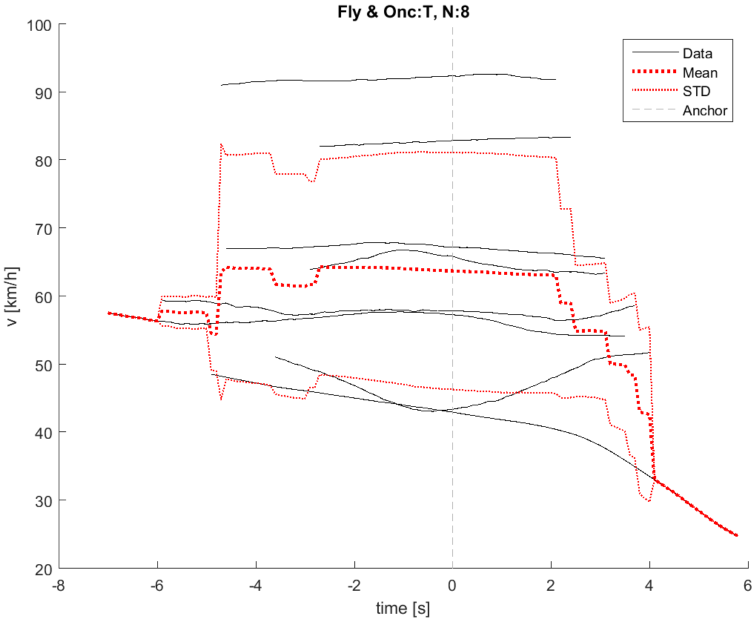


Figure C.40: *Ego vehicle speed for flying overtaking strategy and oncoming traffic present.*

D Data Sorted by Presence of Leading Vehicle

In the following tables and figures 'Lead. Present' implies that a leading vehicle was present and 'Lead. Not Present' implies that a leading vehicle was not present during the execution of the overtaking manoeuvre.

D.1 Tables

Table D.1: CZB [m]. Data sorted by overtaking manoeuvre strategy and Presence of Leading Vehicle.

			Phase 1	Phase 2	Phase 3	Phase 4
Accelerative	Lead Vehicle Present	Mean	55.12	34.00	3.03	-56.66
		STD	±46.93	±45.93	±1.98	±38.00
	Lead Vehicle Not Present	Mean	48.68	14.27	2.69	52.30
		STD	±45.81	±9.81	±1.90	±33.16
Flying	Lead Vehicle Present	Mean	130.67	51.97	3.09	62.46
		STD	±120.09	±64.64	±2.04	±38.53
	Lead Vehicle Not Present	Mean	88.01	33.34	4.44	68.96
		STD	±82.11	±45.54	±4.57	±40.34

Table D.2: p-values for CZB. Phase 1. Data Sorted by Presence of Leading Vehicle.

		Accelerative		Flying	
		Lead. Present	Lead. Not Present	Lead. Present	Lead. Not Present
Accelerative	Lead. Present	-	0.04	**p<0.001**	**p<0.001**
	Lead. Not Present	-	-	**p<0.001**	**p<0.001**
Flying	Lead. Present	-	-	-	**p<0.001**
	Lead. Not Present	-	-	-	-

Table D.3: p-values for CZB. Phase 2. Data Sorted by Presence of Leading Vehicle.

		Accelerative		Flying	
		Lead. Present	Lead. Not Present	Lead. Present	Lead. Not Present
Accelerative	Lead. Present	-	**p<0.001**	**p<0.001**	0.84
	Lead. Not Present	-	-	**p<0.001**	**p<0.001**
Flying	Lead. Present	-	-	-	**p<0.001**
	Lead. Not Present	-	-	-	-

Table D.4: p-values for CZB. Phase 3. Data Sorted by Presence of Leading Vehicle.

		Accelerative		Flying	
		Lead. Present	Lead. Not Present	Lead. Present	Lead. Not Present
Accelerative	Lead. Present	-	0.24	0.89	0.02
	Lead. Not Present	-	-	0.17	**p<0.001**
Flying	Lead. Present	-	-	-	0.02
	Lead. Not Present	-	-	-	-

Table D.5: p-values for CZB. Phase 4. Data Sorted by Presence of Leading Vehicle.

		Accelerative		Flying	
		Lead. Present	Lead. Not Present	Lead. Present	Lead. Not Present
Accelerative	Lead. Present	-	0.07	0.03	**p<0.001**
	Lead. Not Present	-	-	**p<0.001**	**p<0.001**
Flying	Lead. Present	-	-	-	**0.002**
	Lead. Not Present	-	-	-	-

Table D.6: TTC [s]. Data Sorted by Presence of Leading Vehicle.

			Phase 1	Phase 2	Phase 3	Phase 4
Accelerative	Lead Vehicle Present	Mean	2.61	1.21	-0.07	-1.96
		STD	±2.14	±0.90	±0.43	±0.99
	Lead Vehicle Not Present	Mean	2.10	0.65	-0.18	-1.94
		STD	±1.68	±0.46	±0.24	±0.92
Flying	Lead Vehicle Present	Mean	3.04	1.23	-0.18	-2.30
		STD	±2.45	±1.44	±0.20	±2.03
	Lead Vehicle Not Present	Mean	2.38	0.70	-0.11	-1.69
		STD	±2.70	±0.93	±0.21	±0.91

Table D.7: p-values for TTC. Phase 1. Data Sorted by Presence of Leading Vehicle.

		Accelerative		Flying	
		Lead. Present	Lead. Not Present	Lead. Present	Lead. Not Present
Accelerative	Lead. Present	-	**p<0.001**	0.01	0.18
	Lead. Not Present	-	-	**p<0.001**	**0.004**
Flying	Lead. Present	-	-	-	**p<0.001**
	Lead. Not Present	-	-	-	-

Table D.8: p-values for TTC. Phase 2. Data Sorted by Presence of Leading Vehicle.

		Accelerative		Flying	
		Lead. Present	Lead. Not Present	Lead. Present	Lead. Not Present
Accelerative	Lead. Present	-	**p<0.001**	0.89	**p<0.001**
	Lead. Not Present	-	-	**p<0.001**	0.35
Flying	Lead. Present	-	-	-	**p<0.001**
	Lead. Not Present	-	-	-	-

Table D.9: p-values for TTC. Phase 3. Data Sorted by Presence of Leading Vehicle.

		Accelerative		Flying	
		Lead. Present	Lead. Not Present	Lead. Present	Lead. Not Present
Accelerative	Lead. Present	-	0.02	0.69	0.49
	Lead. Not Present	-	-	0.81	**0.005**
Flying	Lead. Present	-	-	-	0.02
	Lead. Not Present	-	-	-	-

Table D.10: p-values for TTC. Phase 4. Data Sorted by Presence of Leading Vehicle.

		Accelerative		Flying	
		Lead. Present	Lead. Not Present	Lead. Present	Lead. Not Present
Accelerative	Lead. Present	-	0.69	0.007	**p<0.001**
	Lead. Not Present	-	-	**p<0.001**	**p<0.001**
Flying	Lead. Present	-	-	-	**p<0.001**
	Lead. Not Present	-	-	-	-

Table D.11: LatC [m]. Data Sorted by Presence of Leading Vehicle.

			Phase 1	Phase 2	Phase 3	Phase 4
Accelerative	Lead Vehicle Present	Mean	0.74	1.25	1.53	1.77
		STD	±0.64	±0.94	±1.00	±0.71
	Lead Vehicle Not Present	Mean	0.15	1.02	1.39	1.32
		STD	±0.66	±0.69	±0.79	±1.14
Flying	Lead Vehicle Present	Mean	1.06	1.50	1.73	1.70
		STD	±0.69	±0.69	±0.67	±0.91
	Lead Vehicle Not Present	Mean	0.55	1.35	1.59	1.44
		STD	±1.05	±0.82	±0.82	±0.96

Table D.12: p-values for LatC. Phase 1. Data Sorted by Presence of Leading Vehicle.

		Accelerative		Flying	
		Lead. Present	Lead. Not Present	Lead. Present	Lead. Not Present
Accelerative	Lead. Present	-	**p<0.001**	**p<0.001**	**0.005**
	Lead. Not Present	-	-	**p<0.001**	**p<0.001**
Flying	Lead. Present	-	-	-	**p<0.001**
	Lead. Not Present	-	-	-	-

Table D.13: p-values for LatC. Phase 2. Data Sorted by Presence of Leading Vehicle.

		Accelerative		Flying	
		Lead. Present	Lead. Not Present	Lead. Present	Lead. Not Present
Accelerative	Lead. Present	-	**p<0.001**	**p<0.001**	0.09
	Lead. Not Present	-	-	**p<0.001**	**p<0.001**
Flying	Lead. Present	-	-	-	**0.002**
	Lead. Not Present	-	-	-	-

Table D.14: p-values for LatC. Phase 3. Data Sorted by Presence of Leading Vehicle.

		Accelerative		Flying	
		Lead. Present	Lead. Not Present	Lead. Present	Lead. Not Present
Accelerative	Lead. Present	-	0.29	0.18	0.68
	Lead. Not Present	-	-	**0.003**	0.06
Flying	Lead. Present	-	-	-	0.24
	Lead. Not Present	-	-	-	-

Table D.15: p-values for LatC. Phase 4. Data Sorted by Presence of Leading Vehicle.

		Accelerative		Flying	
		Lead. Present	Lead. Not Present	Lead. Present	Lead. Not Present
Accelerative	Lead. Present	-	**$p<0.001$**	0.26	**$p<0.001$**
	Lead. Not Present	-	-	**$p<0.001$**	0.02
Flying	Lead. Present	-	-	-	**$p<0.001$**
	Lead. Not Present	-	-	-	-

D.2 Figures

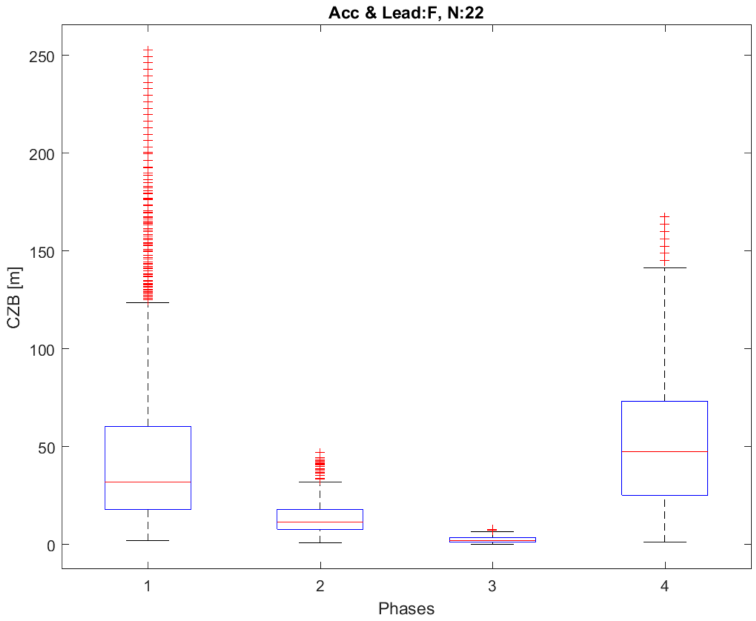


Figure D.1: *Boxplot of minimum distance (CZB) for accelerative overtaking strategy and leading vehicle not present.*

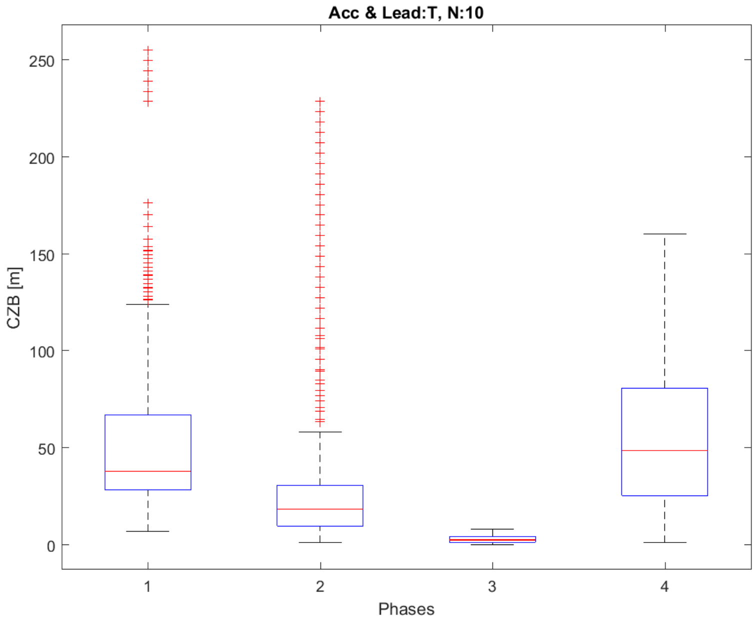


Figure D.2: *Boxplot of minimum distance (CZB) for accelerative overtaking strategy and leading vehicle present.*

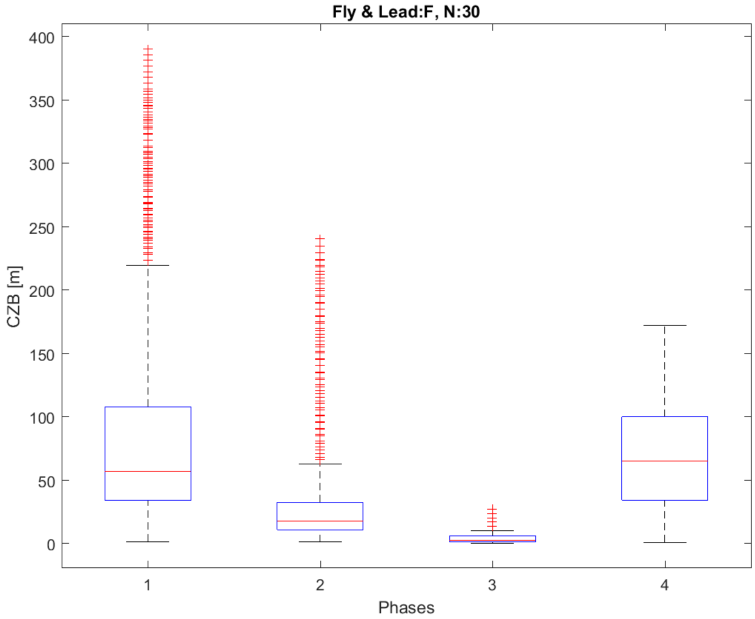


Figure D.3: *Boxplot of minimum distance (CZB) for flying overtaking strategy and leading vehicle not present.*

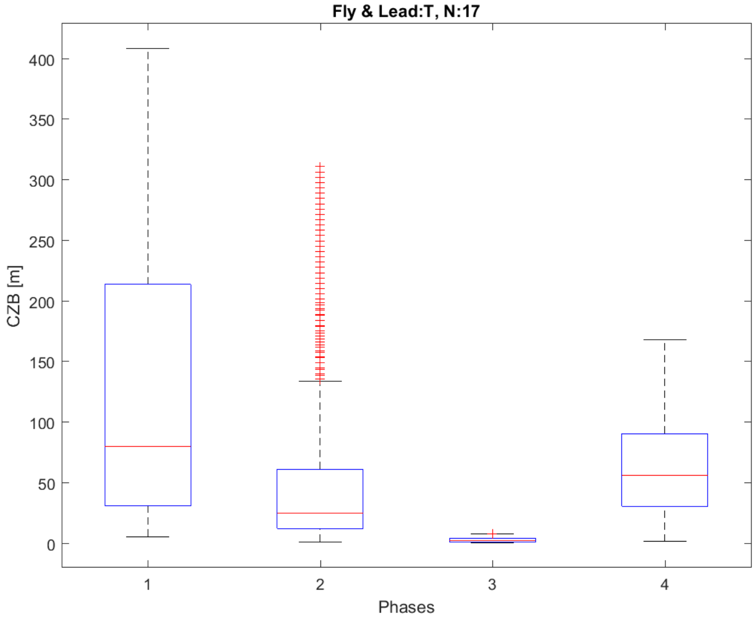


Figure D.4: *Boxplot of minimum distance (CZB) for flying overtaking strategy and leading vehicle present.*

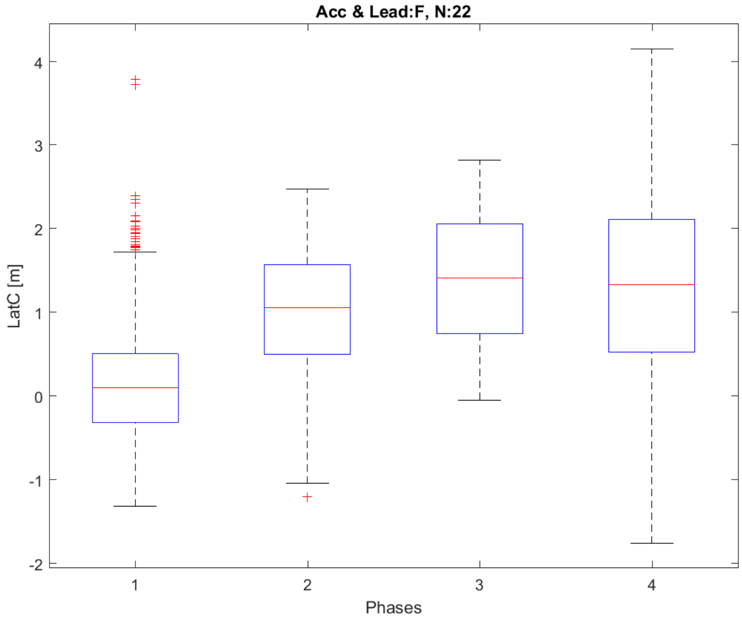


Figure D.5: *Boxplot of lateral clearance for accelerative overtaking strategy and leading vehicle not present.*

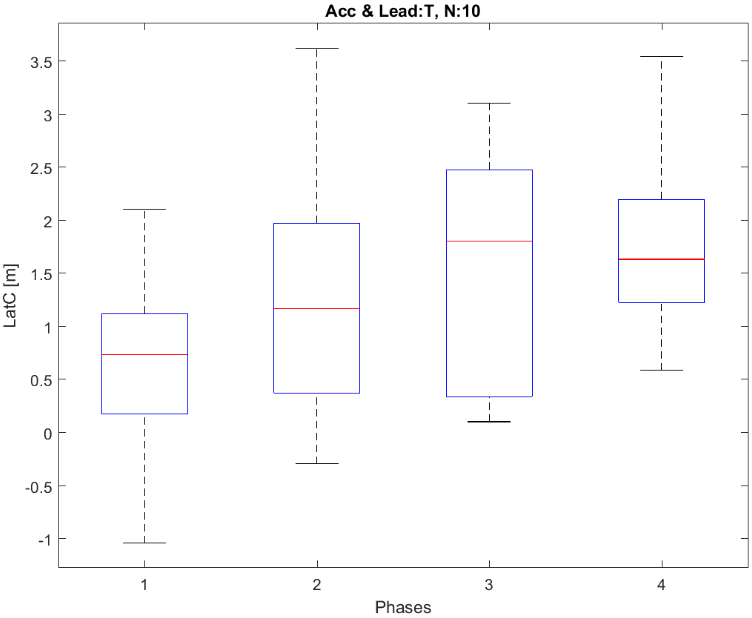


Figure D.6: *Boxplot of lateral clearance for accelerative overtaking strategy and leading vehicle present.*

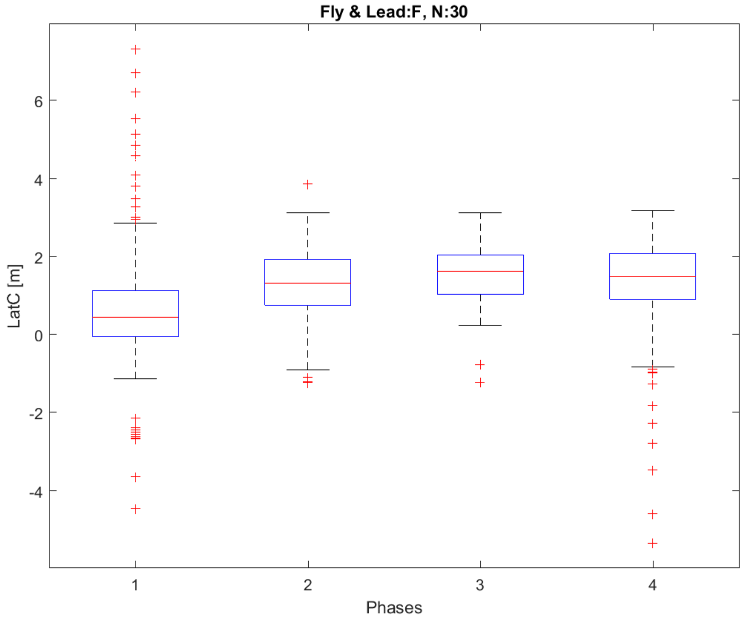


Figure D.7: *Boxplot of lateral clearance for flying overtaking strategy and leading vehicle not present.*

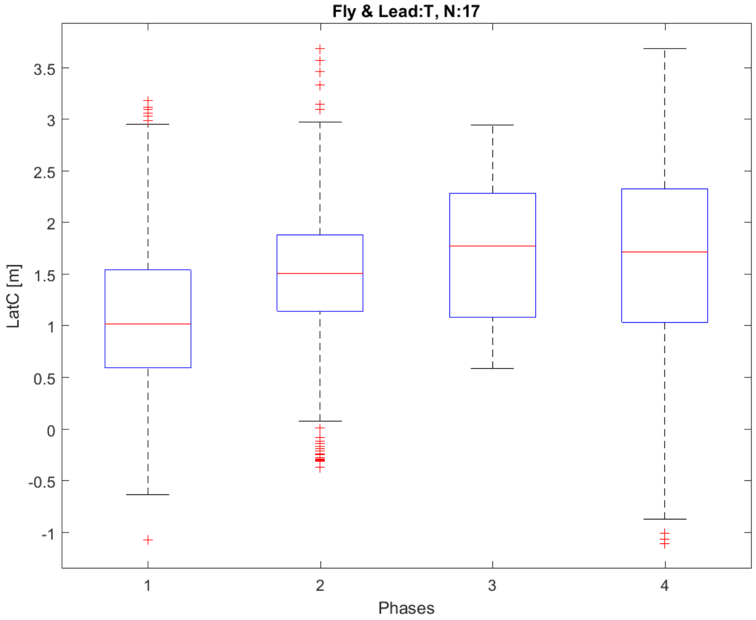


Figure D.8: *Boxplot of lateral clearance for flying overtaking strategy and leading vehicle present.*

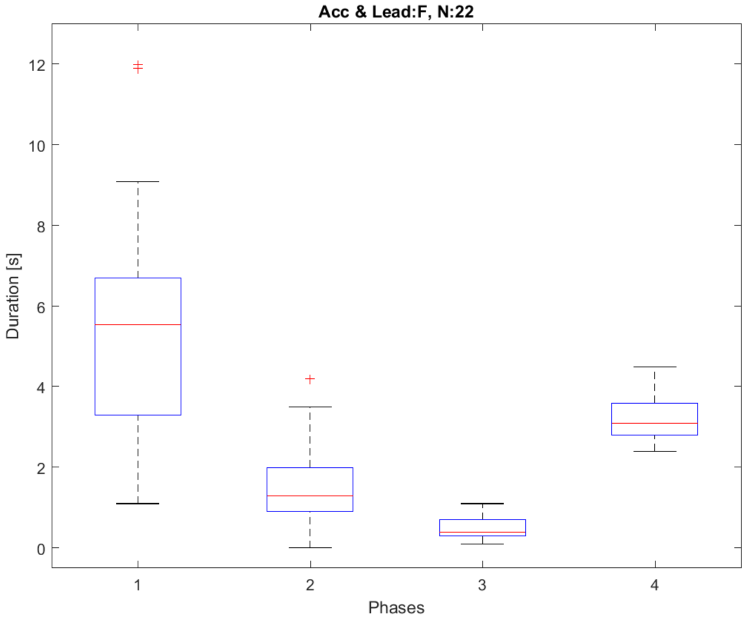


Figure D.9: *Boxplot of phase durations for accelerative overtaking strategy and leading vehicle not present.*

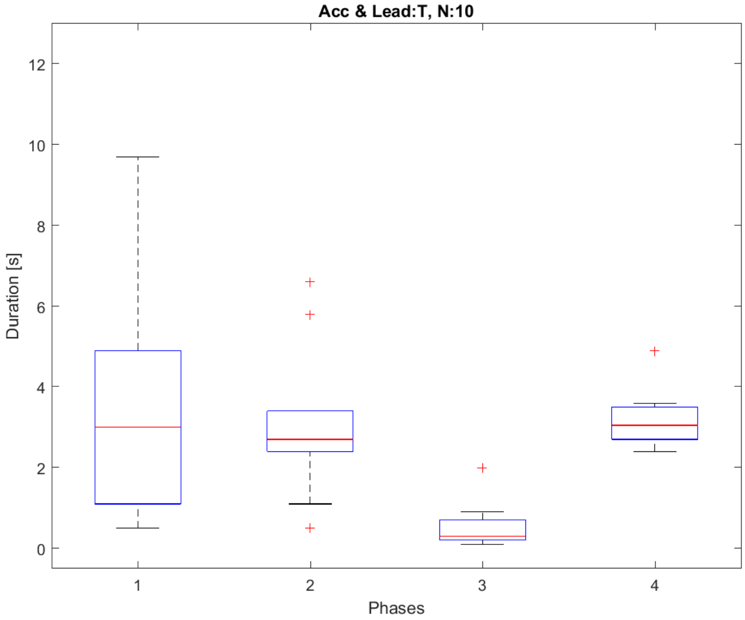


Figure D.10: *Boxplot of phase durations for accelerative overtaking strategy and leading vehicle present.*

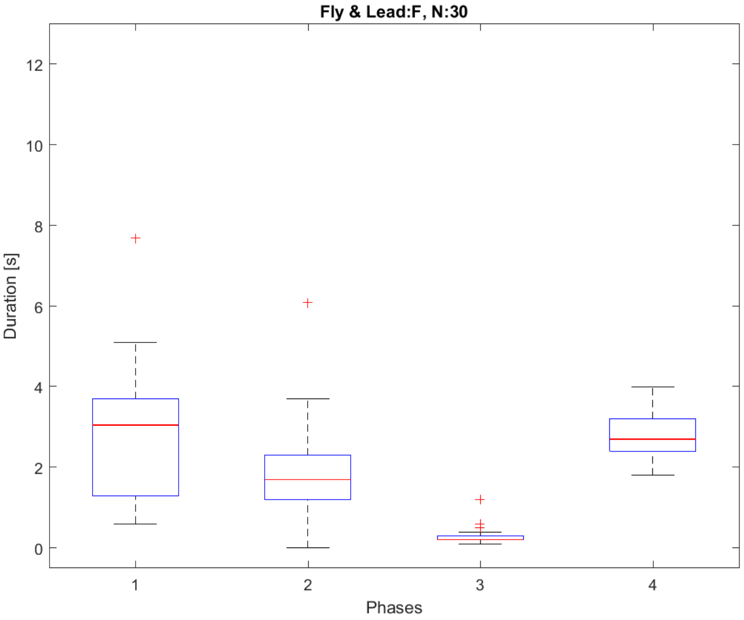


Figure D.11: *Boxplot of phase durations for flying overtaking strategy and leading vehicle not present.*

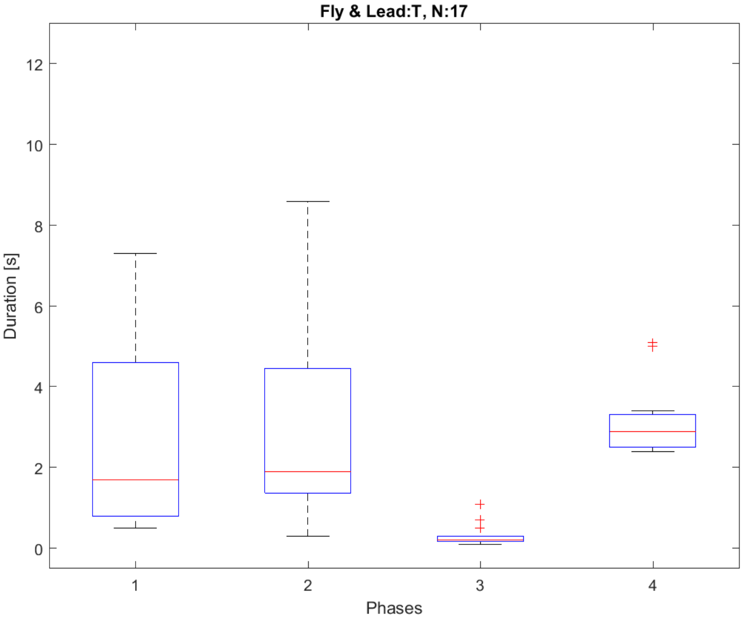


Figure D.12: *Boxplot of phase durations for flying overtaking strategy and leading vehicle present.*

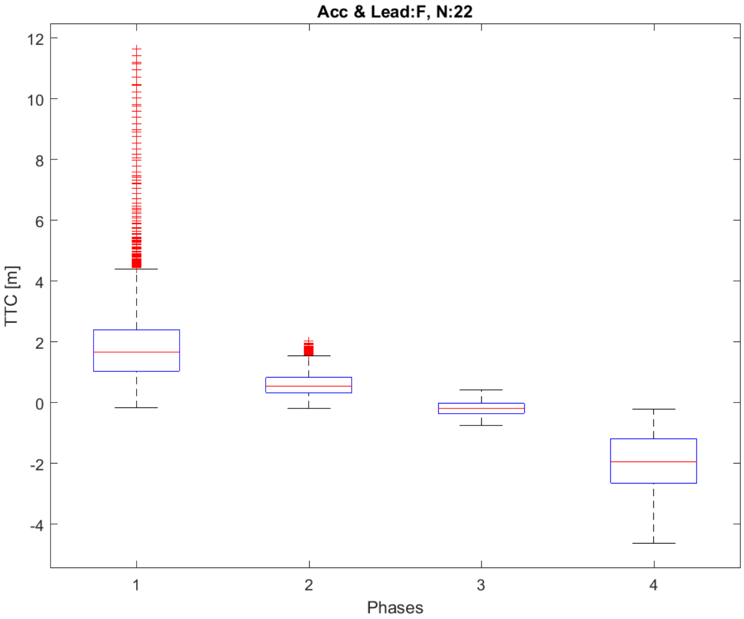


Figure D.13: *Boxplot of time to collision for accelerative overtaking strategy and leading vehicle not present.*

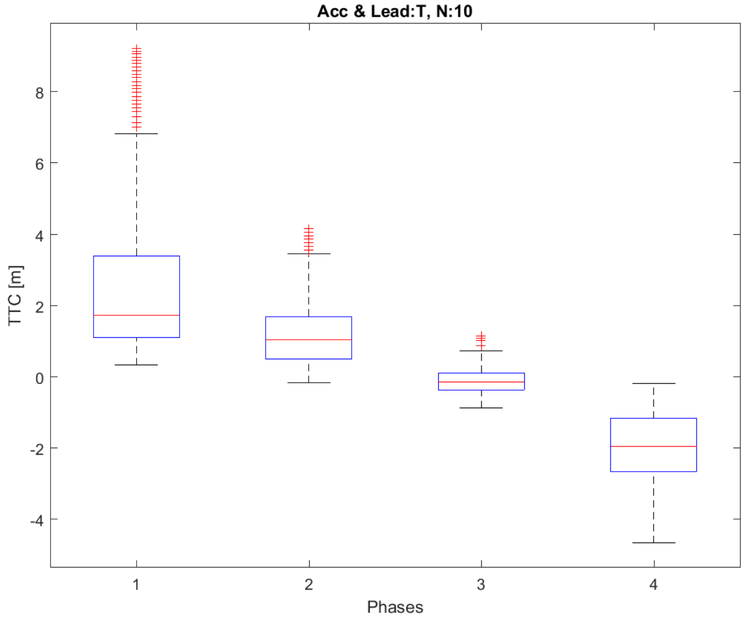


Figure D.14: *Boxplot of time to collision for accelerative overtaking strategy and leading vehicle present.*

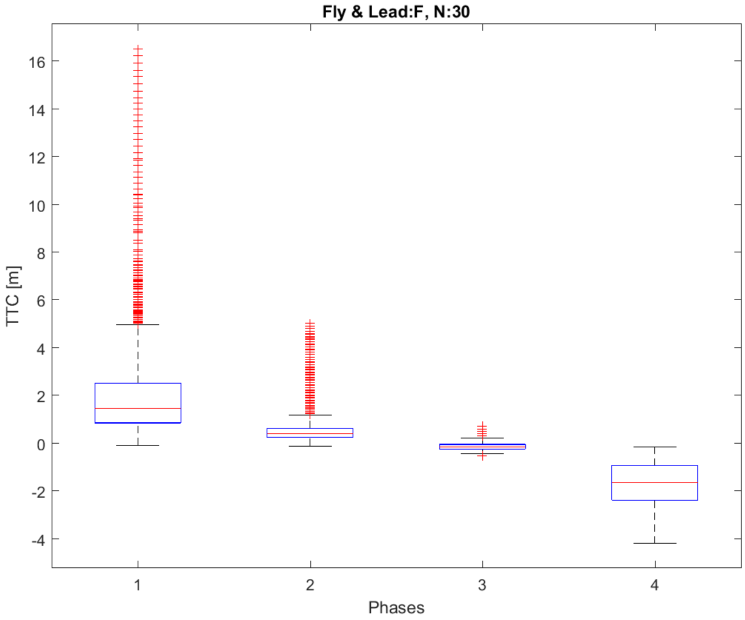


Figure D.15: *Boxplot of time to collision for flying overtaking strategy and leading vehicle not present.*

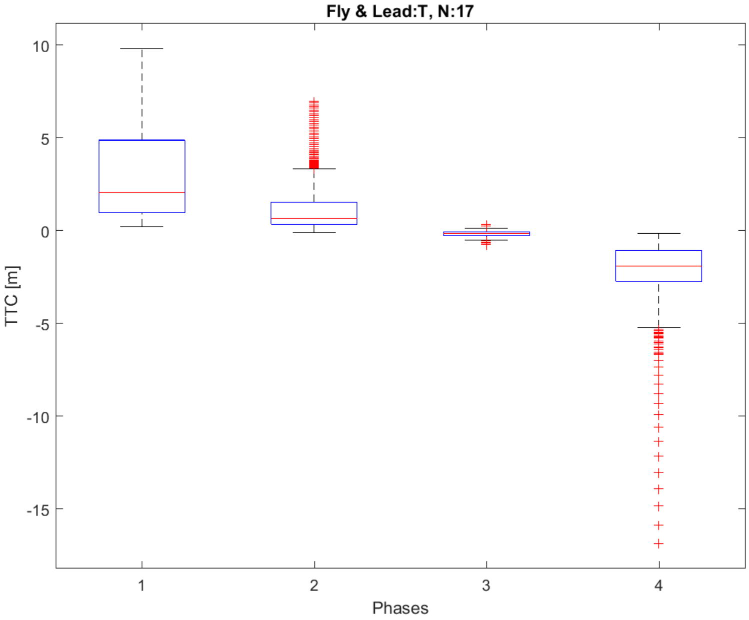


Figure D.16: *Boxplot of time to collision for flying overtaking strategy and leading vehicle present.*

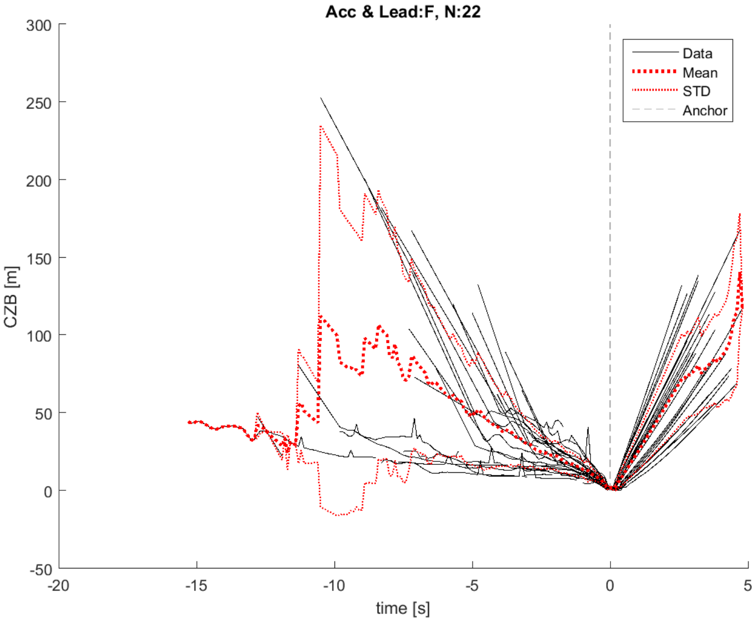


Figure D.17: *Minimum distance (CZB) for accelerative overtaking strategy and leading vehicle not present.*

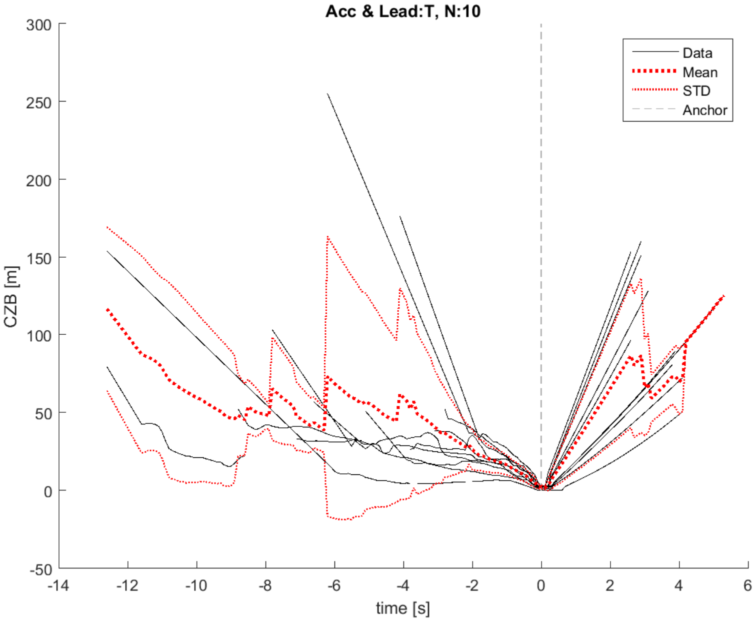


Figure D.18: *Minimum distance (CZB) for accelerative overtaking strategy and leading vehicle present.*

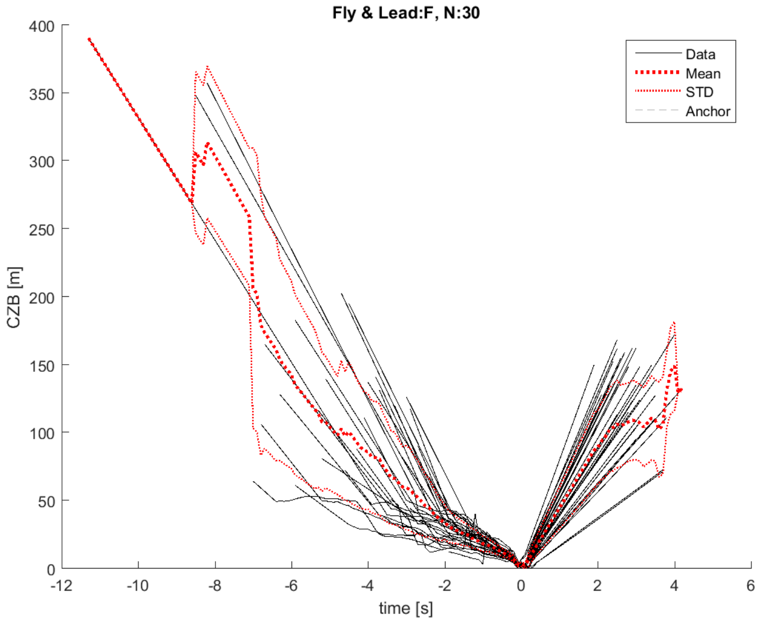


Figure D.19: *Minimum distance (CZB) for flying overtaking strategy and leading vehicle not present.*

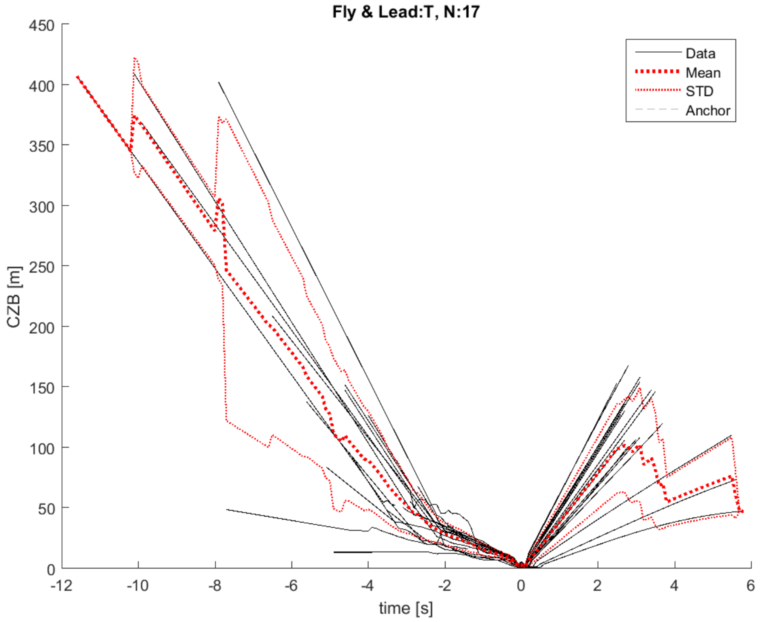


Figure D.20: *Minimum distance (CZB) for flying overtaking strategy and leading vehicle present.*

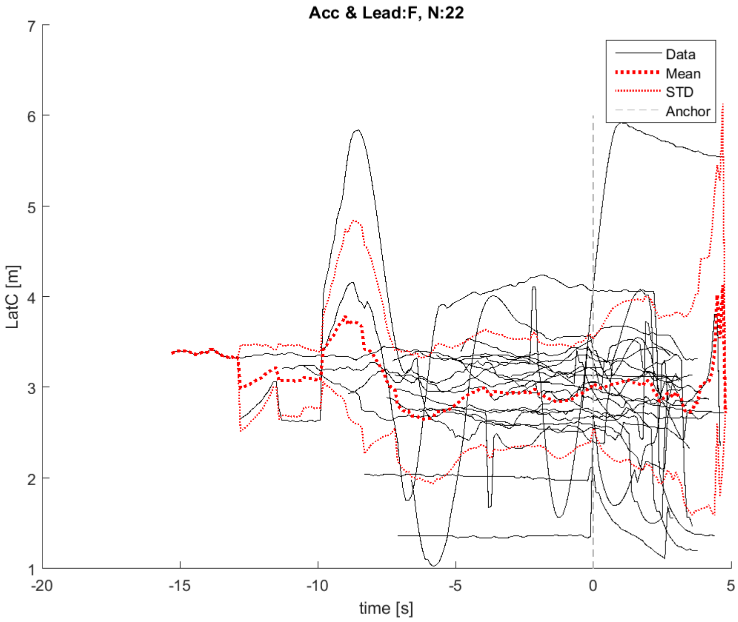


Figure D.21: *Lane width for accelerative overtaking strategy and leading vehicle not present.*

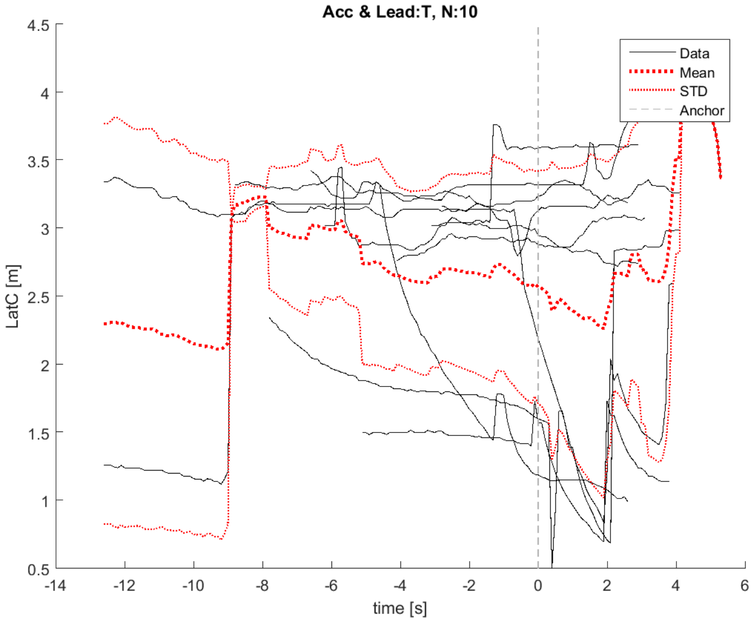


Figure D.22: *Lane width for accelerative overtaking strategy and leading vehicle present.*

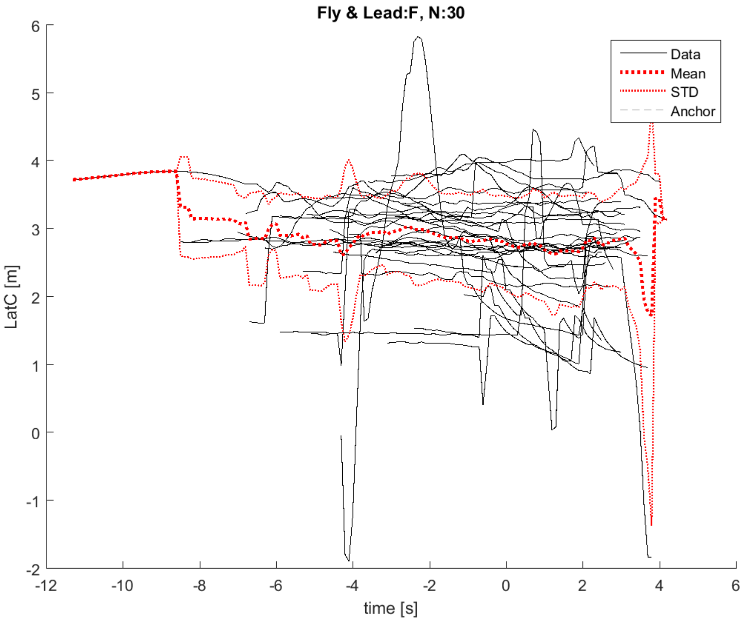


Figure D.23: *Lane width for flying overtaking strategy and leading vehicle not present.*

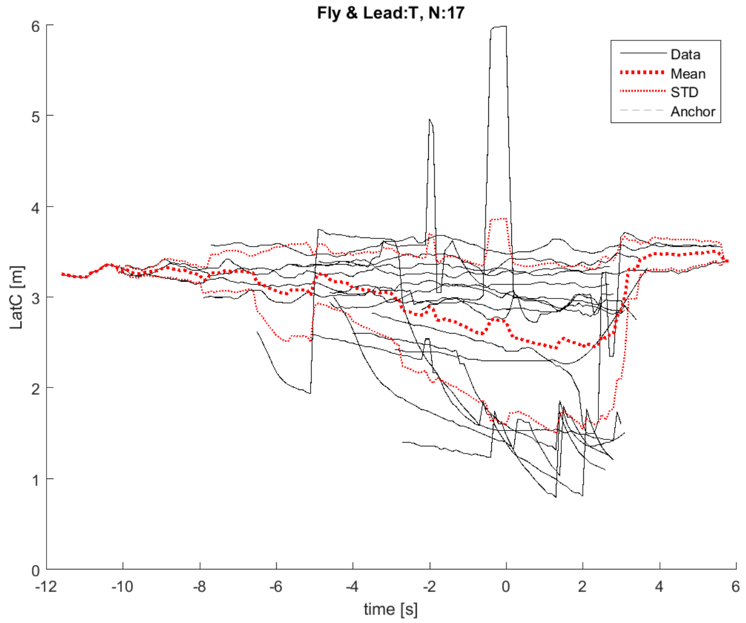


Figure D.24: *Lane width for flying overtaking strategy and leading vehicle present.*

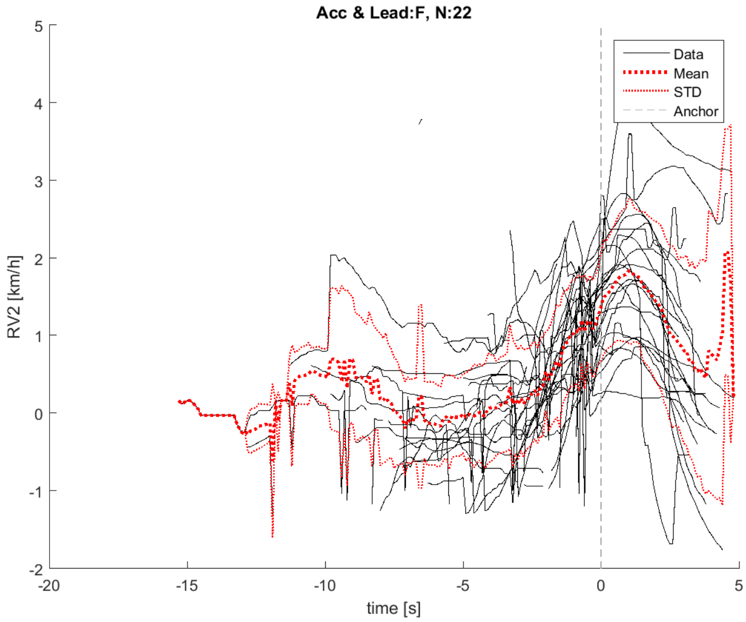


Figure D.25: *Lateral clearance for accelerative overtaking strategy and leading vehicle not present.*

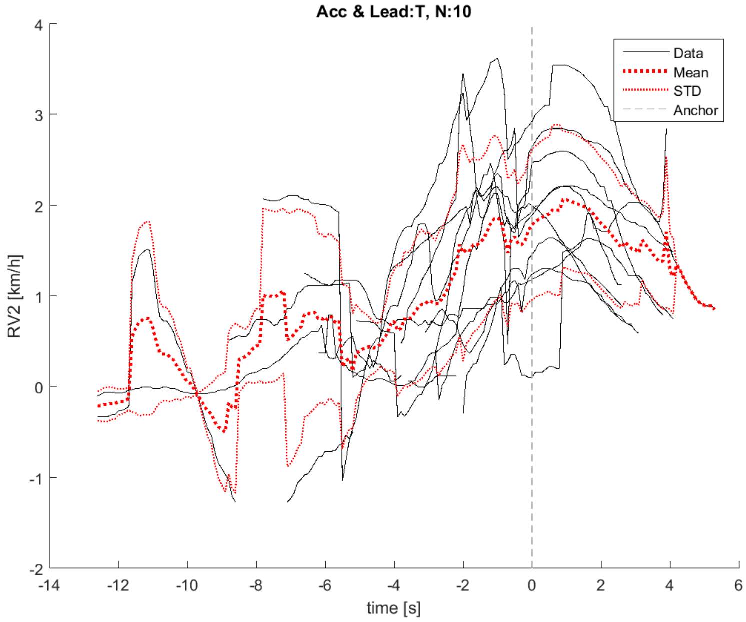


Figure D.26: *Lateral clearance for accelerative overtaking strategy and leading vehicle present.*

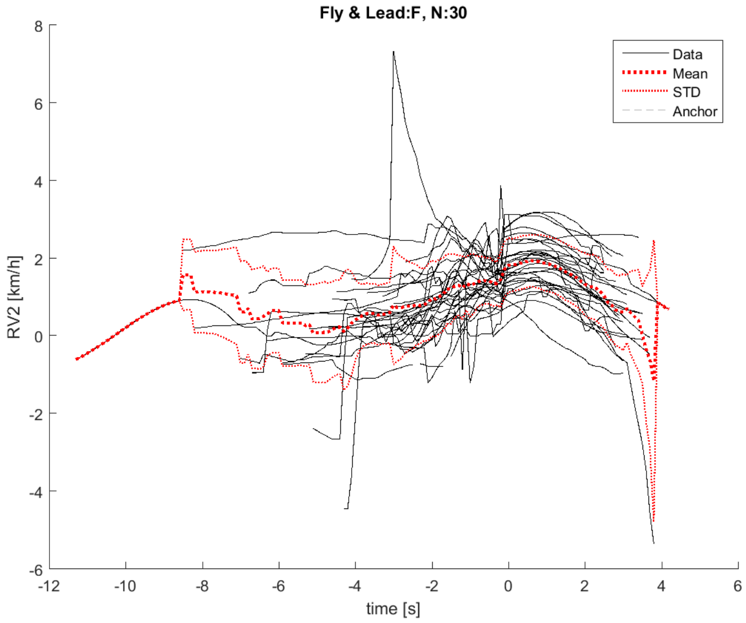


Figure D.27: *Lateral clearance for flying overtaking strategy and leading vehicle not present.*

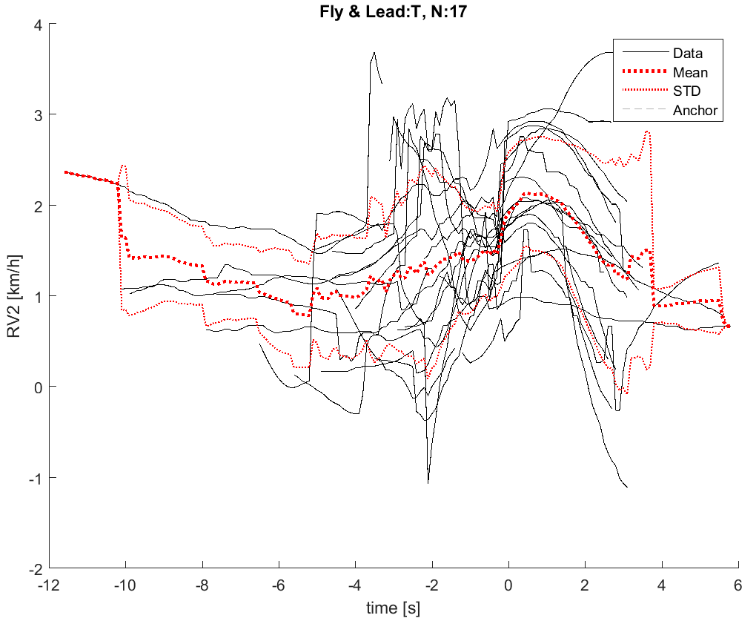


Figure D.28: *Lateral clearance for flying overtaking strategy and leading vehicle present.*

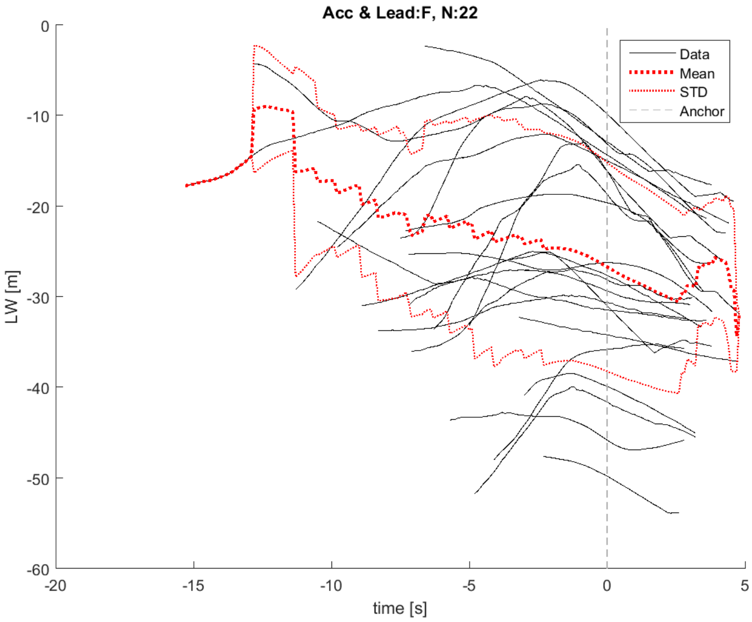


Figure D.29: *Relative velocity to VRU for accelerative overtaking strategy and leading vehicle not present.*

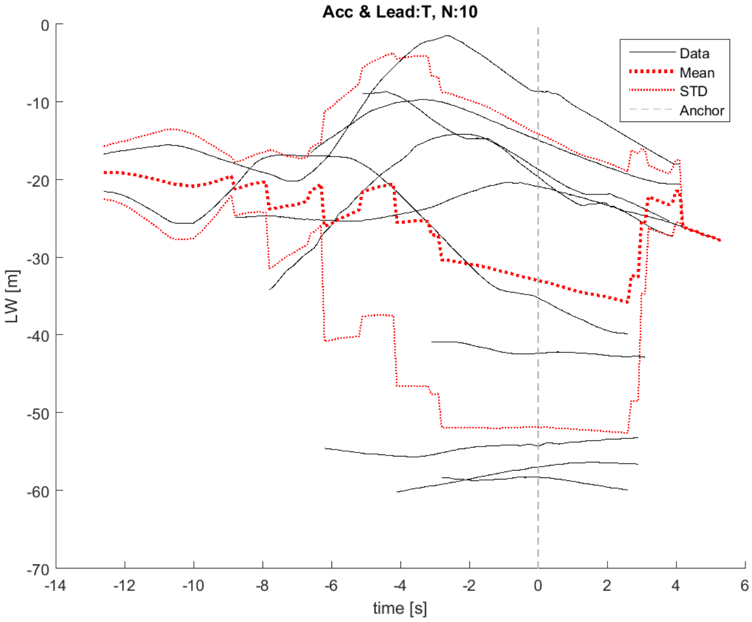


Figure D.30: *Relative velocity to VRU for accelerative overtaking strategy and leading vehicle present.*

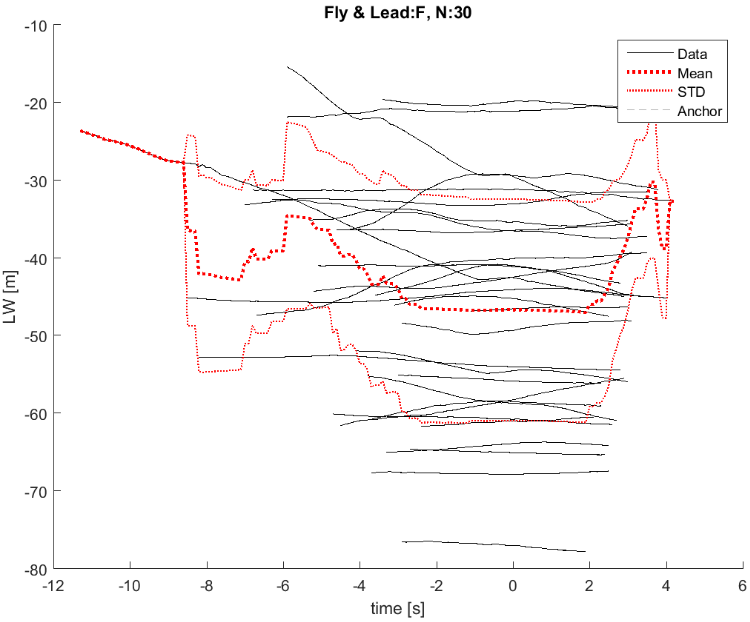


Figure D.31: *Relative velocity to VRU for flying overtaking strategy and leading vehicle not present.*

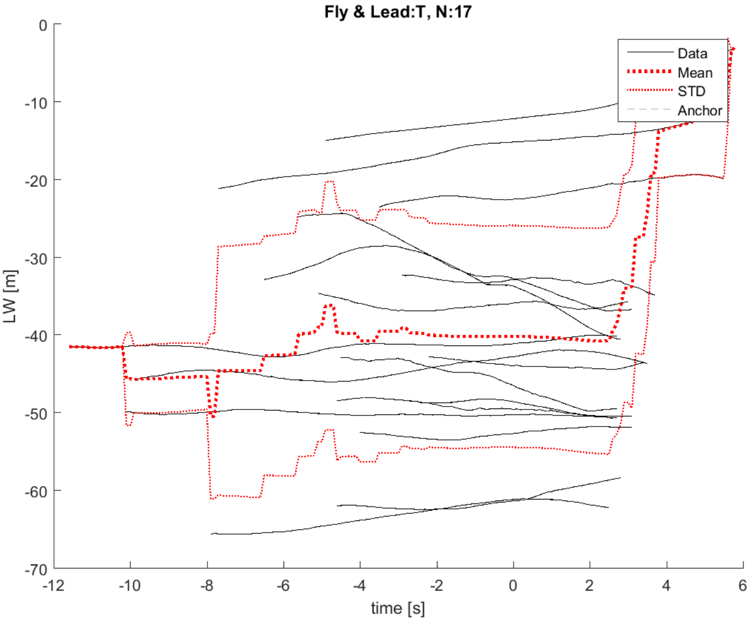


Figure D.32: *Relative velocity to VRU for flying overtaking strategy and leading vehicle present.*

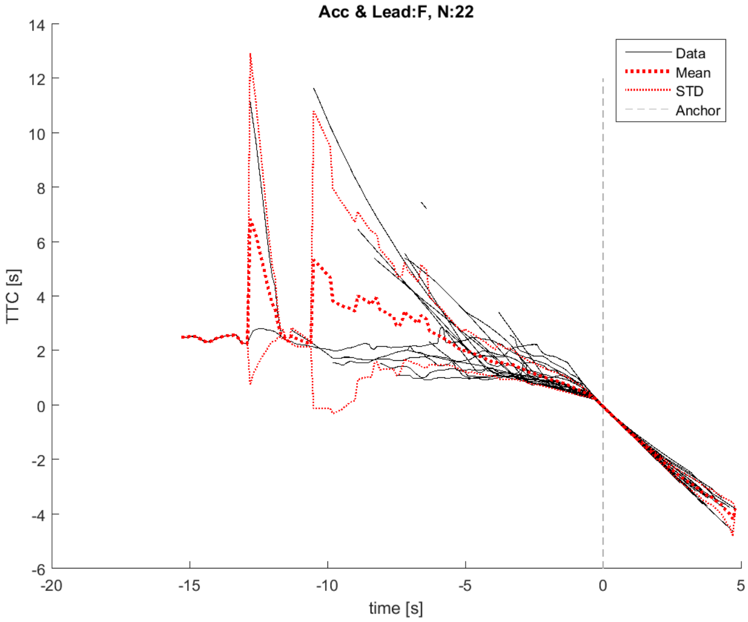


Figure D.33: *Time to collision for accelerative overtaking strategy and leading vehicle not present.*

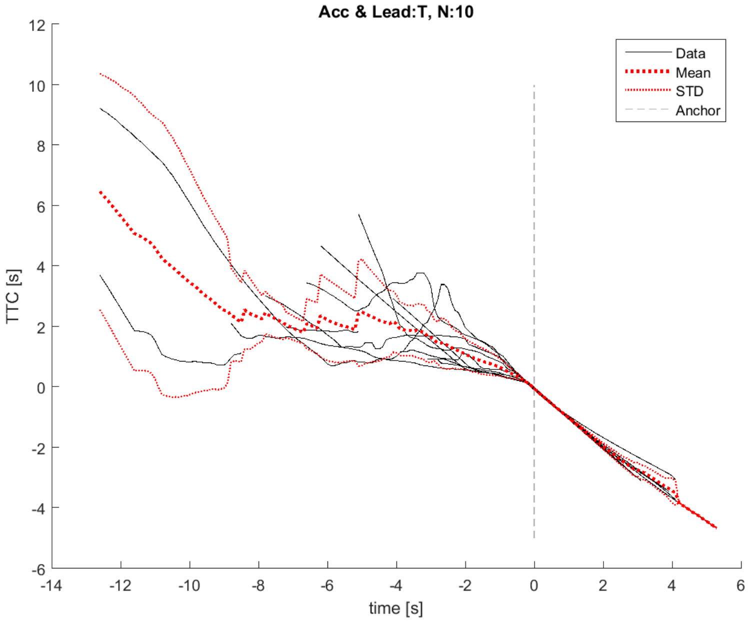


Figure D.34: *Time to collision for accelerative overtaking strategy and leading vehicle present.*

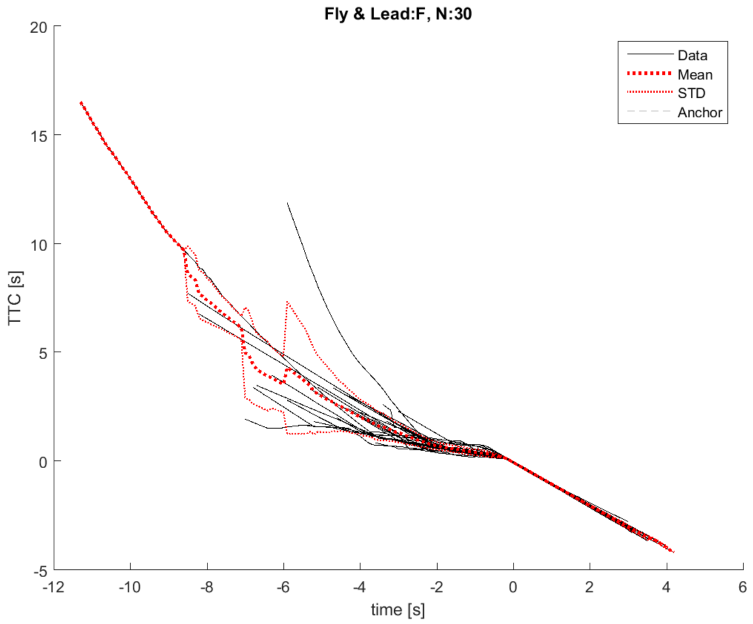


Figure D.35: *Time to collision for flying overtaking strategy and leading vehicle not present.*

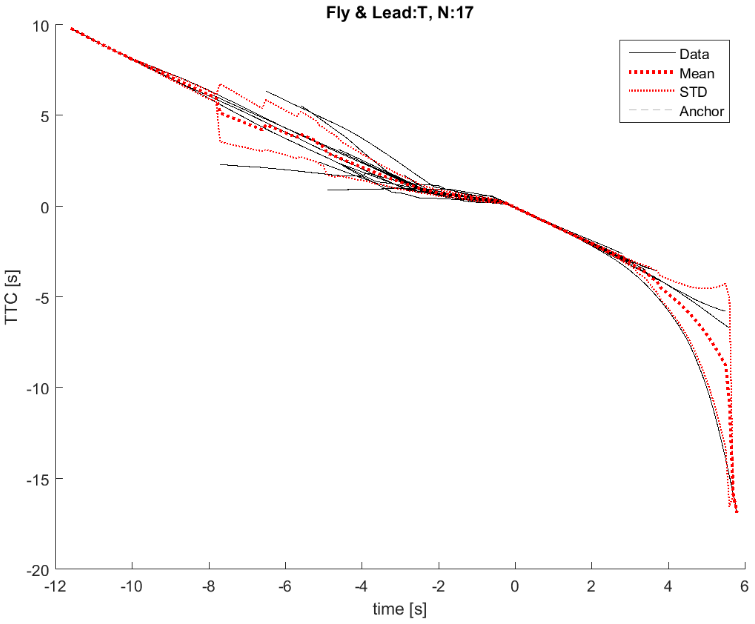


Figure D.36: *Time to collision for flying overtaking strategy and leading vehicle present.*

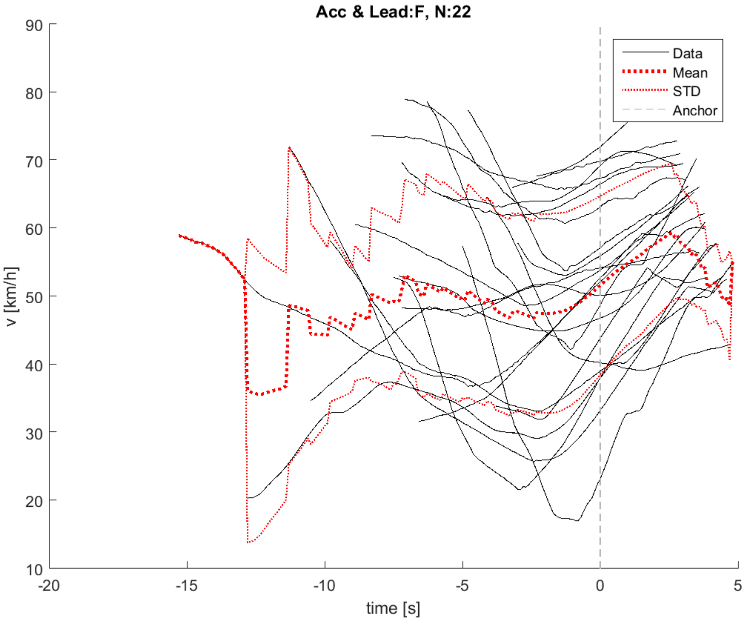


Figure D.37: *Ego vehicle speed for accelerative overtaking strategy and leading vehicle not present.*

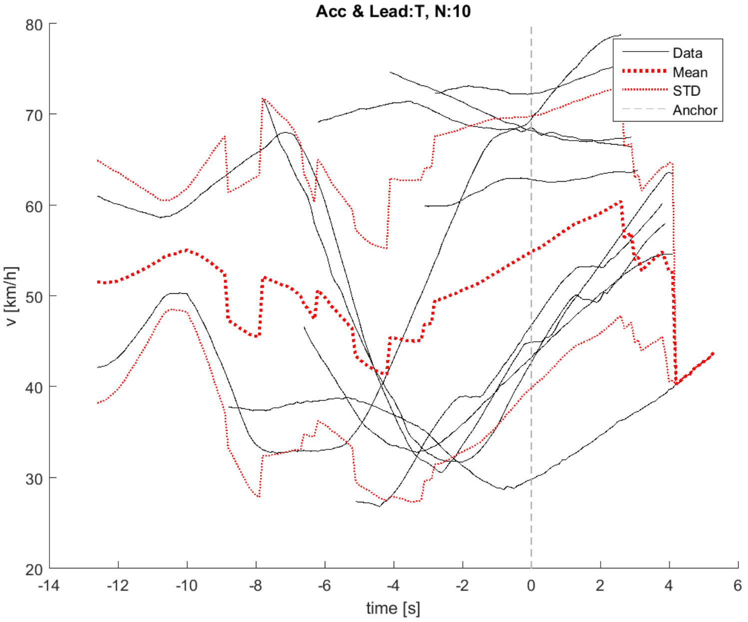


Figure D.38: *Ego vehicle speed for accelerative overtaking strategy and leading vehicle present.*

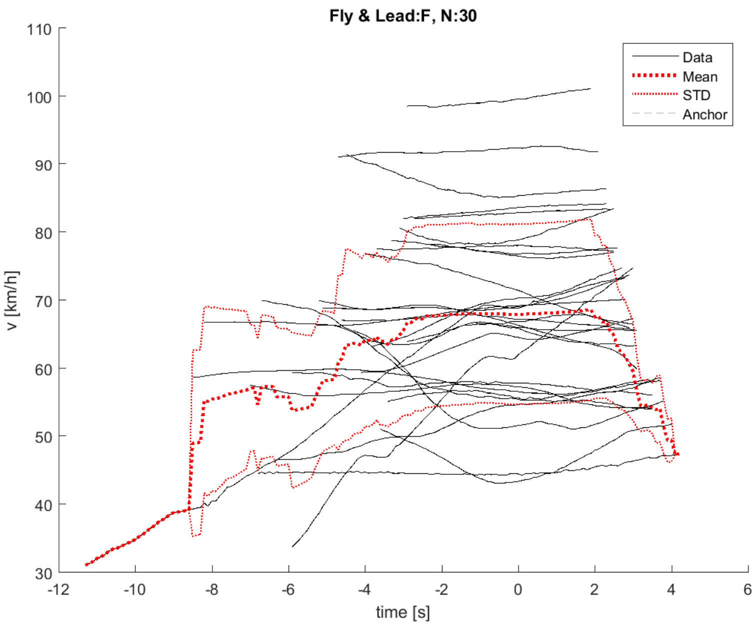


Figure D.39: *Ego vehicle speed for flying overtaking strategy and leading vehicle not present.*

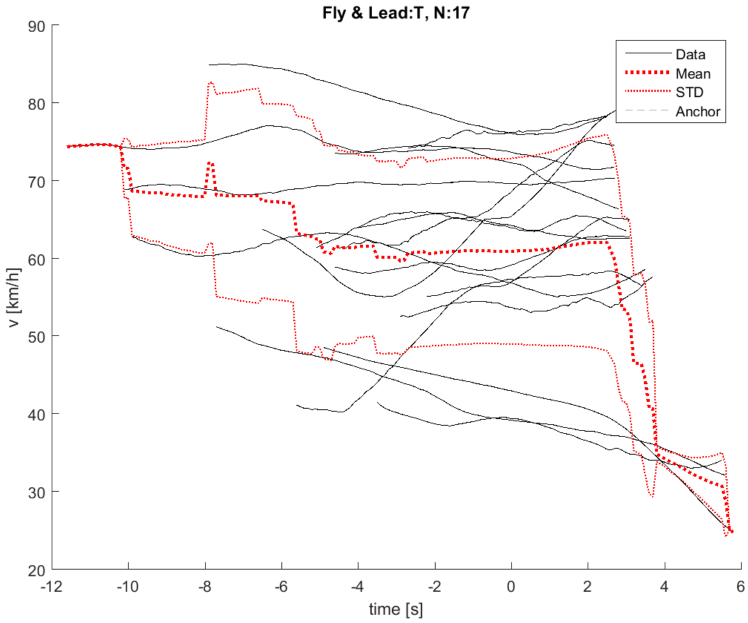


Figure D.40: *Ego vehicle speed for flying overtaking strategy and leading vehicle present.*

E Data Sorted by the Visibility of Lane Markings

In the following tables 'Central Line' implies that only the central lane marking was present, and 'Both Lines' implies that both the central lane marking and the side lane marking was present during the execution of the overtaking manoeuvre.

E.1 Tables

Table E.1: CZB [m]. Data sorted by overtaking manoeuvre strategy and the Visibility of Lane Markings.

			Phase 1	Phase 2	Phase 3	Phase 4
Accelerative	Central Line	Mean	68.90	35.42	3.24	59.11
		STD	±58.59	±48.14	±1.69	±38.01
	Both Lines	Mean	42.34	15.67	2.60	50.53
		STD	±37.41	±14.42	±2.00	±32.34
Flying	Central Line	Mean	53.32	19.94	3.38	65.18
		STD	±35.97	±16.31	±2.57	±38.04
	Both lines	Mean	129.00	54.31	4.34	67.39
		STD	±111.54	±65.68	±4.57	±41.06

Table E.2: p-values for CZB. Phase 1. Data Sorted by the Visibility of Lane Markings.

		Accelerative		Flying	
		Central Line	Both Lines	Central Line	Both Lines
Accelerative	Central Line	-	**p<0.001**	**p<0.001**	**p<0.001**
	Both Lines	-	-	**p<0.001**	**p<0.001**
Flying	Central Line	-	-	-	**p<0.001**
	Both Lines	-	-	-	-

Table E.3: p-values for CZB. Phase 2. Data Sorted by the Visibility of Lane Markings.

		Accelerative		Flying	
		Central Line	Both Lines	Central Line	Both Lines
Accelerative	Central Line	-	**p<0.001**	**p<0.001**	**p<0.001**
	Both Lines	-	-	**p<0.001**	**p<0.001**
Flying	Central Line	-	-	-	**p<0.001**
	Both Lines	-	-	-	-

Table E.4: p-values for CZB. Phase 3. Data Sorted by the Visibility of Lane Markings.

		Accelerative		Flying	
		Central Line	Both Lines	Central Line	Both Lines
Accelerative	Central Line	-	0.03	0.71	0.07
	Both Lines	-	-	0.01	**p<0.001**
Flying	Central Line	-	-	-	0.10
	Both Lines	-	-	-	-

Table E.5: p-values for CZB. Phase 4. Data Sorted by the Visibility of Lane Markings.

		Accelerative		Flying	
		Central Line	Both Lines	Central Line	Both Lines
Accelerative	Central Line	-	**p<0.001**	0.02	**0.001**
	Both Lines	-	-	**p<0.001**	**p<0.001**
Flying	Central Line	-	-	-	0.30
	Both Lines	-	-	-	-

Table E.6: TTC [s]. Data Sorted by the Visibility of Lane Markings.

			Phase 1	Phase 2	Phase 3	Phase 4
Accelerative	Central Line	Mean	2.47	1.01	-0.15	-1.94
		STD	±2.19	±0.85	±0.22	±0.97
	Both Lines	Mean	2.09	0.84	-0.15	-1.95
		STD	±1.59	±0.67	±0.35	±0.92
Flying	Central Line	Mean	1.37	0.55	-0.16	-1.83
		STD	±0.91	±0.61	±0.13	±1.07
	Both lines	Mean	3.25	1.16	-0.11	-2.01
		STD	±2.96	±1.41	±0.24	±1.75

Table E.7: p-values for TTC. Phase 1. Data Sorted by the Visibility of Lane Markings.

		Accelerative		Flying	
		Central Line	Both Lines	Central Line	Both Lines
Accelerative	Central Line	-	**p<0.001**	**p<0.001**	**p<0.001**
	Both Lines	-	-	**p<0.001**	**p<0.001**
Flying	Central Line	-	-	-	**p<0.001**
	Both Lines	-	-	-	-

Table E.8: p-values for TTC. Phase 2. Data Sorted by the Visibility of Lane Markings.

		Accelerative		Flying	
		Central Line	Both Lines	Central Line	Both Lines
Accelerative	Central Line	-	**0.004**	**p<0.001**	0.13
	Both Lines	-	-	**p<0.001**	**p<0.001**
Flying	Central Line	-	-	-	**p<0.001**
	Both Lines	-	-	-	-

Table E.9: p-values for TTC. Phase 3. Data Sorted by the Visibility of Lane Markings.

		Accelerative		Flying	
		Central Line	Both Lines	Central Line	Both Lines
Accelerative	Central Line	-	0.98	0.73	0.30
	Both Lines	-	-	0.78	0.34
Flying	Central Line	-	-	-	0.10
	Both Lines	-	-	-	-

Table E.10: p-values for TTC. Phase 4. Data Sorted by the Visibility of Lane Markings.

		Accelerative		Flying	
		Central Line	Both Lines	Central Line	Both Lines
Accelerative	Central Line	-	0.91	0.11	0.48
	Both Lines	-	-	0.04	0.40
Flying	Central Line	-	-	-	0.03
	Both Lines	-	-	-	-

Table E.11: LatC [m]. Data Sorted by the Visibility of Lane Markings.

			Phase 1	Phase 2	Phase 3	Phase 4
Accelerative	Central Line	Mean	0.49	1.31	2.12	1.88
		STD	±0.73	±0.99	±0.76	±1.23
	Both Lines	Mean	0.17	1.01	1.13	1.21
		STD	±0.67	±0.67	±0.71	±0.86
Flying	Central Line	Mean	0.50	1.25	1.67	1.44
		STD	±0.83	±0.74	±0.89	±0.90
	Both lines	Mean	0.86	1.52	1.62	1.63
		STD	±1.01	±0.76	±0.82	±0.98

Table E.12: p-values for LatC. Phase 1. Data Sorted by the Visibility of Lane Markings.

		Accelerative		Flying	
		Central Line	Both Lines	Central Line	Both Lines
Accelerative	Central Line	-	**p<0.001**	0.95	**p<0.001**
	Both Lines	-	-	**p<0.001**	**p<0.001**
Flying	Central Line	-	-	-	**p<0.001**
	Both Lines	-	-	-	-

Table E.13: p-values for LatC. Phase 2. Data Sorted by the Visibility of Lane Markings.

		Accelerative		Flying	
		Central Line	Both Lines	Central Line	Both Lines
Accelerative	Central Line	-	**p<0.001**	0.42	**p<0.001**
	Both Lines	-	-	**p<0.001**	**p<0.001**
Flying	Central Line	-	-	-	**p<0.001**
	Both Lines	-	-	-	-

Table E.14: p-values for LatC. Phase 3. Data Sorted by the Visibility of Lane Markings.

		Accelerative		Flying	
		Central Line	Both Lines	Central Line	Both Lines
Accelerative	Central Line	-	**p<0.001**	**p<0.001**	**p<0.001**
	Both Lines	-	-	**p<0.001**	**p<0.001**
Flying	Central Line	-	-	-	0.71
	Both Lines	-	-	-	-

Table E.15: p-values for LatC. Phase 4. Data Sorted by the Visibility of Lane Markings.

		Accelerative		Flying	
		Central Line	Both Lines	Central Line	Both Lines
Accelerative	Central Line	-	**p<0.001**	**p<0.001**	**p<0.001**
	Both Lines	-	-	**p<0.001**	**p<0.001**
Flying	Central Line	-	-	-	**p<0.001**
	Both Lines	-	-	-	-

E.2 Figures

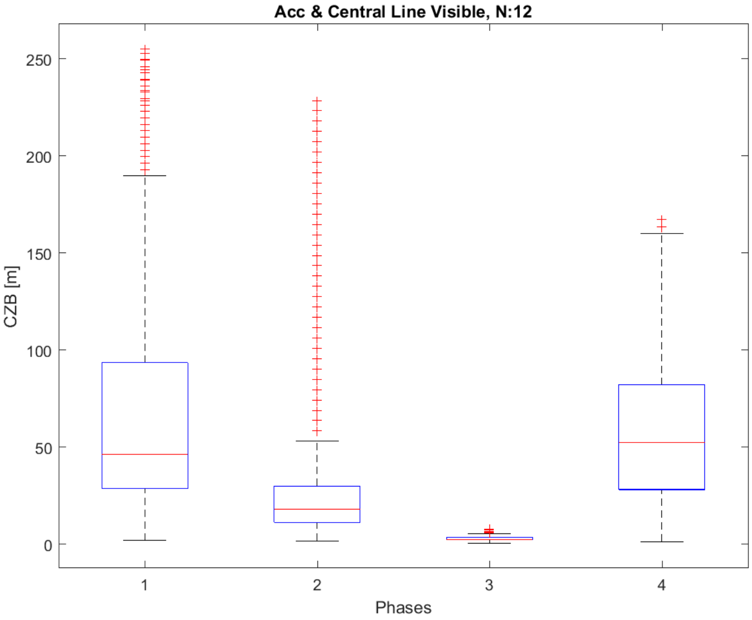


Figure E.1: *Boxplot of minimum distance (CZB) for accelerative overtaking strategy and only central line visible.*

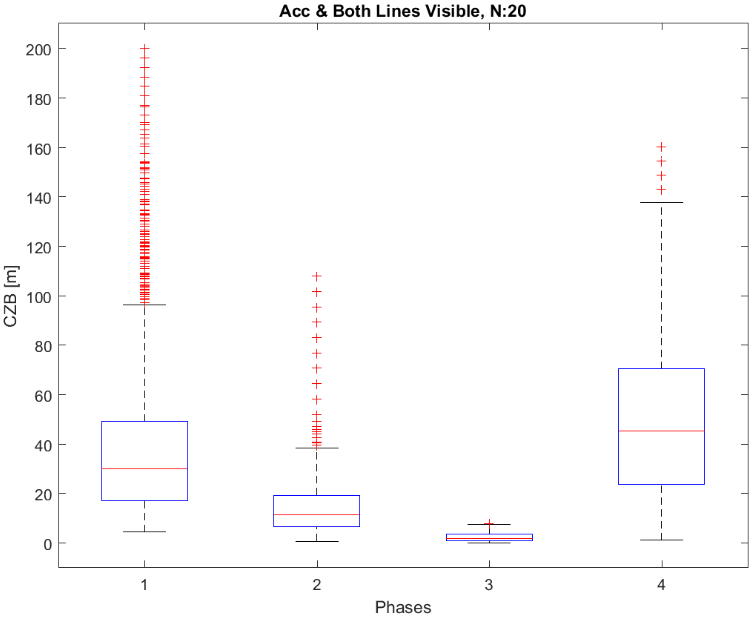


Figure E.2: *Boxplot of minimum distance (CZB) for accelerative overtaking strategy and both lines visible.*

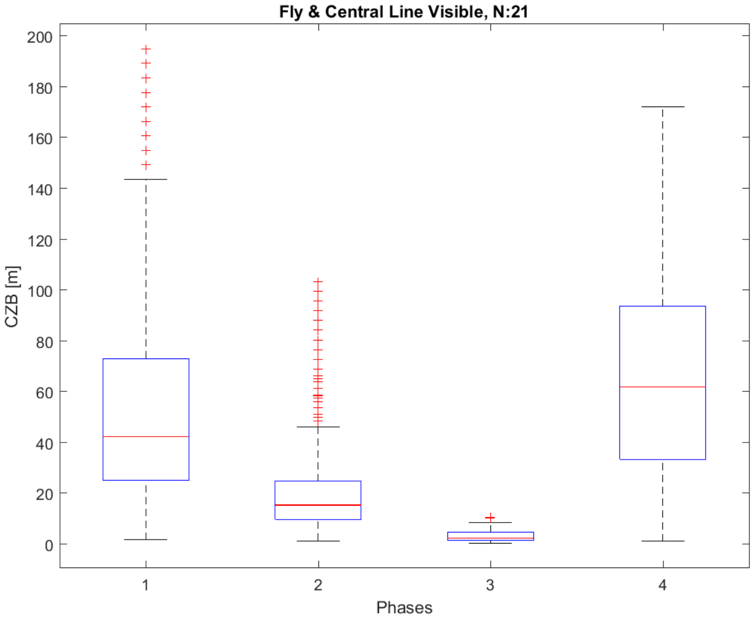


Figure E.3: *Boxplot of minimum distance (CZB) for flying overtaking strategy and only central line visible.*

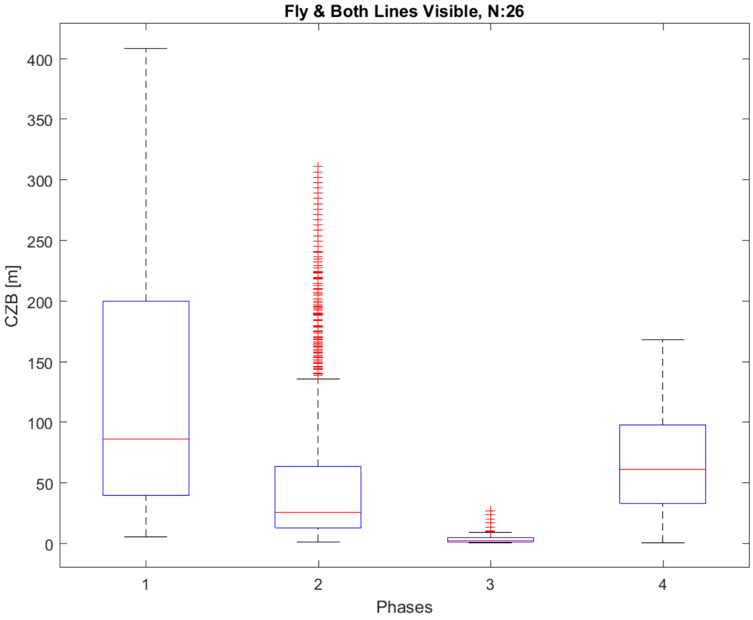


Figure E.4: *Boxplot of minimum distance (CZB) for flying overtaking strategy and both lines visible.*

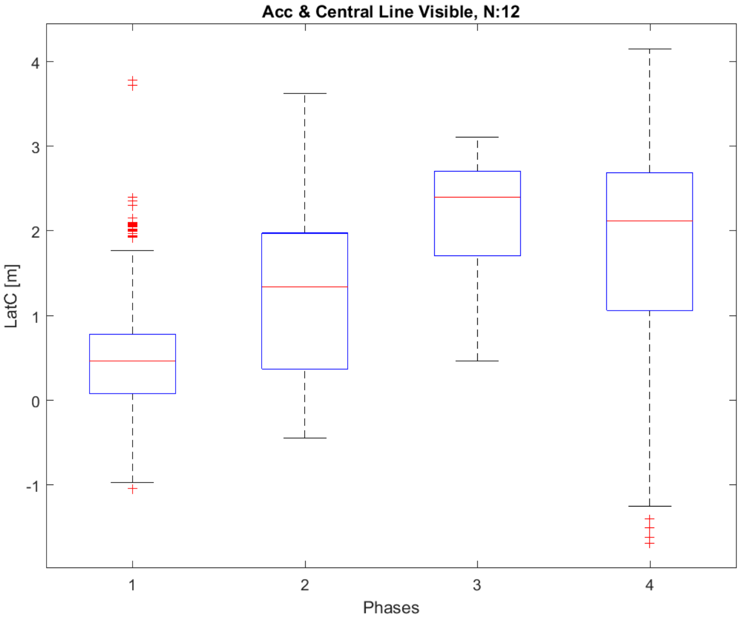


Figure E.5: *Boxplot of lateral clearance for accelerative overtaking strategy and only central line visible.*

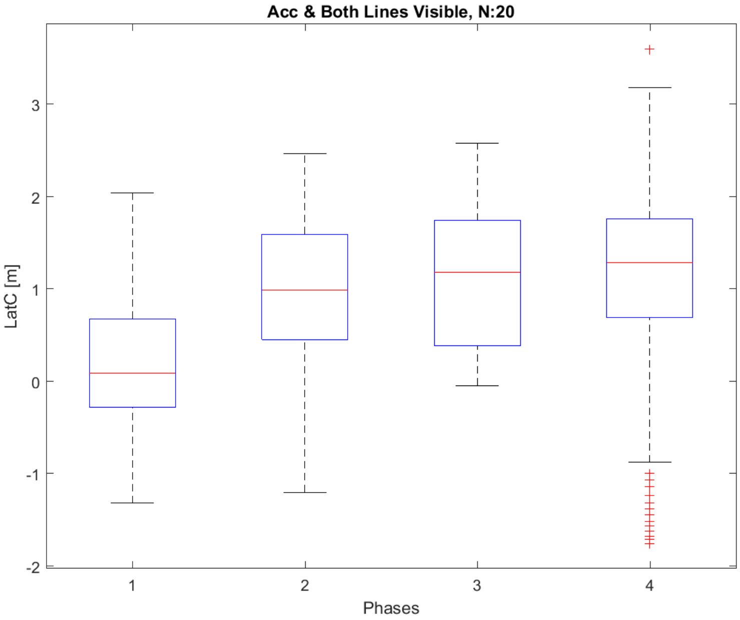


Figure E.6: *Boxplot of lateral clearance for accelerative overtaking strategy and both lines visible.*

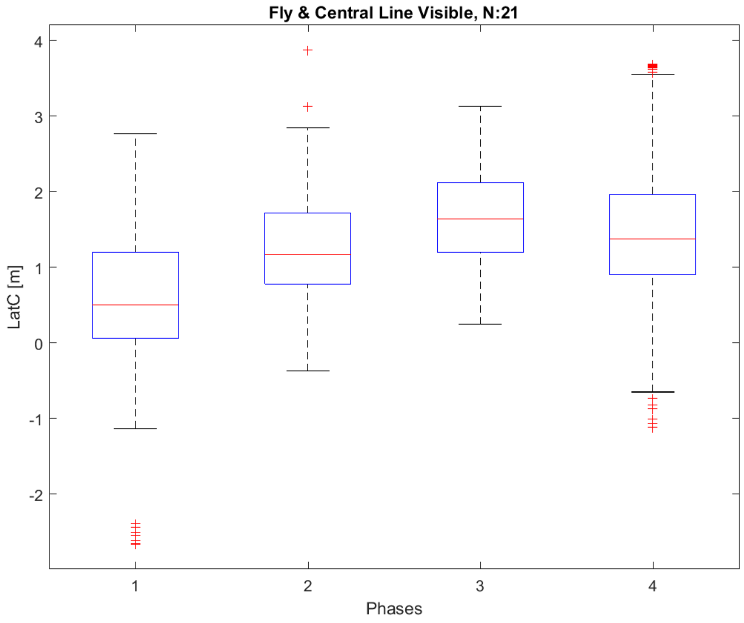


Figure E.7: *Boxplot of lateral clearance for flying overtaking strategy and only central line visible.*

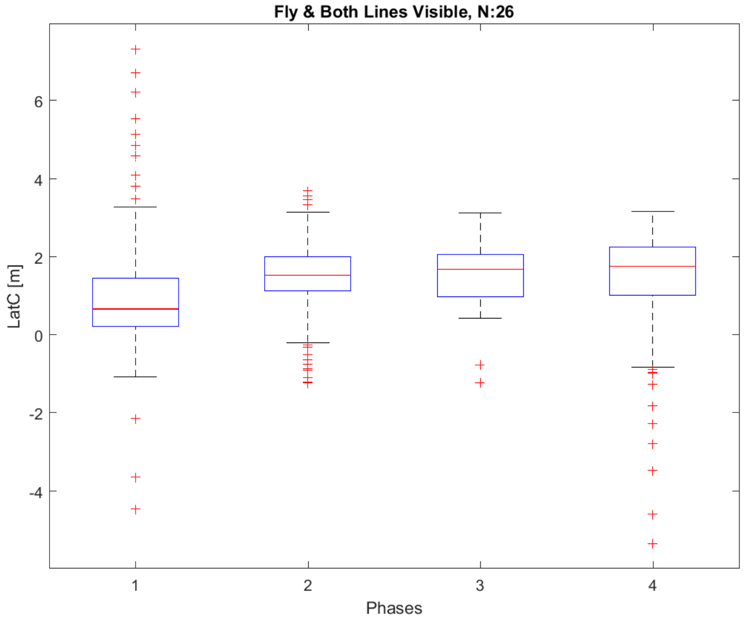


Figure E.8: *Boxplot of lateral clearance for flying overtaking strategy and both lines visible.*

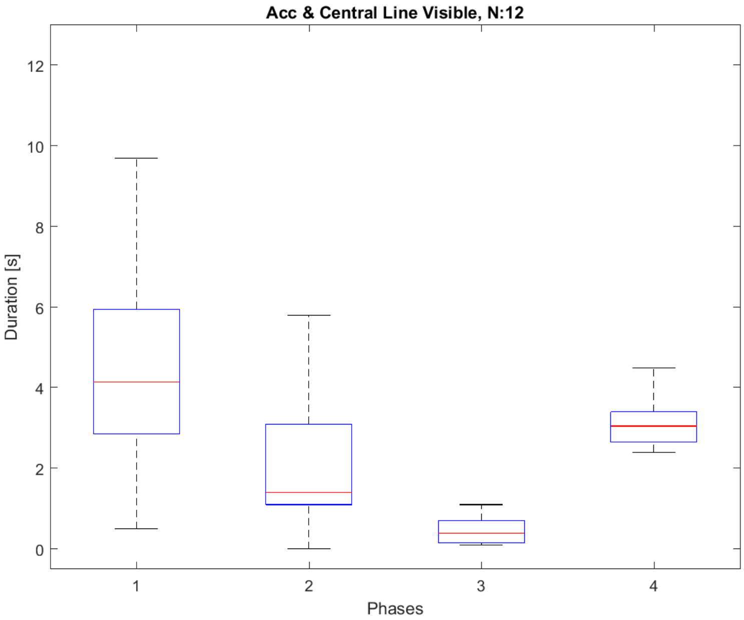


Figure E.9: *Boxplot of phase durations for accelerative overtaking strategy and only central line visible.*

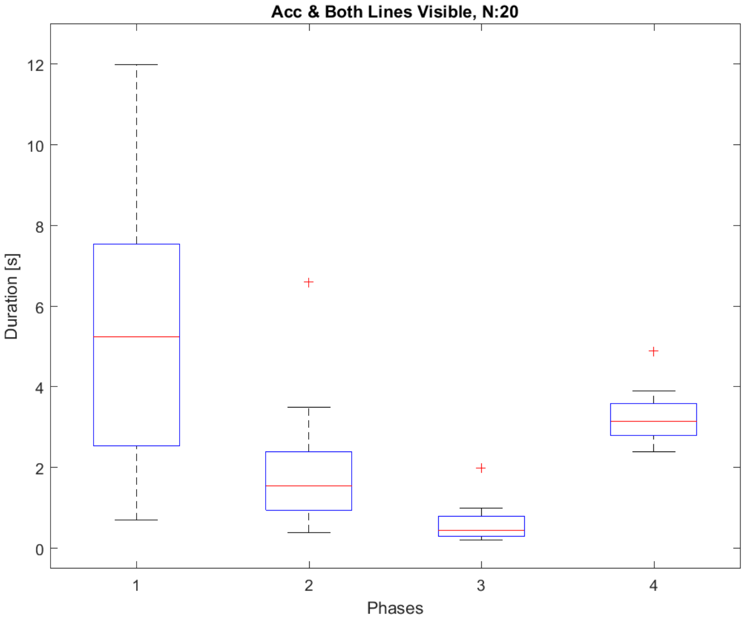


Figure E.10: *Boxplot of phase durations for accelerative overtaking strategy and both lines visible.*

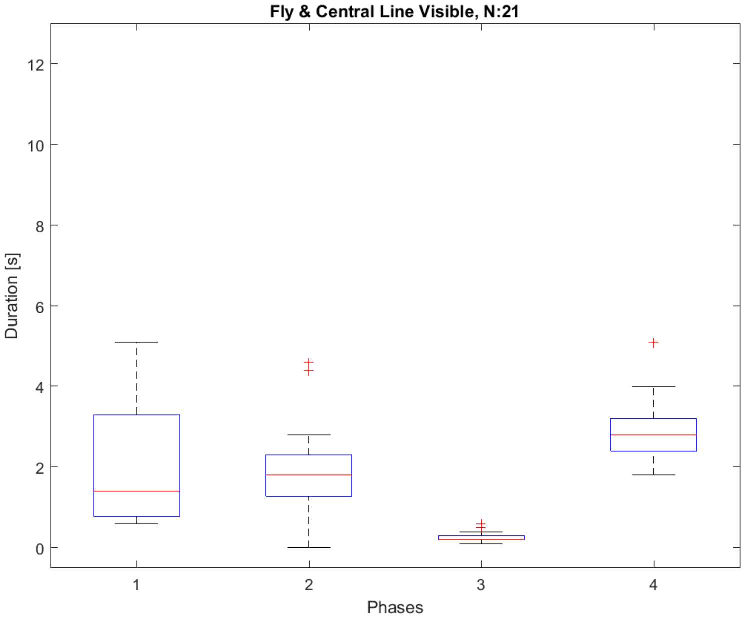


Figure E.11: *Boxplot of phase durations for flying overtaking strategy and only central line visible.*

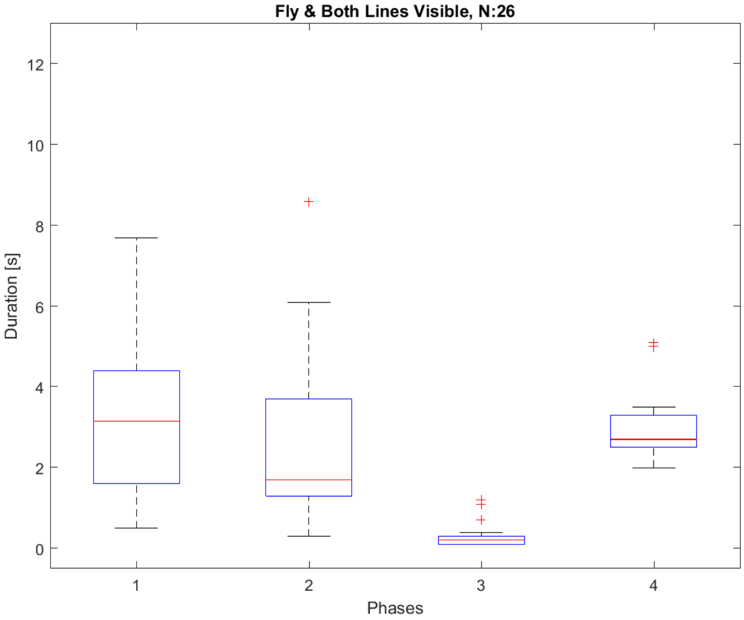


Figure E.12: *Boxplot of phase durations for flying overtaking strategy and both lines visible.*

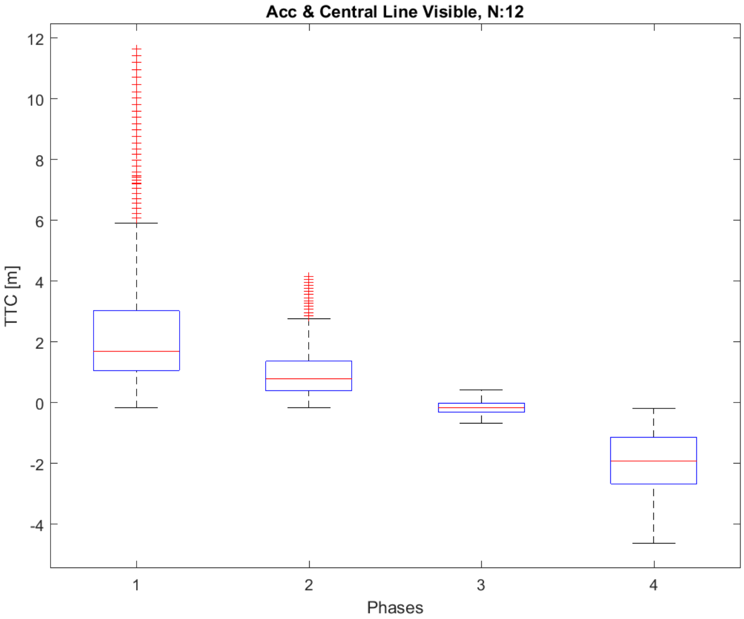


Figure E.13: *Boxplot of time to collision for accelerative overtaking strategy and only central line visible.*

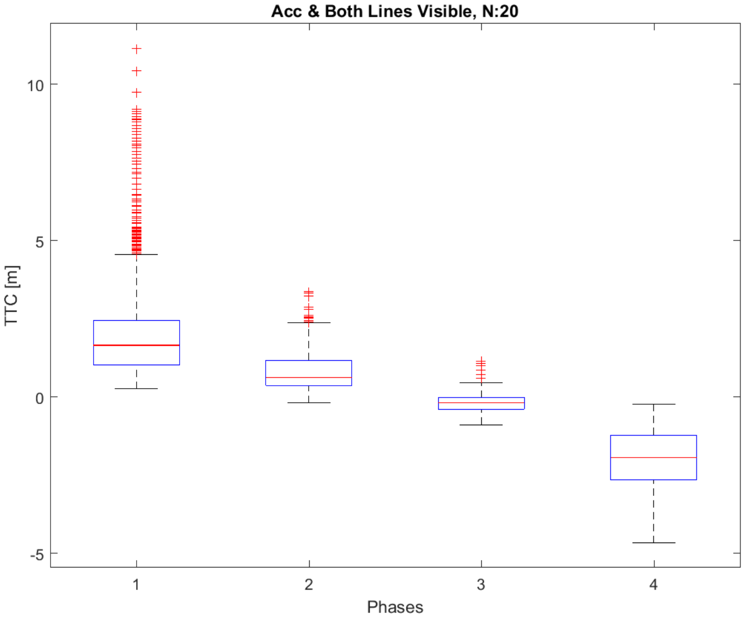


Figure E.14: *Boxplot of time to collision for accelerative overtaking strategy and both lines visible.*

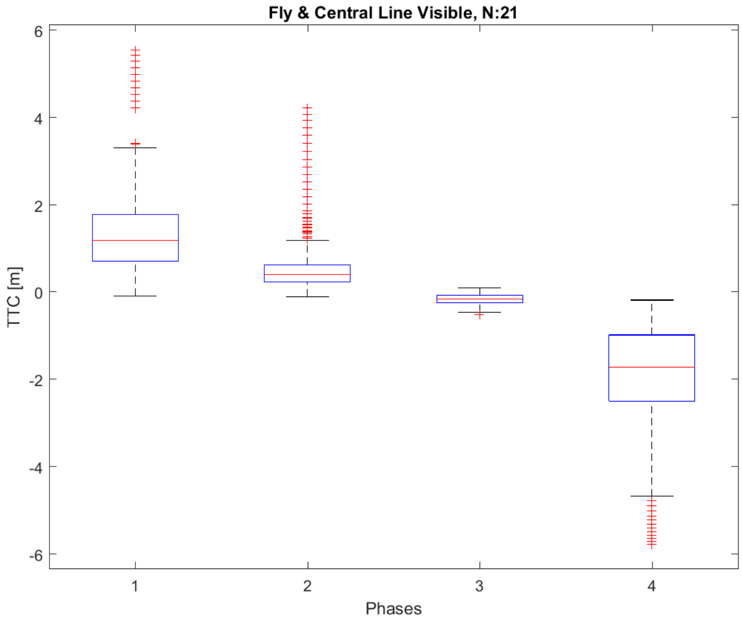


Figure E.15: *Boxplot of time to collision for flying overtaking strategy and only central line visible.*

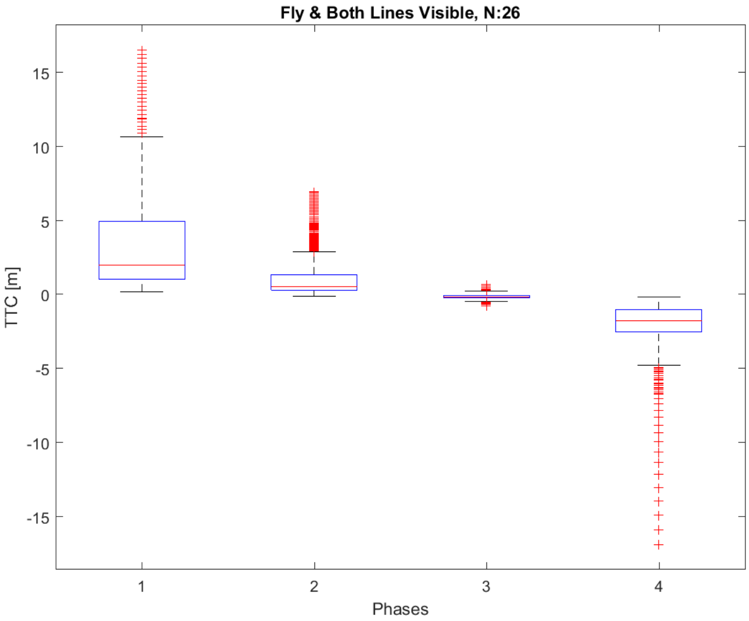


Figure E.16: *Boxplot of time to collision for flying overtaking strategy and both lines visible.*

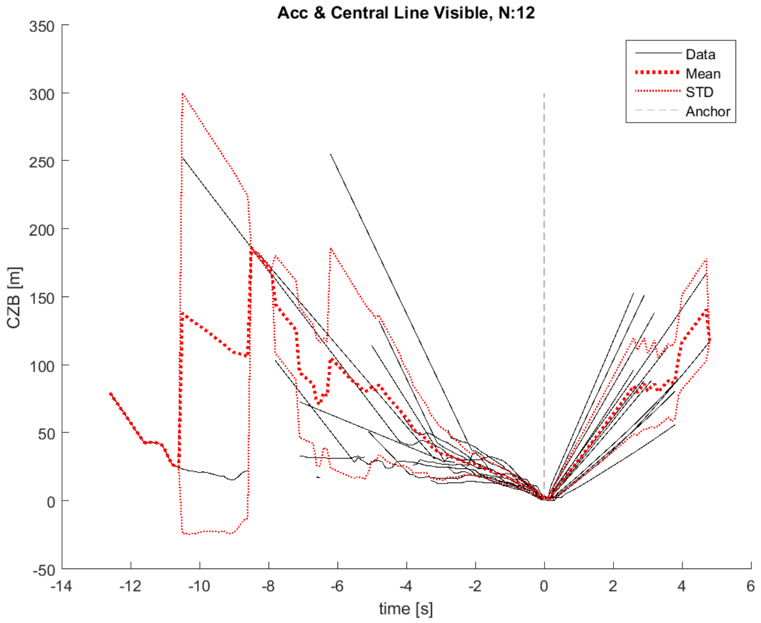


Figure E.17: *Minimum distance (CZB) for accelerative overtaking strategy and only central line visible.*

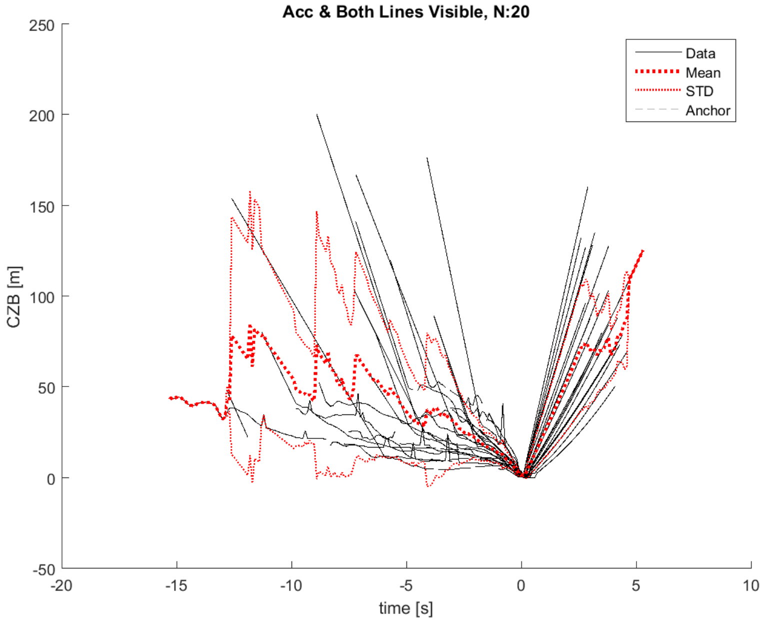


Figure E.18: *Minimum distance (CZB) for accelerative overtaking strategy and both lines visible.*

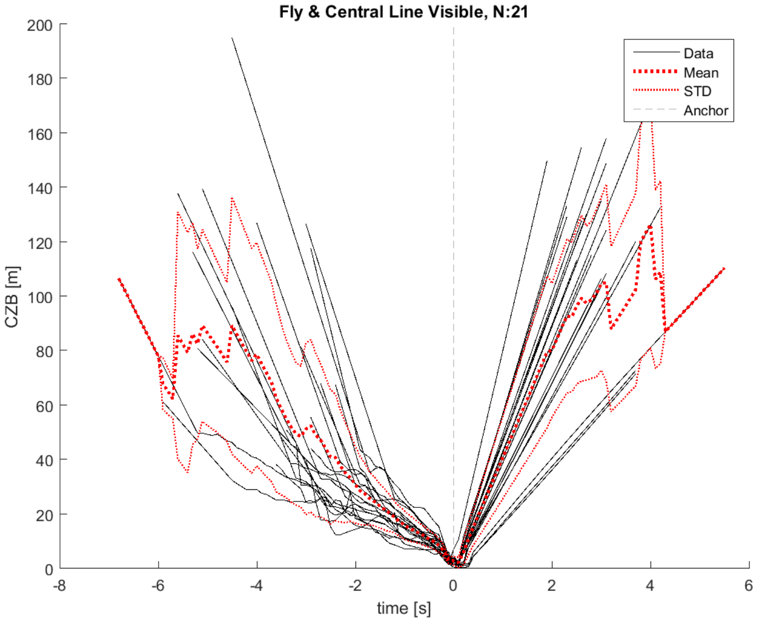


Figure E.19: *Minimum distance (CZB) for flying overtaking strategy and only central line visible.*

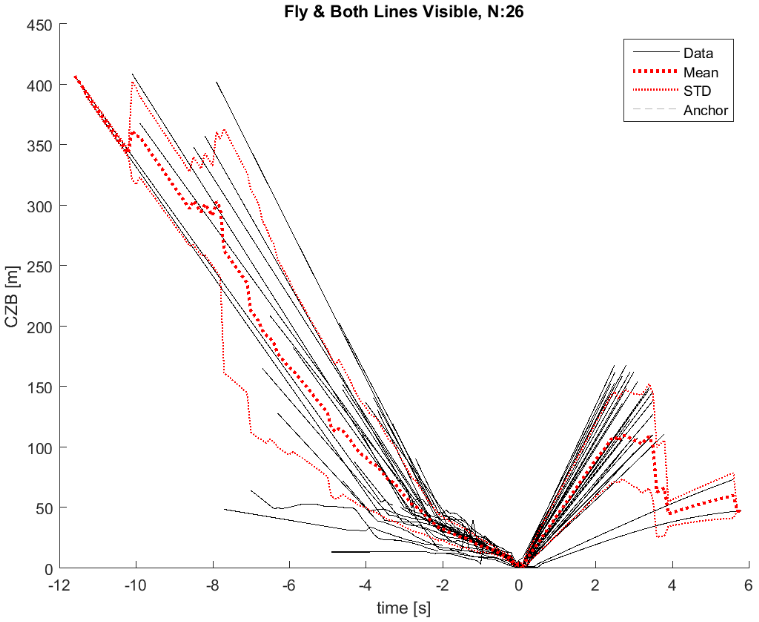


Figure E.20: *Minimum distance (CZB) for flying overtaking strategy and both lines visible.*

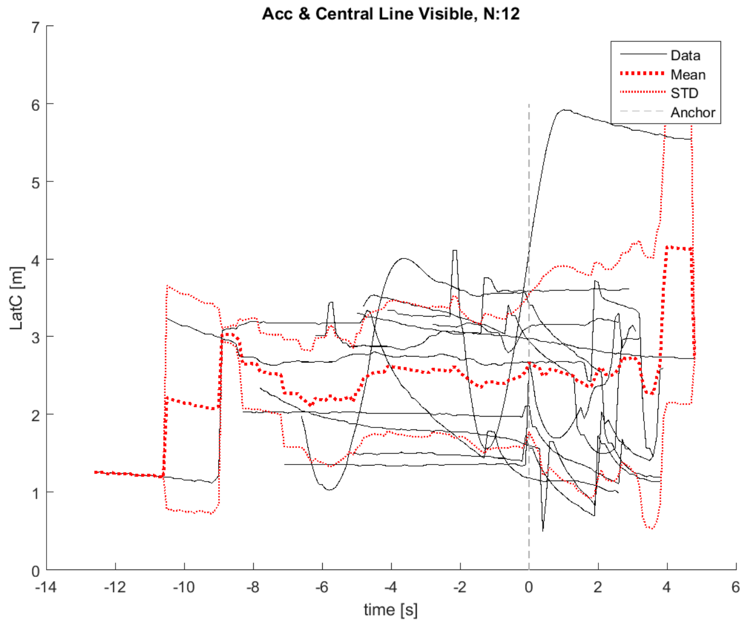


Figure E.21: *Lane width for accelerative overtaking strategy and only central line visible.*

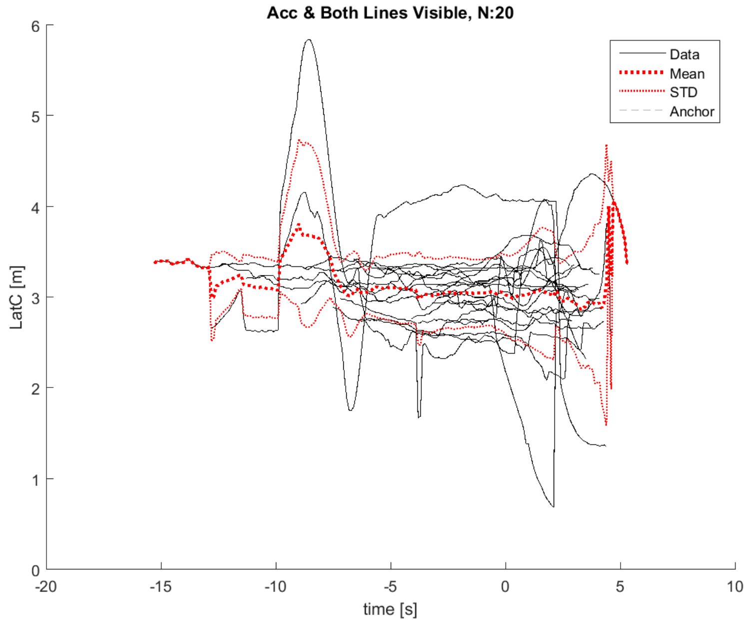


Figure E.22: *Lane width for accelerative overtaking strategy and both lines visible.*

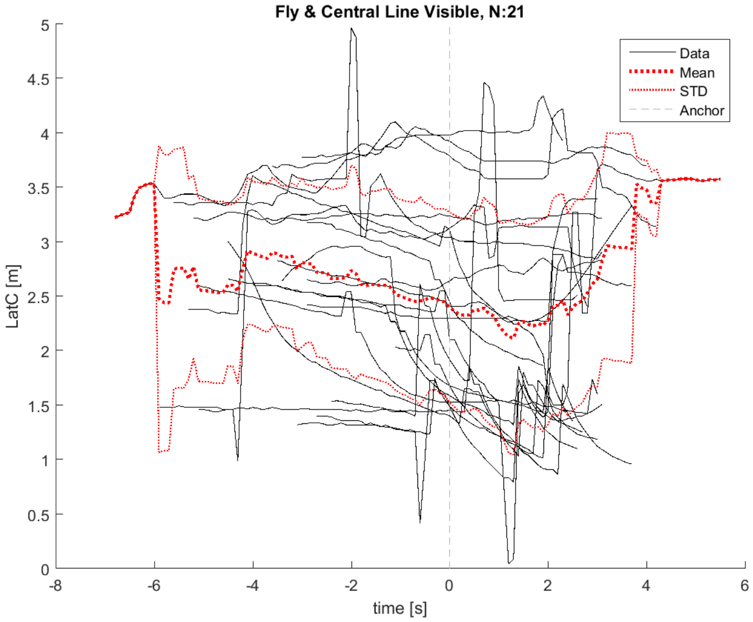


Figure E.23: *Lane width for flying overtaking strategy and only central line visible.*

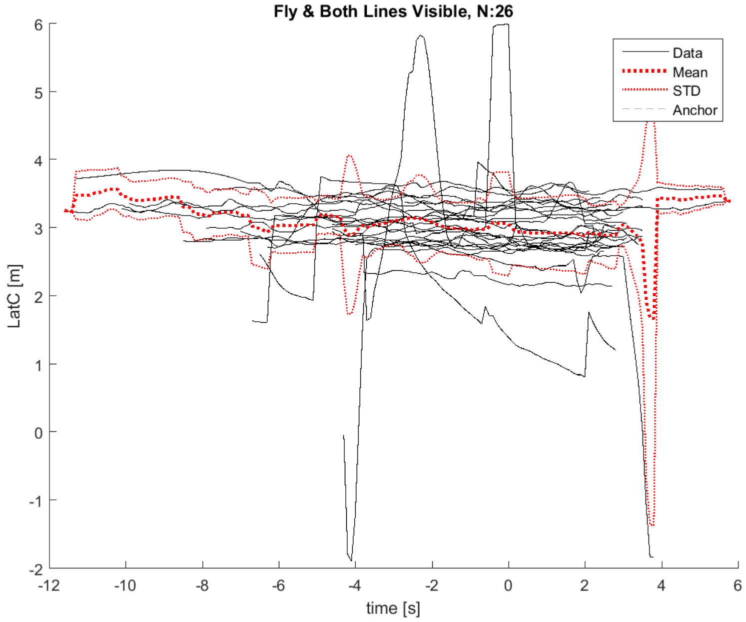


Figure E.24: *Lane width for flying overtaking strategy and both lines visible.*

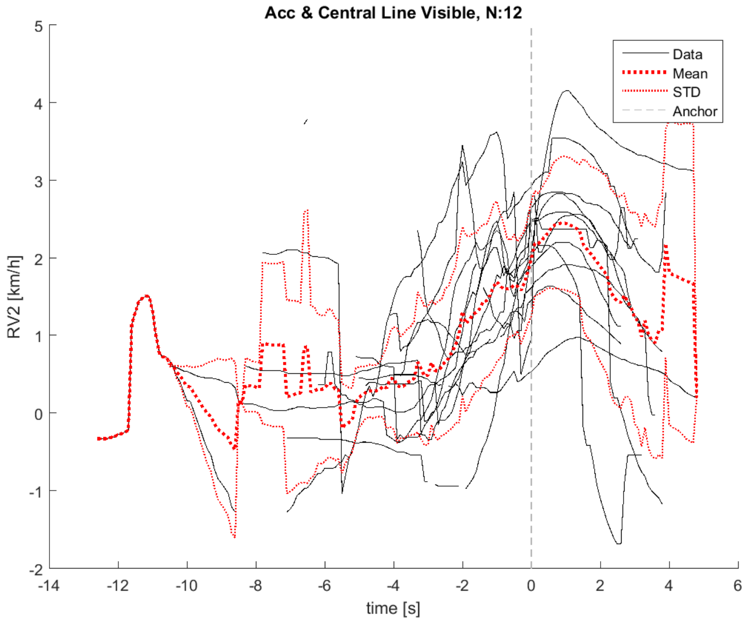


Figure E.25: *Lateral clearance for accelerative overtaking strategy and only central line visible.*

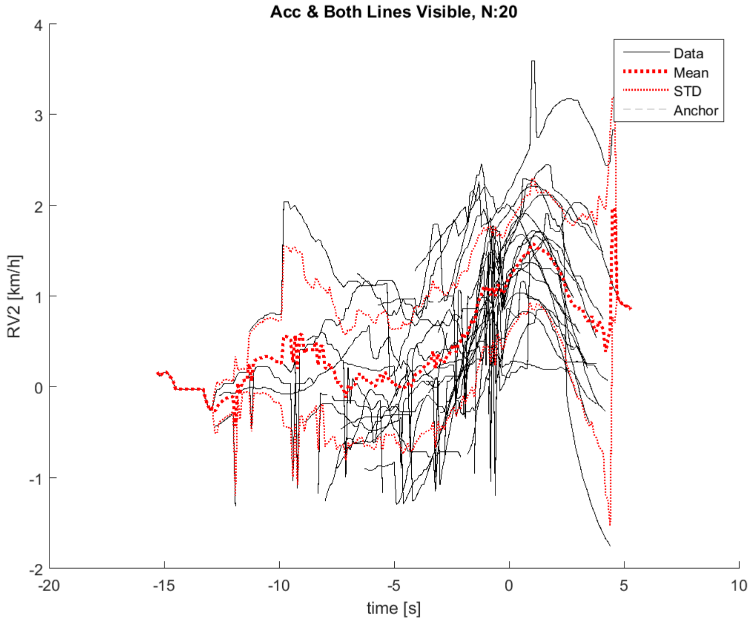


Figure E.26: *Lateral clearance for accelerative overtaking strategy and both lines visible.*

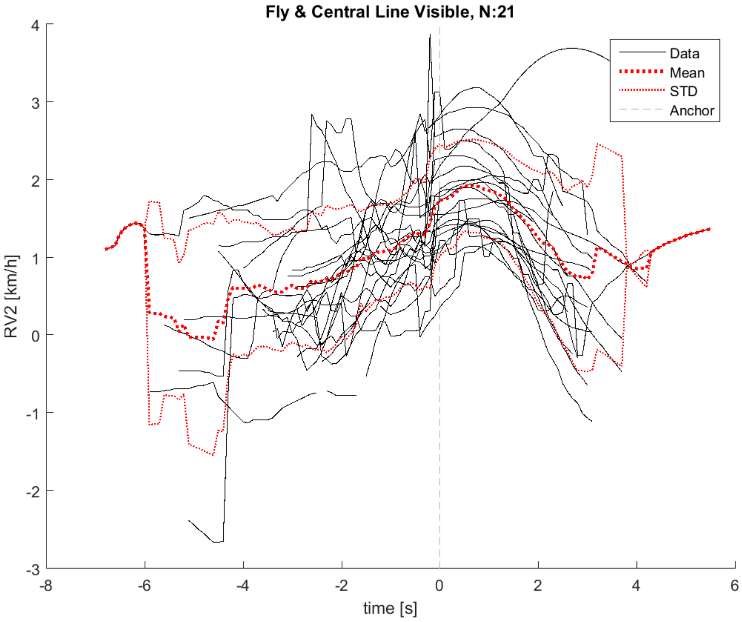


Figure E.27: *Lateral clearance for flying overtaking strategy and only central line visible.*

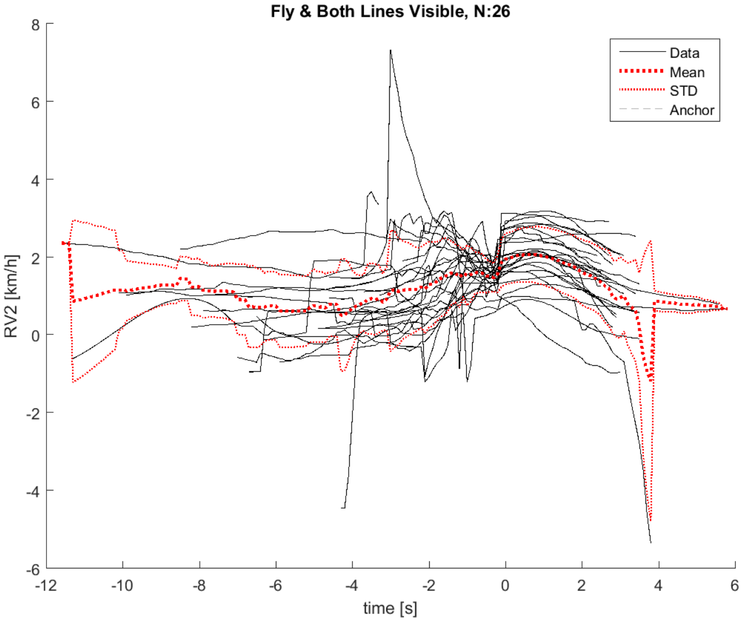


Figure E.28: *Lateral clearance for flying overtaking strategy and both lines visible.*

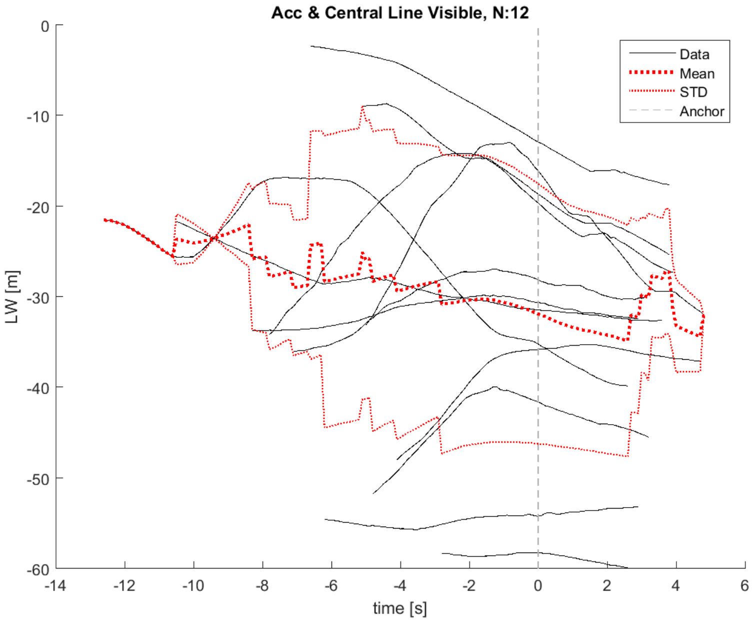


Figure E.29: *Relative velocity to VRU for accelerative overtaking strategy and only central line visible.*

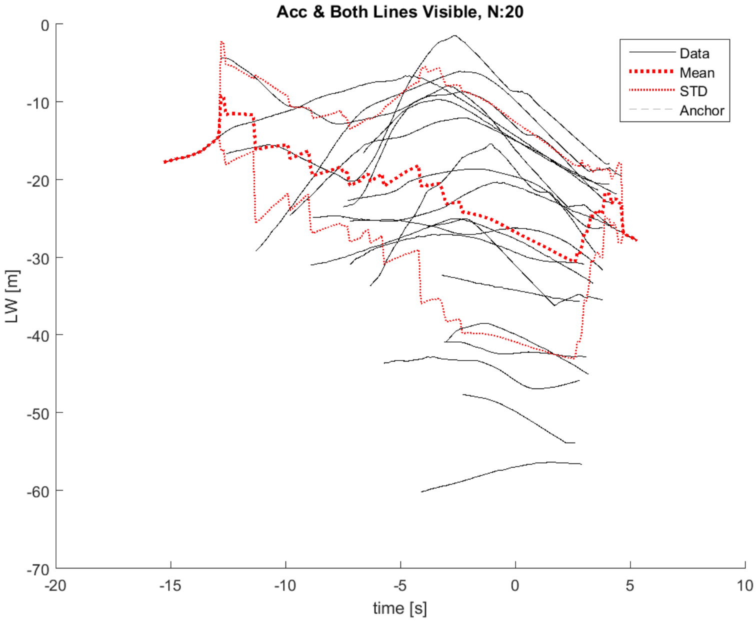


Figure E.30: *Relative velocity to VRU for accelerative overtaking strategy and both lines visible.*

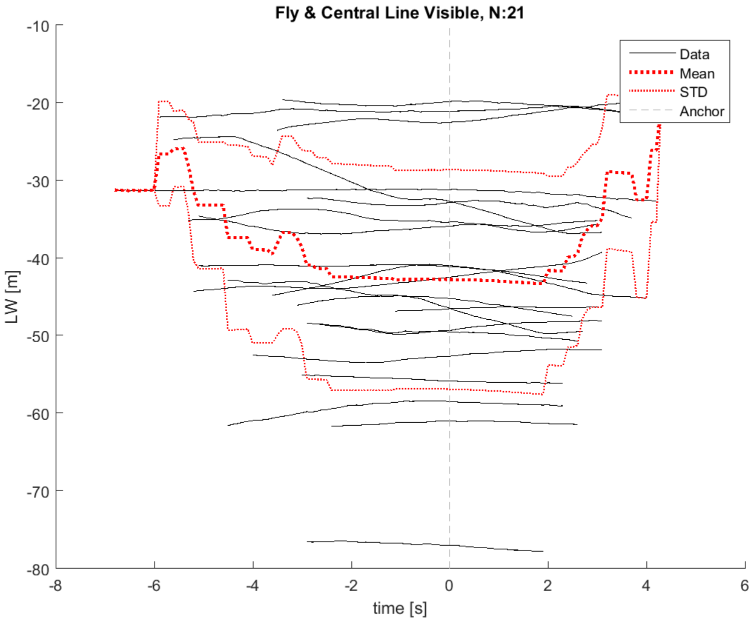


Figure E.31: *Relative velocity to VRU for flying overtaking strategy and only central line visible.*

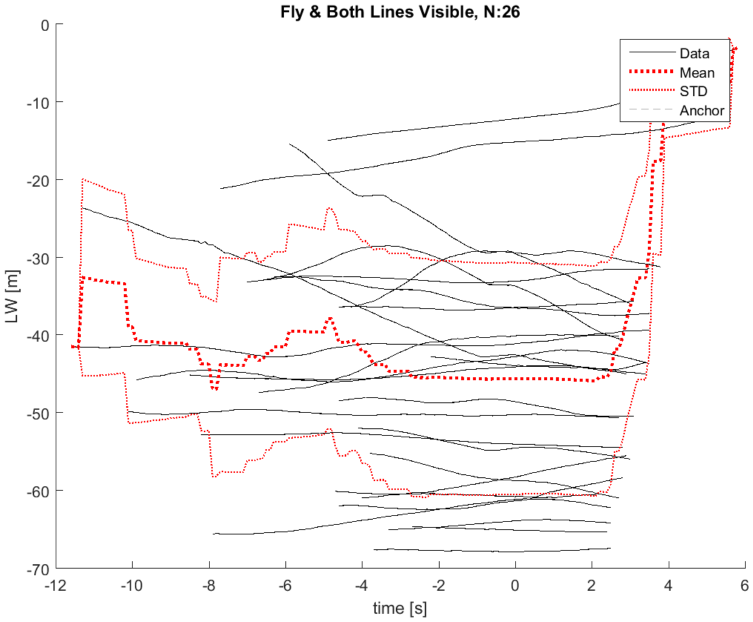


Figure E.32: *Relative velocity to VRU for flying overtaking strategy and both lines visible.*

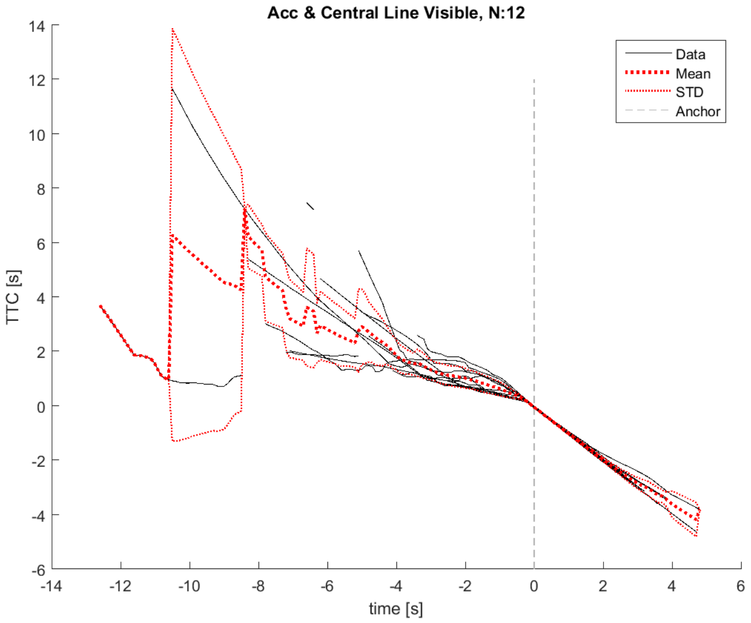


Figure E.33: *Time to collision for accelerative overtaking strategy and only central line visible.*

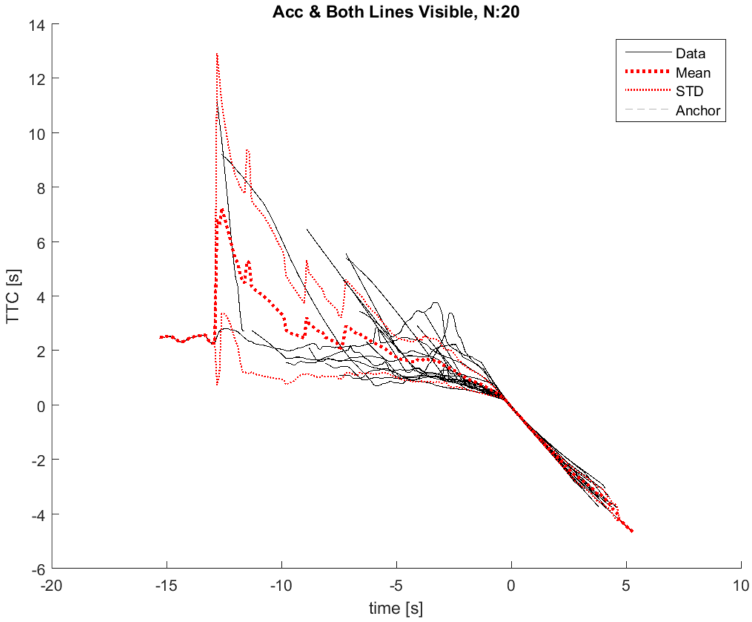


Figure E.34: *Time to collision for accelerative overtaking strategy and both lines visible.*

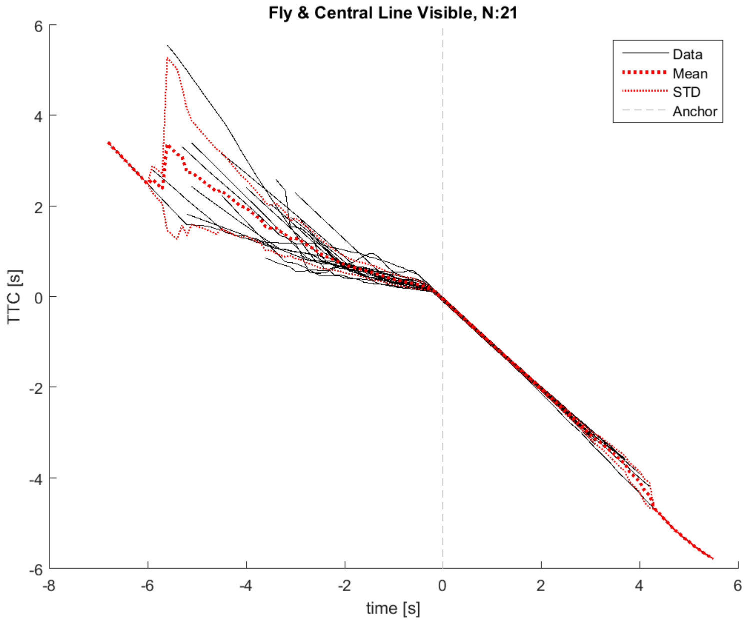


Figure E.35: *Time to collision for flying overtaking strategy and only central line visible.*

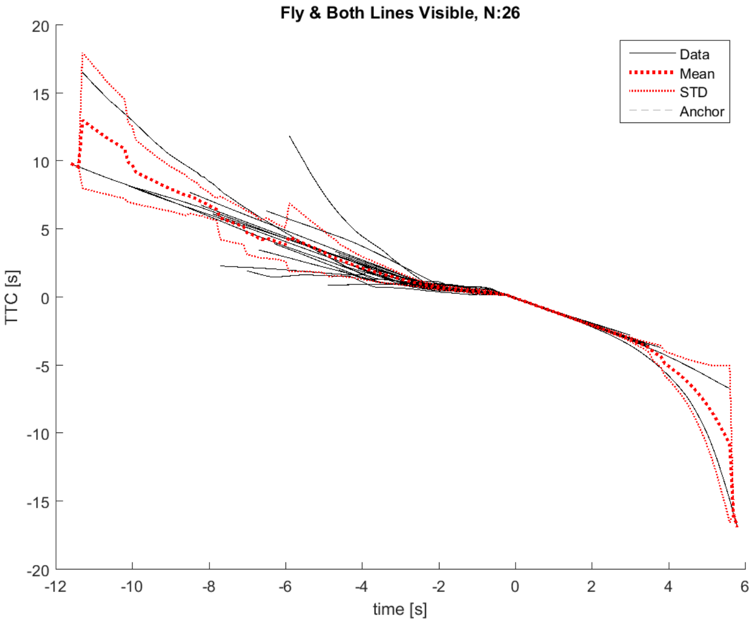


Figure E.36: *Time to collision for flying overtaking strategy and both lines visible.*

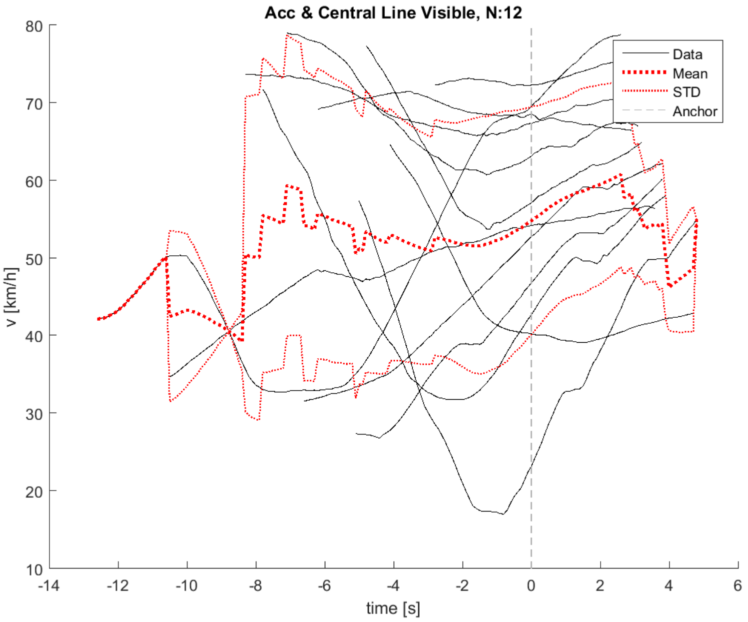


Figure E.37: *Ego vehicle speed for accelerative overtaking strategy and only central line visible.*

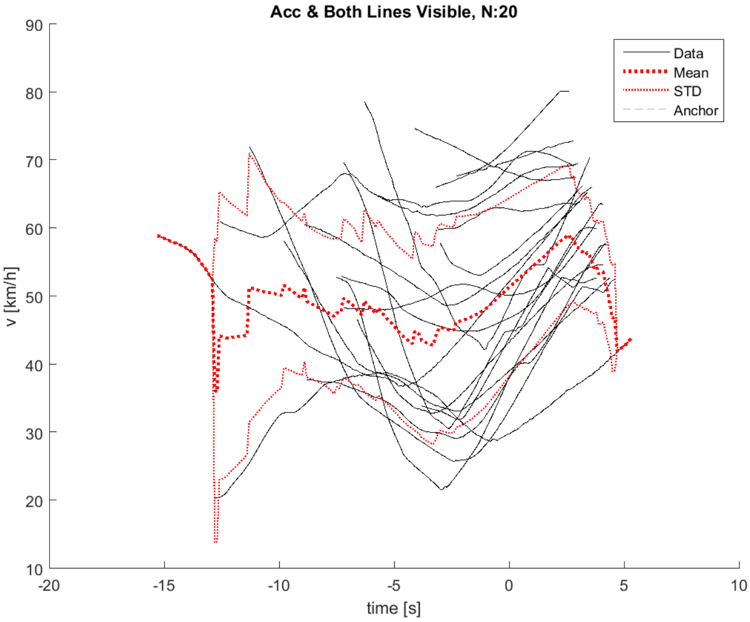


Figure E.38: *Ego vehicle speed for accelerative overtaking strategy and both lines visible.*

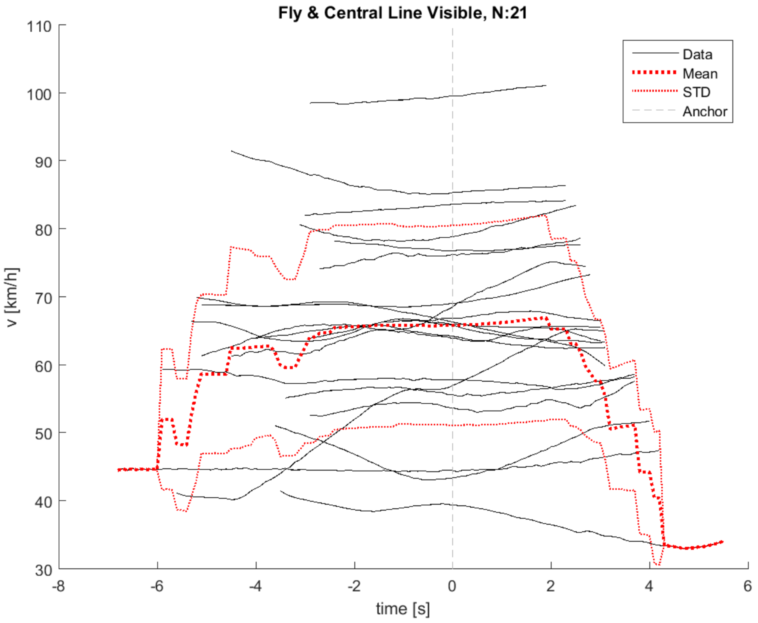


Figure E.39: *Ego vehicle speed for flying overtaking strategy and only central line visible.*

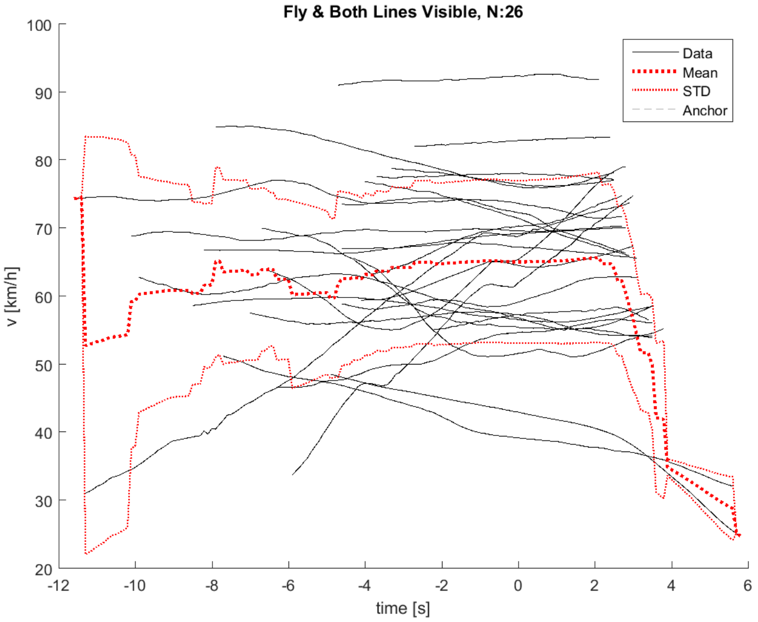


Figure E.40: *Ego vehicle speed for flying overtaking strategy and both lines visible.*