Hannes Geiser, BSc

Analysis of Criteria for Objectifying the Driving Style of Racecar Drivers

MASTER'S THESIS

to achieve the university degree of

Diplom-Ingenieur(in)

Master's degree programme: Mechanical Engineering

submitted to

Graz University of Technology

Supervisor

Assoc.Prof. Dipl.-Ing. Dr.techn., Arno Eichberger

Institute of Automotive Engineering

Julian von Schleinitz MSc, BMW Motorsport

Graz, September 2019

Acknowledgement

I would like to thank BMW Motorsport and especially Michael Graf for giving me the chance and trust to conclude my studies with a master's thesis that handles a topic in this special environment, namely motorsports. Furthermore, I would like to thank my supervisor Julian von Schleinitz for the great support on my work. Whenever I needed some advice, I could go upstairs and he was ready to help out immediately. Not rarely I lost myself into details where he opened up my horizon to have a second thought on the bigger picture. Thank you as well to Arno Eichberger for proofreading this thesis and for the useful inputs in our meetings.

Of course big thanks goes to my family that gave me the chance to go to the university and study the subjects I am interested in. I received lots of support during my studies which I am really thankful for.

Finally, I would like to thank my better half Mara. She supported me through my studies and cheered me up in the difficult phases.

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master‘s thesis.

.....	.....
Date	Signature

Abstract

In motor racing, the behaviour of drivers and their interactions with the vehicle has a big influence on the competitivity of racing teams. By manipulating the vehicle properties to specifically suit the driving behaviour of drivers, the adaptability and the peak performance of the driver-vehicle interactions may be improved significantly. Common approaches for describing driver specific properties rely on empiric knowledge which require lots of experience that is very time extensive to obtain. This thesis treats the problem of objectively describing driving styles of professional race car drivers with driver metrics. The validation of the metrics is done by a driver classification problem using established machine learning algorithms. A method is presented to calculate the theoretical maximum vehicle performance in terms of acceleration based on simulator data. Thereupon, metrics are developed to yield information on the vehicle performance exploitation by the drivers. The correlation to other metrics gives valuable information regarding their performance relevance. The transferability of driver metrics among two racing series with different vehicle classes is analyzed. Some metrics can be successfully transferred, while others deviate among the racing series.

Kurzfassung

Im Motorsport hat der Fahrstil von Rennfahrern und deren Interaktion mit dem Fahrzeug einen großen Einfluss auf die Wettbewerbsfähigkeit von Rennteams. Durch spezielle Anpassung der Fahrzeugeigenschaften auf die spezifischen Ansprüche der Fahrer, kann deren Anpassungsfähigkeit erhöht und mehr Potential vom Fahrzeug ausgenutzt werden. Bestehende Ansätze zur Beschreibung fahrerspezifischer Eigenschaften basieren auf empirischen Erkenntnissen, welche viel Erfahrung voraussetzen. Diese wissenschaftliche Arbeit behandelt das Problem der objektiven Beschreibung des Fahrstils von professionellen Rennfahrern durch Metriken. Um die Metriken zu validieren, wird eine Fahrerklassifizierung unter Verwendung etablierter Algorithmen für maschinelles Lernen durchgeführt. Es wird eine Methode vorgestellt, welche die maximal mögliche Fahrzeugbeschleunigung basierend auf Simulatordaten berechnet. Daraufhin werden Metriken definiert, die Aufschluss über die Ausnutzung der Fahrzeugperformance von den Fahrern geben. Zusätzlich gibt die Korrelationsanalyse mit anderen Metriken eine Aussage über deren Performance Relevanz. Weiters, wird die Übertragbarkeit von Metriken zwischen zwei Rennserien mit verschiedenen Fahrzeugklassen analysiert. Einige Metriken können mit zufriedenstellenden Ergebnissen übertragen werden, während andere Unterschiede zwischen den Rennserien aufweisen.

Contents

Abbreviations

CoG	Center of Gravity
IC	kinematic Instant Center
DIL	Driver-in-the-Loop
GPS	Global Positioning System
RMS	Root Mean Square
DOF	Degrees of Freedom
NMPC	Nonlinear Model Predictive Control
NaN	Not a Number
ANOVA	Analysis of Variance
NMPC	Nonlinear Model Predictive Control

Symbols

Parameters and constants

r_{Thr}	threshold for throttle point detection
$r_{\mathrm{Thr}_{\mathrm{MaxStat}}}$	maximum static value of throttle signal
$r_{\mathrm{Thr}_{\mathrm{MinStat}}}$	minimum static value of throttle signal
$th_{1\mathrm{l}}$	first lower threshold for sector merging
$th_{2\mathrm{l}}$	second lower threshold for sector merging
th_{u}	upper threshold for sector merging
g	gravity acceleration on earth
m	vehicle mass
l_{F}	longitudinal lever from center of gravity to front axle
l_{R}	longitudinal lever from center of gravity to rear axle
br_{dist}	brake pressure distribution
ϵ_1	brake distribution constraint approximation parameter
ϵ_2	brake distribution constraint offset parameter
ω_{lat}	lateral constraint normalization weighting factor
ω_{T}	torque constraint normalization weighting factor
α_{max}	maximum side slip angle constraint
κ_{max}	maximum slip ratio constraint
$T_{\mathrm{Engine}_{\mathrm{max}}}$	maximum engine torque
f_{tol}	stepsize tolerance of the solver
f_{max}	maximum iteration number
con_{tol}	relative constraint violation tolerance
opt_{tol}	first order optimality tolerance
τ	fixed time interval
S	fixed distance
ρ_{crit}	correlation threshold for the initial metric selection
N	number of samples
K	number of groups for the F-value calculation
n	number of values within groups for the F-value calculation
$P_{\mathrm{Engine}_{\mathrm{max}}}$	maximum power output of the powertrain

Variables

α_{F}	side slip angle of front axle

α_{R}	side slip angle of rear axle
β	chassis side slip angle
δ_{f}	steering angle of front axle
δ	steering wheel angle
a_{r}	radial acceleration
a_{x}	longitudinal acceleration
a_{y}	lateral acceleration
R	path radius
κ	path curvature
v	resultant vehicle velocity
v_{x}	longitudinal vehicle velocity
$\dot{\psi}$	yaw angular velocity
s_{tr}	driven trajectory
μ_{β}	mean of the chassis side slip angle
σ_{β}	standard deviation of the chassis side slip angle
$\mu_{\mathrm{e}_{\kappa}}$	mean of the relative error of the approximated trajectory curvature
$\sigma_{\mathrm{e}_{\kappa}}$	standard deviation of the relative error of the approximated trajectory curvature
$\mu_{\mathrm{e}_{\dot{\psi}}}$	mean of the relative error of the approximated yaw rate
$\sigma_{\mathrm{e}_{\dot{\psi}}}$	standard deviation of the relative error of the approximated yaw rate
$\dot{\beta}$	chassis side slip angular velocity
$\ddot{\beta}$	chassis side slip angular acceleration
y	general placeholder for functions
x	general placeholder for variables
e_{x}	error of x
e_{y}	error of y
$e_{\cos\beta}$	error of small angle approximation
$e_{\dot{\beta}}$	error of neglecting
$\tilde{x}$	median of data
$Q1$	first quartile or $25^{\mathrm{th}}\%$ quantile of data
$Q3$	third quartile or $75^{\mathrm{th}}\%$ quantile of data
IQR	interquartile range of data
t	time
t_0	beginning of time interval
t_1	end of time interval
$\ddot{y}$	second derivative with respect to the time of signal y
r_{Thr}	throttle ratio
p_{Brk}	brake pressure
a_{Str}	steering wheel angle
d_{Str}	steering wheel angle rate
a_{StrHp}	highpass filtered steering wheel angle
d_{StrHp}	highpass filtered steering wheel angle rate

y_{lin}	linearized signal y
a_{yExit}	lateral acceleration on the corner exit
f_{s}	samplerate
t_{Lap}	laptime
t_{sector}	time spent in a sector
δ	steering wheel angle
κ_{fl}	front left tire slip ratio
κ_{fr}	front right tire slip ratio
κ_{rl}	rear left tire slip ratio
κ_{rr}	rear right tire slip ratio
Fz_{fl}	front left tire vertical load
Fz_{fr}	front right tire vertical load
Fz_{rl}	rear left tire vertical load
Fz_{rr}	rear right tire vertical load
α_{fl}	front left tire side slip angle
α_{fr}	front right tire side slip angle
α_{rl}	rear left tire side slip angle
α_{rr}	rear right tire side slip angle
μ_{fl}	front left tire dynamic toe angle
μ_{fr}	front right tire dynamic toe angle
μ_{rl}	rear left tire dynamic toe angle
μ_{rr}	rear right tire dynamic toe angle
b_{fl}	front left half track
b_{fr}	front right half track
b_{rl}	rear left half track
b_{rr}	rear right half track
γ_{fl}	front left tire inclination angle
γ_{fr}	front right tire inclination angle
γ_{rl}	rear left tire inclination angle
γ_{rr}	rear right tire inclination angle
D_{x}	longitudinal aerodynamic drag
D_{y}	lateral aerodynamic drag
r_{fl}	front left tire static radius
r_{fr}	front right tire static radius
r_{rl}	rear left tire static radius
r_{rr}	rear right tire static radius
n_{Engine}	engine angular velocity
i_{Tot}	total gearing ratio
vx	longitudinal vehicle velocity
vy	lateral vehicle velocity
${}_{\text{I}}vx_{\text{Tfl}}$	front left tire longitudinal velocity on chassis coordinate system
${}_{\text{I}}vx_{\text{Tfr}}$	front right tire longitudinal velocity on chassis coordinate system
${}_{\text{I}}vx_{\text{Trl}}$	rear left tire longitudinal velocity on chassis coordinate system

${}_{\mathrm{I}}vx_{\mathrm{Trr}}$	rear right tire longitudinal velocity on chassis coordinate system
${}_{\mathrm{I}}vy_{\mathrm{Tfl}}$	front left tire lateral velocity on chassis coordinate system
${}_{\mathrm{I}}vy_{\mathrm{Tfr}}$	front right tire lateral velocity on chassis coordinate system
${}_{\mathrm{I}}vy_{\mathrm{Trl}}$	rear left tire lateral velocity on chassis coordinate system
${}_{\mathrm{I}}vy_{\mathrm{Trr}}$	rear right tire lateral velocity on chassis coordinate system
${}_{\mathrm{T}}vx_{\mathrm{Tfl}}$	front left tire longitudinal velocity on tire coordinate system
${}_{\mathrm{T}}vx_{\mathrm{Tfr}}$	front right tire longitudinal velocity on tire coordinate system
${}_{\mathrm{T}}vx_{\mathrm{Trl}}$	rear left tire longitudinal velocity on tire coordinate system
${}_{\mathrm{T}}vx_{\mathrm{Trr}}$	rear right tire longitudinal velocity on tire coordinate system
${}_{\mathrm{T}}vy_{\mathrm{Tfl}}$	front left tire lateral velocity on tire coordinate system
${}_{\mathrm{T}}vy_{\mathrm{Tfr}}$	front right tire lateral velocity on tire coordinate system
${}_{\mathrm{T}}vy_{\mathrm{Trl}}$	rear left tire lateral velocity on tire coordinate system
${}_{\mathrm{T}}vy_{\mathrm{Trr}}$	rear right tire lateral velocity on tire coordinate system
σ_{y}	lateral tire relaxation length
Δt	stepsize of discretization
Fx_{Ext}	external longitudinal forces
Fy_{Ext}	external lateral forces
Mz_{Tfl}	front left tire aligning moment
Mz_{Tfr}	front right tire aligning moment
Mz_{Trl}	rear left tire aligning moment
Mz_{Trr}	rear right tire aligning moment
$\ddot{\psi}$	yaw angular acceleration
δ_{Tfl}	front left tire steering angle
δ_{Tfr}	front right tire steering angle
δ_{Trl}	rear left tire steering angle
δ_{Trr}	rear right tire steering angle
ω_{tr}	trajectory angular velocity
$\dot{\omega}_{\mathrm{tr}}$	trajectory angular acceleration
T_{fl}	front left tire torque
T_{fr}	front right tire torque
T_{rl}	rear left tire torque
T_{rr}	rear right tire torque
T_{f}	front axle torque
T_{r}	rear axle torque
s	distance
s_{x}	distance in longitudinal direction of the chassis
t	time
ra_{GO}	relative acceleration margin left in the vehicle
a_{GO}	optimized theoretical maximum acceleration
$\Delta a_{\mathrm{GO_{Long}}}$	absolute difference between the optimized maximum acceleration and the acceleration used by the driver in longitudinal direction
$\Delta a_{\mathrm{GO_{Lat}}}$	absolute difference between the optimized maximum acceleration and the acceleration used by the driver in lateral direction

$e_{\mathrm{a_x}}$	error of the longitudinal acceleration induced by the steering error
$e_{\mathrm{a_y}}$	error of the lateral acceleration induced by the steering error
$e_{\ddot{\psi}}$	error of the yaw angular acceleration induced by the steering error
$e_{\delta_{\mathrm{Tfl}}}$	error of the front left tire steering angle
$e_{\delta_{\mathrm{Tfr}}}$	error of the front right tire steering angle
ρ	Pearson correlation factor
ρ_{s}	Spearman correlation factor
μ	mean
σ	standard deviation
rg	rank
Δi_{T}	decrease of an impurity measure
i_{T}	impurity measure
s_{T}	split in decision tree
t_{T}	node in decision tree
g_{i}	Gini index

Vectors

$\mathbf{x}_{\mathrm{in}}$	input state vector from data
$\mathbf{x}_{\mathrm{solv}}$	solver vector
${}_{\mathrm{I}}\mathbf{v}_{\mathrm{as}}$	tire velocity vector on chassis coordinate system
${}_{\mathrm{T}}\mathbf{v}_{\mathrm{as}}$	tire velocity vector on tire coordinate system

Matrices

$\mathbf{T}_{\mathrm{TIas}}$	Transition matrix from chassis coordinate system to the corresponding tire coordinate system
$\mathbf{R}$	correlation matrix

1. Introduction

1.1. Background and Motivation

Since the existence of mankind competitions always played a big role in the society for entertainment purposes of the individuals. The roman empire had gladiator fights, the medieval period used jousting tournaments for entertaining the people and since the beginning of the 20^{th} century, when the automotive industry started to rise, motor racing became a fascinating discipline with machines whose capabilities amazed the people. The more attention was given to motorsport by the society, the more it got attractive for big automotive companies, as it is a very good advertising factor to have the fastest car within given regulations compared to other competitors.

As the budget of companies for motorsport applications started to reach significant amounts, also the effort invested in the research of racing performance rised significantly. Thus, increasing the need for carrying out physical tests to validate the theoretical methods was a necessary factor to achieve the performance goal of lowering the laptimes or increasing the travelled distance within a given amount of time. However, nowadays the number of physical tests is decreasing which leads to an increase of virtual tests carried out in computer simulations. Simulations of vehicles are already well understood. As the computational resources are increasing continuously, the physical properties can be modeled with a high accuracy resulting in more and more sophisticated simulations. However, racing performance is not depending solely on the vehicle, but it is more an interaction of the driver, the vehicle and the environment. The interactions among these systems can be described by the principles of a control loop in control theory where the driver is the controller which gives inputs to the vehicle. The vehicle gives then a certain output to the environment fulfilling the driving task and furthermore, the signal is fed back to the driver and thereby closing the loop [BTTP17]. Thereupon, for improving the performance, besides the vehicle also the driver and the environment have to be considered and especially their interactions with each other. The environment performance factors are given on the properties of the track surface, impeding obstacles (traffic) and air pressure and density, just to name some of them. Even if race track conditions are varying continuously, the variations appear equally for all competitors. Therefore, advantages regarding the environment can only be taken by optimally timing runs, hence, selecting the right tires regarding the conditions or avoiding impeding competitors on qualification runs.

The drivers have a large influence on the performance of the dynamical system *driver-*

vehicle-environment. Knowing the behaviour and characteristics of the drivers, the vehicle can be adapted in such a way that it suits the requirements of them. This results in facilitating the drivers control tasks as they can operate in a more natural way, without having to focus on the driving style adaption to the vehicle. In addition, also the amount of time needed for the adjustment to new conditions is reduced, leading to a maximized performance. However, it is not trivial, to describe the characteristics and properties of the drivers seperately, as they are always closely related to the driven vehicle and environment. The driver characterization usually relies on subjective statements and needs a big amount of experience to choose meaningful and relevant assumptions. This leads to the need of objective descriptions of driver properties that are able to be implemented in the virtual development process, to adapt the vehicle to the driver requirements from the early design stages on.

The scope of this thesis lies in the extension and analyzation of an established work for a specific racing series at BMW Motorsport which describes driving styles of race drivers by objective criteria and proves the method with a driver classification problem. The objective criteria should enable a track independent characterization of driving styles. New objective criteria should be developed as well as racing series specific ones. The transferability of the existing objective criteria to another racing series have to be proven and compared respecting the different vehicle classes. Additionally, the application of encroaching objective criteria should be analyzed.

1.2. State of the Art

In the second half of the 20^{th} century, investigations about the behaviour of drivers while fulfilling the driving task were made by the introduction of first driver models. Fiala et al. [Fia66] contributed to this topic with first scientific publications.

In the last decades, more detailed approaches of driver models gained popularity, building the basis for recently published researches. Rasmussen et al. [Ras83] introduces a three level approach with varying cognitive demandings of each level in a human workflow which ranges from simple everyday repeated tasks over unexpected challenges until rarely occured extraordinary cases. This approach initially was developed for experienced participants with a completed study phase, but established to be applicable on different learning processes as well. Complex unexpected situations which demand actions that were never trained by a participant lead to the level of *knowledge-based* behaviour. This behaviour consists in weighting the different available acting options for the best possible outcome, based on available or still acquirable knowledge. The chosen acting option can then be set as rule for similar cases appearing in the future. In contrast, the *rule-based* behaviour is defined for situations which already occured several times where the participant could derive rules, based on the past outcomes of the chosen actions. Thereupon, the acting option with the subjectively best outcome is retrieved. The third level is defined as *skill-based* behaviour. It consists of knee jerk reactions based

on a stimulation and a resulting automatically, unconsciously performed action which has to be trained in a relative extensive amount of time. This behaviour is the most time effective one and typical for processes which have a given routine. Even a certain capacity for the processing of other not necessarily related secondary tasks is left over. Rasmussen's work relies on the principles of engineering psychology and is related to general human workflows with goal targeted activities.

Donges et al. [Don82] developed a three level approach of the driving task from an engineering point of view. The three defined levels are navigation, guidance and stabilization. The navigational level includes the appropriate route from a given road infrastructure while also taking an approximate time needed for the route into account. Additional information of traffic, incidents or roadworks can be considered. In an unknown environment, the navigation is a process of the consciously choosing actions and is therefore correlated to the knowledge-based behaviour of Rasmussen. The main driving process is defined in the guidance and stabilization level. The guidance consists in deriving a target trajectory and target velocity, considering the planned driving route and the constantly changing conditions of the traffic environment. According to the derived targets, control actions are chosen in an anticipatory manner (open loop) to yield the best possible initial condition with the least deviations from the targets. On the stabilization level the driver has to take corrections on the vehicle controls to stabilize the deviations to the target at an acceptable level. As current states are given as input information for the driver in reference to the target, this level is considered a closed loop behaviour. Figure 1.1 represents the proposed model by Donges.

A common approach to characterize driving behaviour is to classify driving states with machine learning algorithms according to driver inputs. Samiee et al. [SAK+14] detects driver drowsiness states of road car driving simulator data, by using artificial neural networks. He uses image processing-based techniques as well as driver-vehicle interactions to improve the robustness of the method. Arefnezhad et al. [ASEN19] classifies drowsy and awake states using a support vector machine with statistical features of the steering wheel angle and the steering wheel angle rate as inputs. The feature selection is based on a combination of filter and wrapper algorithms using a neuro-fuzzy inference system. Particle swarm optimization is applied to adjust parameters of an adaptive fuzzy system for exploiting the accuracy of the classifier. The method is proved on 20.5h of road car driving simulator data and shows a higher accuracy than recent available algorithms.

Kegelman et al. [KHG17] observes the repeatability of professional race car driver performance by investigating the statistical dispersion of their driven trajectories, based on a public database from vintage race cars. He concludes that similar lap times can be achieved by employing distinct driving styles, as driving at the limit allows a family of solutions in terms of speed and paths.

Segers et al. [Seg14] evaluates driving behaviour of race car drivers, by introducing measures based on the performance, smoothness, response and consistency of the driver input signals, as shown in figure 1.2. The measures are used for simple comparisons among different drivers and to objectively detect differences between them.

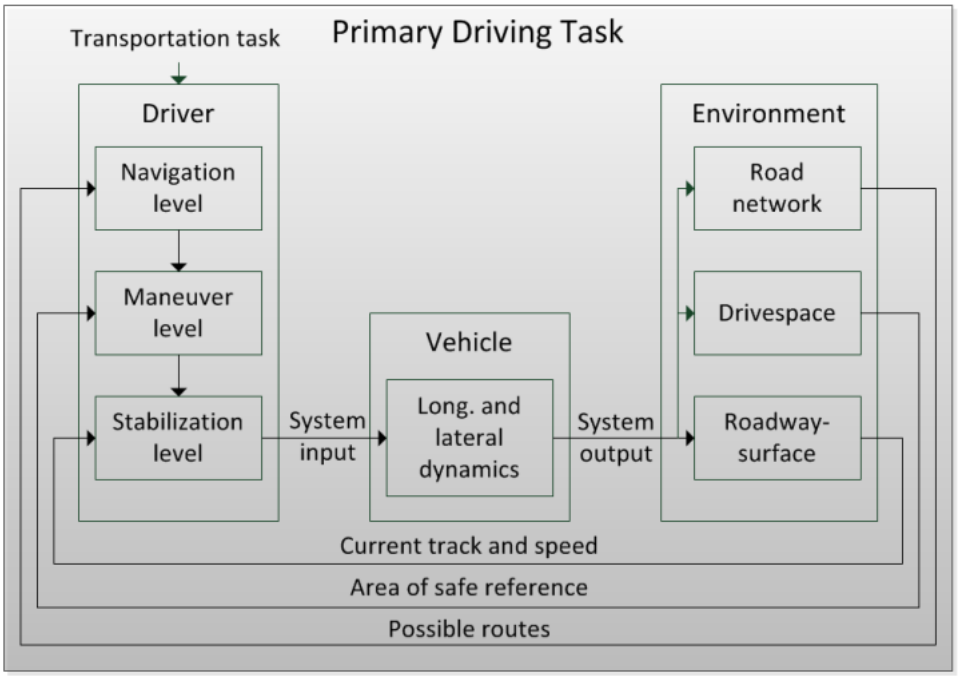


Figure 1.1.: Three level driving approach for the driving task [Don82]

Wörle et al. [WGE18] defines objective criteria of race drivers in professional motorsport, based on their driving control inputs and on their chosen driving trajectories hence, using the principles of the stabilization and guidance level of Donges' method. The algorithm automatically detects cornering maneuvers and uses pattern recognition to find characteristic points within the corners which define the calculation intervals for the objective criteria. To proof the method, objective criteria are supplied as features to a random forest decision tree classifier, resulting in an average driver identification accuracy of 77%.

Similarly, Schleinitz et al. [vSWG+19] derives 14 manually engineered metrics from time series data by using throttle and brake pedal interactions of race drivers in a professional motorsport environment. Supervised learning models in form of multi layered perceptron artificial neural networks and random forest decision trees are used for classifying drivers. With the input data of one corner a classifier accuracy of 77% is reached. Moreover a driver can be detected reliably over different sessions, tracks and seasons.

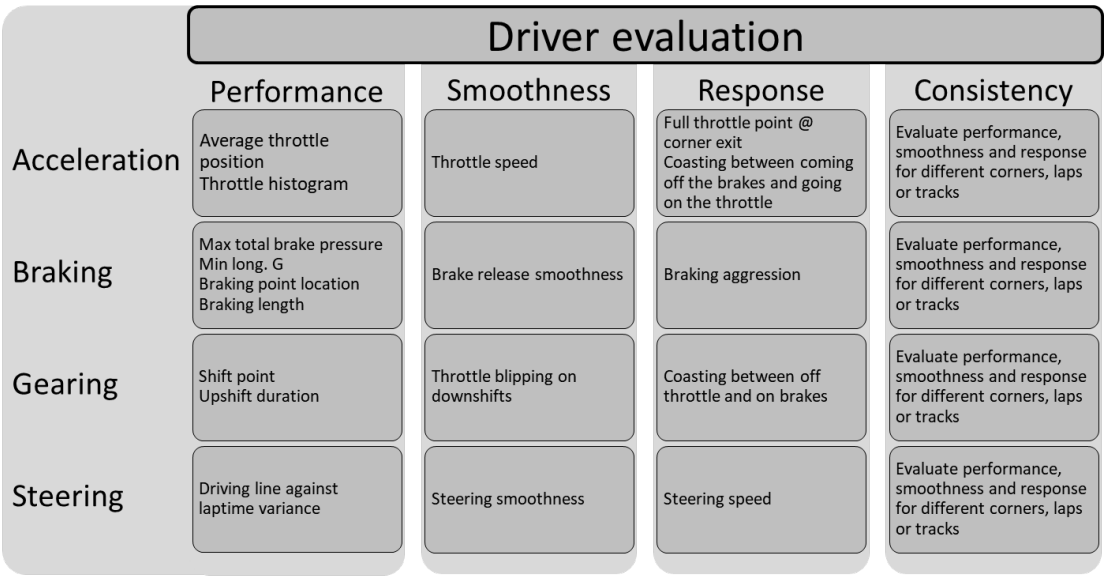


Figure 1.2.: Schematic presentation of the measures used to compare drivers objectively, adapted from [Seg14]

1.3. Overview

This chapter gives an overview of the following topics in this thesis. Four chapters organize the presented contents. First, information of the motivation and task for this work is given with a presentation of state of the art methods relating to similar problem formulations as in this work. In the next chapter, the methodology is presented used to develop results from the given task. The results group an own chapter where objective descriptions of the presented data occurs and the resulting data is discussed and interpreted.

First of all, the data recorded during racing weekends, test events and Driver-in-the-Loop (DIL) simulator sessions is preprocessed to filter out irrelevant information. The presented methods are only valid for drivers operating at their performance limits, therefore, only the fastest laps of sessions during racing events are considered for further processing. The measures of interest are recorded with a common data acquisition system for motor racing application. However, there are some laps with incomplete data channels which are approximated if possible.

Next, objective criteria are developed containing only the most relevant information regarding the description of driving styles, referred to as metrics or features. The calculation interval of the metrics, named sector, plays an important role and is defined for

grip limited corners with applied braking. Grip limited states typically are the phases where the most driving style related properties are requested, as the drivers have to react the vehicle behaviour on its performance limit which is highly nonlinear. The sectors are further divided into characteristic throttle, brake and steering points, to additionally select the calculation interval containing the most driving style related information for each metric. Sectors are detected and numbered subsequently and as the vehicle, environment or driver conditions change, also the sector detection changes. Hence, the sector counting does not always correlate to the appearance of the sector on the track. Thereupon, a sector merging algorithm is applied to reassign numbers to the sectors regarding to their spatial position on the track. Driver behaviours are then compared directly on specific sectors of a race track which highlights the different driving approaches.

Driver metrics are based on the throttle, brake and steering wheel signals as well as on spatial trajectory related measures. Easy interpretable metrics are derived as gradients, distances or duration times of the latter signals in the sector intervals and are the most suitable to describe driving styles with the methods of this work. More abstract metrics are based on statistical quantities of the steering wheel angle and the steering wheel angle rate. Additionally, the statistical calculation methods are also applied for the highpass filtered signals of the steering wheel angle and the steering wheel angle rate, to avoid trajectory related information in these metrics. The filtered metrics show a better performance than the trajectory related measures, but cannot reach the importances of the metrics which are based on pedal gradients. Gear shift or energy management metrics are developed specifically for one racing series each and cannot be applied on the other one. Especially gearshift properties are able to distinct driver behaviour for the corresponding racing series.

To use the advantage that every physical measure of the vehicle in data recorded during DIL simulator sessions is available, a method is presented which determines the theoretical maximum vehicle performance potential in terms of acceleration. On every data sample the absolute vehicle acceleration is maximized using the measured data as starting point and starting direction for an optimizer. Further constraints are used to assure that the vehicle is able to follow the path trajectory and to replicate the powertrain properties with simplifications in the vehicle model. Metrics are then developed based on the acceleration margin to the theoretical maximum absolute vehicle acceleration, unused by the drivers. The metrics based on this method are not the most suitable for characterizing driving styles. Nevertheless, with the acceleration being a performance indicator, the correlation of the latter metrics to other metrics which are difficult to classify regarding their performance relevance is of special interest.

For validation purposes, the driving style characterization problem is introduced to the machine learning environment. Assuming that drivers have their own unique driving styles, a random forest algorithm is used to classify drivers based on the driver metrics. For better classification results and to limit the information amount, only the most relevant metrics in terms of their ability to describe driving styles are kept and used for the classifier. The metrics are first selected according to their correlation with each

other, correlated metrics are dropped subsequently. Of the remaining metrics only the first twenty, regarding to their importance factors derived directly from the random forest algorithm, are kept and used for the classification method.

Driving styles are usually closely related to the driven vehicle, leading to differences for the results of the objective criteria among the different racing series. Therupon, for the application of the method on several racing series with different vehicle classes, the transferability of the metrics is analyzed and a method is developed to normalize the metric values in respect of the vehicle class in the racing series.

2. Methodology

The following chapter describes methods to objectively characterize driving styles in a motor racing environment. Objective driving style criteria are developed and applied on different racing series. To obtain performance relevant driving style criteria, a method is elaborated where the theoretical vehicle exploitation regarding the maximum accelerations by a racing driver is calculated. Another point of interest lies in the transferability of the criteria among different racing series with the occurrence of different vehicle classes. The objective driving style criteria are referred as *metric* or *feature* throughout this work.

Driving style descriptions usually rely on subjective statements. To successfully characterize driving styles, timeseries data is transformed into objective criteria and calculated in predefined sectors as described in detail in section 2.2.

In motorsports, the main goal is to achieve the lowest time needed for covering a given distance or to cover the maximum distance in a fixed time period, respecting given boundaries, defined by specific rules of the race series. These goals correlate to the accelerations reached by the vehicle, hence, accelerations can be seen as performance measures. Thereupon, it makes sense to develop metrics based on the exploitation of the vehicle acceleration as elaborated with the method of section 2.3. Besides the property of giving driving style information, correlations to other metrics are calculated which are not trivial to interpret in terms of their performance relevance.

Assuming that each racing driver has its own distinct driving style, the ability of the objective criteria to describe driving styles is validated by predicting drivers using no other information than the developed criteria. The driver prediction task is approached by machine learning which methods are explained in section 2.4.3.

The expected close relation of the driving styles to the vehicle is suspected for difficulty in transferring driver metrics from one racing series with a specific vehicle class to another racing series with a different vehicle class. In section 2.5, a comparison of the metrics and the driver predictions for different racing series is done and normalization methods are introduced to adapt the metrics for a better suitability in describing driving styles among different racing series. The racing series are referred to as racing series A and racing series B. As a Driver-in-the-Loop (DIL) simulator is used for the drivers to train racing scenarios, additional, valuable driving style information is gathered and used for the method. A transferability analysis of the metrics from the real car data to DIL simulator data is inevitable, though, and done accordingly to gain information on the validity of the additional data in terms of the driving style description.

2.1. Data Preprocessing

The data used in this work were recorded by different sensors mounted on the vehicle during race weekends or testing events which consist of several sessions with several runs for each session. The driver inputs (throttle, brake and steering wheel), vehicle accelerations and the yaw angular velocity are the main measures used for the methods described in this work and usually available for every run of the vehicle. Chassis side slip angles and gearshift or recuperation pedal actuations are further measures of interest, but were not recorded for every racing session. The usage of chassis side slip angle sensors is not allowed during race weekends and is therefore only present for testing sessions. The gearshift or other special pedals are only used for specific racing series, hence, these measures are also not present in every outing. To capture driving trajectory specific information a Global Positioning System (GPS) is used in every session for racing series B and not allowed to use in racing series A, being therefore only available at testing sessions.

Data recorded in DIL simulator sessions has the advantage of always containing all the interested measures. Furthermore, data which in a real car cannot be measured with the current state of technology, are available at DIL simulator sessions as they are calculated by the vehicle model during the simulation.

During the vehicle runs, some recorded laps are marked by comments of special occurrences. The fastest laps within a run are marked as *best lap*. The median of all best laps within a session is defined as a target time. Laps with higher laptimes than 105% and laptimes lower than 95% of the target time are excluded. The thresholds are chosen empirically to exclude non physical low laptimes and laps where the drivers are not competing on their performance limit. For some sessions of racing series B, no best lap markers are available, therefore the target time is defined as the fifth quantile of all uncommented laps with laptimes above an empiric threshold of 40 seconds.

2.1.1. Approximation of Missing signals

Some laps have missing data channels for the path curvature κ and the yaw angular velocity $\dot{\psi}$ which are necessary for the calculation of specific driver metrics. These channels are approximated by other measured quantities and simple kinematic relations from a vehicle model on a given spatial path s_{tr} as visualized in figure 2.1. Furthermore, an error analysis of the approximations is done where the results are presented in section 3.1.1.

The curvature κ, is defined as the reciprocal of the path radius R. The radius is related to the radial acceleration a_{r} and the vehicle velocity v. Assuming small chassis side slip angles β, $\cos\beta$ can be approximated by using the small-angle approximation of (2.2). For side slip angles $|\beta| < 8.1°$ the error is lower than one percent of the approximated value. Using the measured lateral acceleration a_{y} and the measured longitudinal velocity

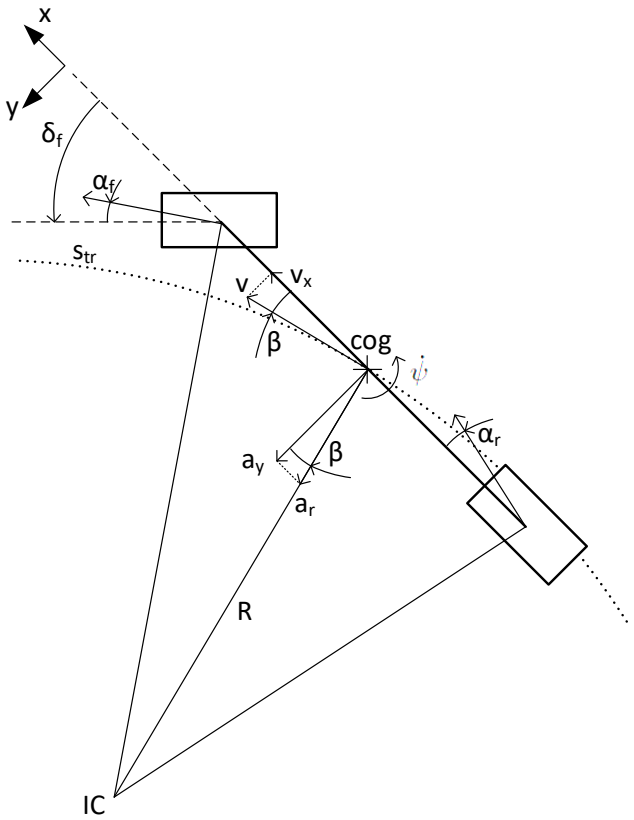


Figure 2.1.: Exemplary single track vehicle model

v_x leads then to (2.3).

$$\kappa = \frac{1}{R} = \frac{a_\mathrm{r}}{v^2} \ , \tag{2.1}$$

$$\text{with } a_\mathrm{y} = a_\mathrm{r} \cdot \cos\beta,$$

$$v_\mathrm{x} = v \cdot \cos\beta \tag{2.2}$$

$$\text{and } \cos\beta \approx 1,$$

$$\kappa \approx \frac{a_\mathrm{y}}{v_\mathrm{x}^2} \tag{2.3}$$

The resulting approximation can be seen in figure 2.2 where the path curvature calculated from known GPS data is compared with the approximated calculation method from a simulator session.

The yaw rate $\dot{\psi}$, as defined by (2.4), can be approximated by assuming the chassis side slip angle rate $\dot{\beta}$ as much smaller than the ratio $\frac{a_\mathrm{r}}{v}$ as follows in (2.5) which then yields

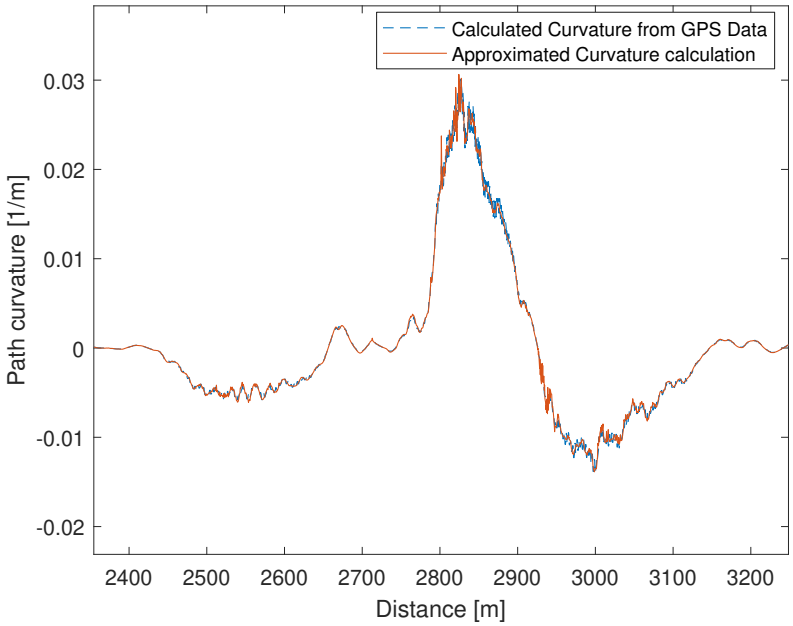


Figure 2.2.: Comparison of curvature calculation methods

(2.6). Figure 2.3 represents a comparison of the measured and the approximated yaw rate in a cornering maneuver.

$$\dot{\psi} = \frac{a_\mathrm{r}}{v} + \dot{\beta}, \tag{2.4}$$

$$\begin{aligned} \text{with } a_\mathrm{y} &= a_\mathrm{r} \cdot \cos\beta \ , \\ v_\mathrm{x} &= v \cdot \cos\beta \\ \text{and } \dot{\beta} &\ll \frac{a_\mathrm{r}}{v}, \end{aligned} \tag{2.5}$$

$$\dot{\psi} \approx \frac{a_y}{v_x} \tag{2.6}$$

By using the latter approximations, a systematic error is introduced. The error propagation gets further analyzed by using a Taylor series expansion. When an arbitrary error e_x is assumed to be small, terms of higher order can be neglected leading to equation (2.7).

$$\begin{aligned} y(x+e_x) &= y(x) + \frac{1}{1!}\frac{\partial y}{\partial x}e_\mathrm{x} + \frac{1}{2!}\frac{\partial^2 y}{\partial x^2}e_\mathrm{x}^2 + \dots \ , \\ e_\mathrm{y} &= y(x+e_\mathrm{x}) - y(x) = \frac{\partial y}{\partial x}e_\mathrm{x} \end{aligned} \tag{2.7}$$

With the errors $e_{\cos\beta} = \cos\beta - 1$ and $e_{\dot{\beta}} = \dot{\beta}$ of the corresponding approximations, the error distribution of the path curvature e_κ and yaw rate $e_{\dot{\psi}}$ within a lap is determined

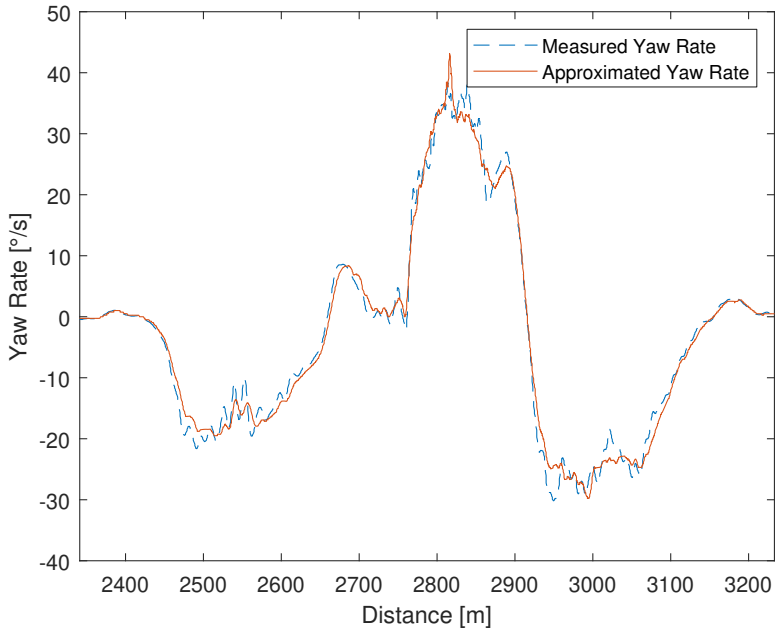


Figure 2.3.: Comparison of measured and calculated yaw rate

by (2.8) and (2.9).

$$\begin{aligned} e_\kappa &= \frac{\partial \kappa}{\partial(\cos\beta)} \cdot e_{\cos\beta}, \\ e_\kappa &= \frac{a_\mathrm{y}}{v_\mathrm{x}^2} \cdot (\cos\beta - 1), \end{aligned} \tag{2.8}$$

$$\begin{aligned} e_{\dot\psi} &= \frac{\partial \dot\psi}{\partial \dot\beta} \cdot e_{\dot\beta}, \\ e_{\dot\psi} &= 1 \cdot \dot\beta. \end{aligned} \tag{2.9}$$

2.2. Metric Definition

In this chapter, the metric definitions for the driver classification used in this work are presented. Before calculating metrics, the metric calculation intervals have to be defined, which are named sectors. The sectors are automatically detected and chosen with the subjective assumption in containing the most valuable information regarding the driving style characterization. Then, the developed metric selection of Wörle et al. [WGE18] gets presented, followed by the metrics of Arefnezhad et al. [ASEN19] and Segers et al. [Seg14]. Lastly, racing series specific metrics are presented relying on data only available for one racing series. All the introduced metrics are used for the metric selection and driver classification methods of section 2.4.3.

2.2.1. Sector Interval Definition

The driver metrics are calculated only in sectors which are considered containing the most valuable information regarding driving style characteristics. Generally, the boundaries of the driver - vehicle - environment system are defined as being limited by the power output of the vehicle powertrain, hence *power limited*, or by the maximum acceleration capabilities of the vehicle - environment interactions, named as *grip limited*. Grip limitations typically occur during cornering events where the tires are the limiting factor in terms of the vehicle accelerations. Grip limited corners can be further divided into occurrences of simple throttle lift-offs without braking or throttle lift offs followed by a brake application in approach to a corner. The sectors are defined as intervals of grip limited corners with applied braking.

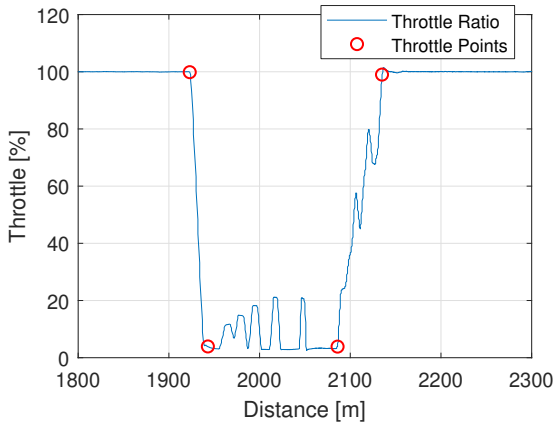
The automatic sector detection algorithm observes the throttle percentage and brake pressure signals. In the first step, extreme values of the signal are calculated as the fifth and the 95^{th} quantile of the signal range. Next, static intervals can be defined where the median of all values that are five percent higher or lower than the respective extremum value are not exceeded. Each detected interval is cut on both sides by 0.5 seconds to avoid short spikes in the signal interval. In addition, intervals smaller than one second are not detected. The maximum static value and the minimum static value is then defined as the median of all values within the respective static interval. Contrary to using absolute values, this approach ignores offset and gain errors that might appear during the data logging. Knowing the static values, characteristic driver input points can be defined within a sector: the *off-throttle* and *no-throttle* points are defined as the last sample satisfying equation (2.10) and (2.11) respectively. Similarly, the *on-throttle* point is the last throttle sample below (2.11) where the *full-throttle* point is the first sample satisfying condition (2.10).

$$r_{\text{Thr}} \geq r_{\text{Thr}_{\text{MaxStat}}} - 0.01 \cdot r_{\text{Thr}_{\text{MaxStat}}}, \tag{2.10}$$

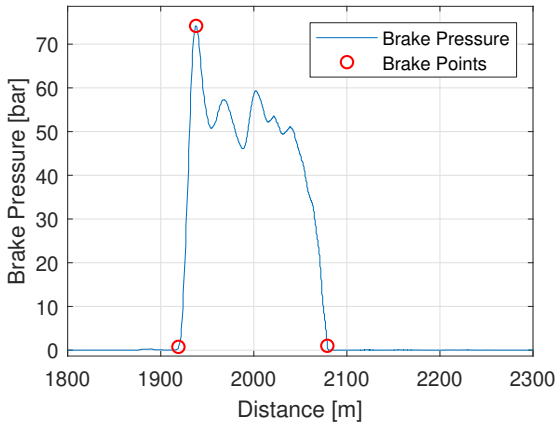
$$r_{\text{Thr}} \leq r_{\text{Thr}_{\text{MinStat}}} + 0.01 \cdot r_{\text{Thr}_{\text{MaxStat}}}. \tag{2.11}$$

By detecting the no-throttle point before the off-throttle point, small lift offs are not detected. Similarly, to avoid the detection of taps on the trottle, the full-throttle point is detected before the on-throttle point. This detection sequence improves the robustness of the sector recognition. Contrary to the throttle signal, no static maximum is present in the brake pressure signal. Therefore, braking events get detected if 15 percent of the 99^{th} quantile of the brake pressure gets exceeded. From thereon the *on-brake*, *max-brake* and *no-brake* points get detected in a similar way as the throttle points. The max-brake point is characterized by the first peak of the signal which often also represents the maximum brake pressure within the sector. There are some sectors though where the max-brake point is followed by a rise of the brake pressure, resulting in a different point, representing the maximum brake pressure within the sector as appears in figure 2.5. This behaviour may already be an indicator of unique driving styles. Characterizing the steering wheel signal needs a different approach than for the throttle and brake pressure

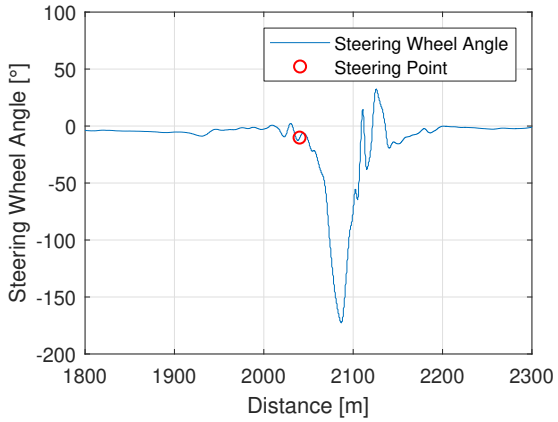
signal. A *turn-in* point gets defined as the point where the integral of the steering wheel signal reaches five percent of its extremum value. The extremum value is defined as the maximum absolute value of the steering wheel angle between the detected throttle points. In order to take into account the different signs of the signal at left and right turns, the turn-in point must have the same sign as the extremum value. By this approach, steering wheel movements induced by the driver in order to stabilize the car (typically in braking phases) are ignored and not mistakenly defined as turn-in points [WGE18]. Figure 2.4 shows exemplary throttle, brake and steering points in the respective signals.



(a) off-throttle, no-throttle, on-throttle and full-throttle points in the throttle signal



(b) on-brake, max-brake and no-brake points in the brake pressure signal



(c) turn-in point in the steering wheel signal

Figure 2.4.: Detected throttle, brake and steering points in the driver input signals

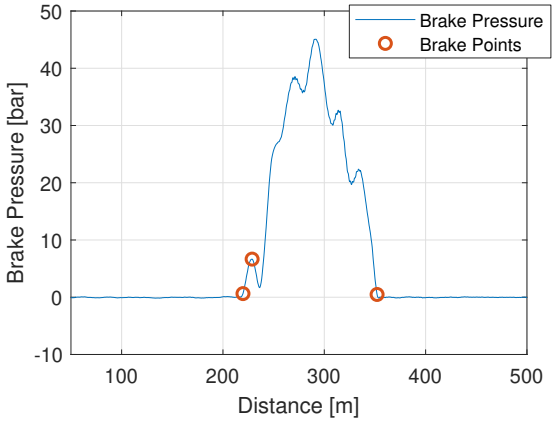


Figure 2.5.: Brake pressure signal with the max-brake point not corresponding to the maximum brake pressure in the sector

2.2.1.1. Sector-merging Algorithm

The sector detection algorithm recognizes sectors successively and might not always detect the same amount sectors on a given race track. This may be explained by different race situations. For example as tires wear out, a power limited corner can turn into a grip limited one or if different drivers have a different amount of recognized sectors, driving style differences and changing track conditions also come into effect. Observing driver metrics on the same sectors leads to driver comparisons without track influences, besides the changing environmental conditions. A sector merging algorithm is introduced to identify and number the sectors according to their appearances on the track from the start-finish line. Figure 2.6 represents boxplots [MTL78] of the successively initial numbered sectors with the corresponding track distances. The median is shown as horizontal red line, the first and third quartiles as a blue box and outliers as red dotted crosses. The amount of outliers indicates that many sectors are not numbered according to their track distances.

First, the sectors are checked for data that should belong to a farther or to a new sector. If datapoints lie closer to the median of the next sector than to the median of the current sector, they are reassigned correspondingly. In the next step, sectors are checked for datapoints below two defined thresholds $th_{1\mathrm{l}}$, $th_{2\mathrm{l}}$ and should therefore define a new sector. The thresholds are chosen empirically in equation (2.12) and (2.13), where $\tilde{x}_{\mathrm{n}}$ represents the median of the n^{th} sector distances, Q_1 the first quartile of sector n and IQR the interquartile range of the data in sector n. For defining a new sector at least ten percent of the datapoints from the mean of the datapoints of the other sectors must satisfy the conditions.

$$th_{1\mathrm{l}} \;=\; \tilde{x}_{\mathrm{n}} \;-\; 0.3 \cdot (\tilde{x}_{\mathrm{n}} - \tilde{x}_{\mathrm{n}-1}) \tag{2.12}$$

$$th_{2\mathrm{l}} \;=\; Q1_{\mathrm{n}} \;-\; 2 \cdot IQR_{\mathrm{n}} \tag{2.13}$$

In addition, datapoints which do not satisfy the conditions for a new sector but still lie closer to the median of the sector before, are moved correspondingly. When moving datapoints from one sector to another, statistical quantities like the first quartile and the interquartile range change hence the algorithm has to be iterated consecutively. For the last two iterations the data is checked again for sectors that could not be recognized from the previous steps. If at least ten percent of datapoints of the mean from the other sectors lie above the upper threshold th_{u}, a new sector is detected. $Q3_{\mathrm{n}}$ in equation (2.14) represents the third quartile of sector n.

$$th_{\mathrm{u}} = Q3_{\mathrm{n}} \;+\; 4 \cdot IQR_{\mathrm{n}} \tag{2.14}$$

After the last iteration, sectors with medians that still lie closer than 50 meters to each other are merged and datapoints farther away than 70 meters of their median are defined as outliers and marked with a negative sector number. Negative sector numbers are ignored for further analyzes. Figure 2.7 shows the resulting sector assignment on the same race track as figure 2.6.

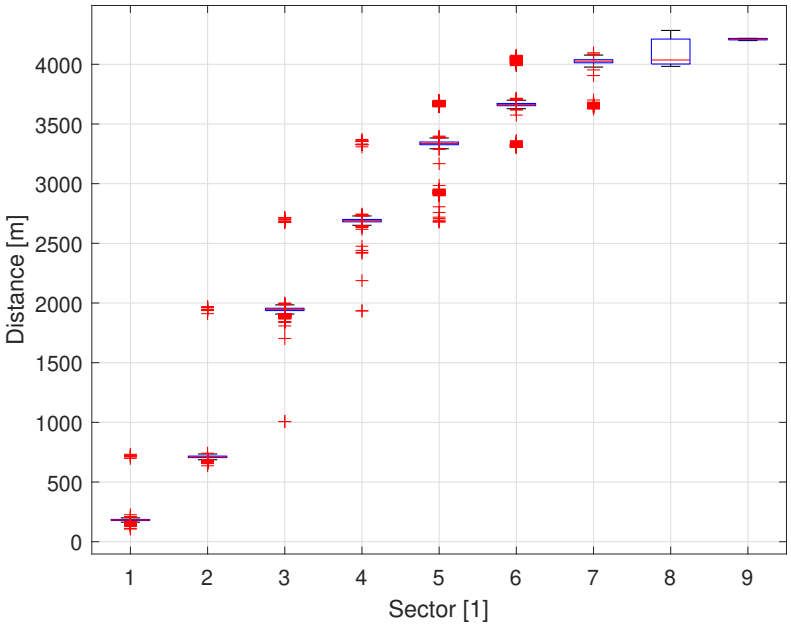


Figure 2.6.: Initial sector counting

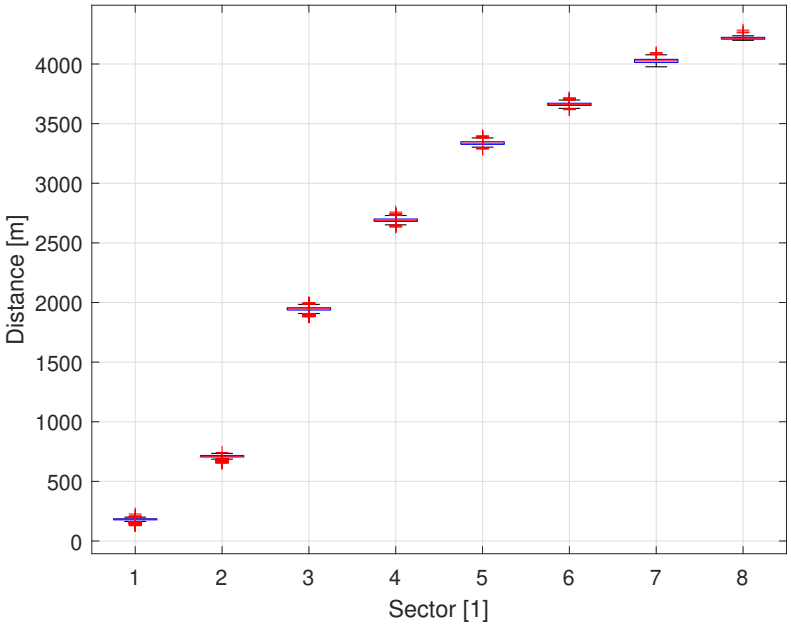


Figure 2.7.: Modified sector numbers

2.2.2. Grip Limited Sector Metrics

The following metrics are based on grip limited corners with the defined throttle, brake and steering points as a reference according to Wörle et al. [WGE18]. As already mentioned, the throttle signal is indexed with the off-throttle, the no-throttle, the on-throttle and the full-throttle points, where in the brake pressure signal the on-brake, the max-brake and the no-brake points are marked. From the steering wheel angle the turn-in point is defined. Table 2.1 lists the metrics selected by Wörle et al. [WGE18]. The linearized gradients are defined as

$$dPdlPhase \;=\; \frac{y(t_1) - y(t_0)}{t_1 - t_0}, \tag{2.15}$$

where y is a placeholder for the throttle percentage r_{Thr} or the brake pressure p_{Brk} signal. t_0 and t_1 represent the beginning and the ending of the corresponding gradient interval. Pdl is a placeholder for the used driver pedal inputs as throttle ratio r_{Thr} or brake pressure p_{Brk}. $Phase$ denotes whether the gradient gets calculated at the pedal activation (On) or deactivation phase (Off).

As the gradient metrics represent the derivative of a linearized throttle action, a measure for describing the validity of the assumed linearity is of interest. The Root Mean Square (RMS) error of the linearity deviation of the throttle and the brake gradients is defined by (2.16).

$$PdlRmseLin \;=\; \sqrt{\frac{1}{t_1 - t_0}\sum_{t=t_0}^{t_1}(y(t) - y_{\mathrm{lin}}(t))^2}, \tag{2.16}$$

with y_{lin} as the linear approximation of the signal according to (2.17):

$$y_{\mathrm{lin}}(t) = y(t_0) + \frac{y(t_1) - y(t_0)}{t_1 - t_0} \cdot t. \tag{2.17}$$

The RMS of the second derivative, $rd2PdlRmse$, is calculated in a similar way with the difference of using the second derivative $\ddot{y}$ of the signal instead of the linearity deviation as written in (2.18).

$$PdlRmseLin \;=\; \sqrt{\frac{1}{t_1 - t_0}\sum_{t=t_0}^{t_1}\ddot{y}^2} \tag{2.18}$$

The distance metrics are denoted with s followed by the corresponding pedal and phase. Finally, a notation is added whether the metric is calculated as an absolute measure or relative to the entire sector distance. Time metrics are named similarly with the exception that instead of the s a t for time is used.

The *V-Angle* metrics $aVCurv$ and $aVApex$ rely on the manoeuvre level of the three level driving task approach [Don82]. The curvature V-Angle $aVCurv$ is defined as the angle between the vectors $V1$ and $V2$. Vector $V1$ represents the trajectory curvature signal

from the turn-in point to the point with the maximum curvature. Whereas Vector $V2$ is defined from the point with the maximum curvature to the full-throttle point, as can be seen in figure 2.8. In an analogous way, the V-Angle of the apex distance $aVApex$ can be defined based on the distance from the vehicle to the inner track limit. The tip of vector $V1$ and the shank of vector $V2$ are then defined as the point with the minimum distance value (figure 2.9).

Additionally, the minimum and maximum vehicle velocities $vCarMin$ and $vCarMax$ are calculated for each sector.

Table 2.1.: Metric selection by Woerle et al.

Metric	Description
drThrOff	Gradient of linearized throttle deactivation
drThrOn	Gradient of linearized throttle activation
rThrRmseLin	Deviation of linearity throttle activation
rd2ThrRmse	Root mean square error of second derivative of throttle deact.
sThrOffRel	Relative no throttle distance
dpBrkOff	Gradient of linearized brake deactivation
dpBrkOn	Gradient of linearized brake activation
pBrkRmseLin	Deviation of linearity brake deactivation
rd2BrkRmse	Root mean square error of second derivative of brake act.
tBrkDlyAbs	Time delay between throttle deact. and brake act.
sBrkRel	Relative braking distance
rd2StrRmse	Root mean square error of second derivative of steering signal
sTurnInRel	Relative distance between off-throttle and turn-in
sTrailRel	Relative distance between turn-in and off-brake
sRollRel	Relative distance between off-brake and on-throttle
cMax	Maximum of absolute curvature
sCurvRel	Relative distance between off-throttle and maximum curvature
aVCurv	V-angle of curvature from turn-in to the maximum curvature and back to the full-throttle point
aVApex	V-angle of distance to inner track limit between turn-in to minimum distance and back to the full-throttle point
xApexMin	Minimum distance to inner track limit
sApexRel	Relative distance between off-throttle and apex
vCarMin	Minimum velocity
vCarMax	Maximum velocity

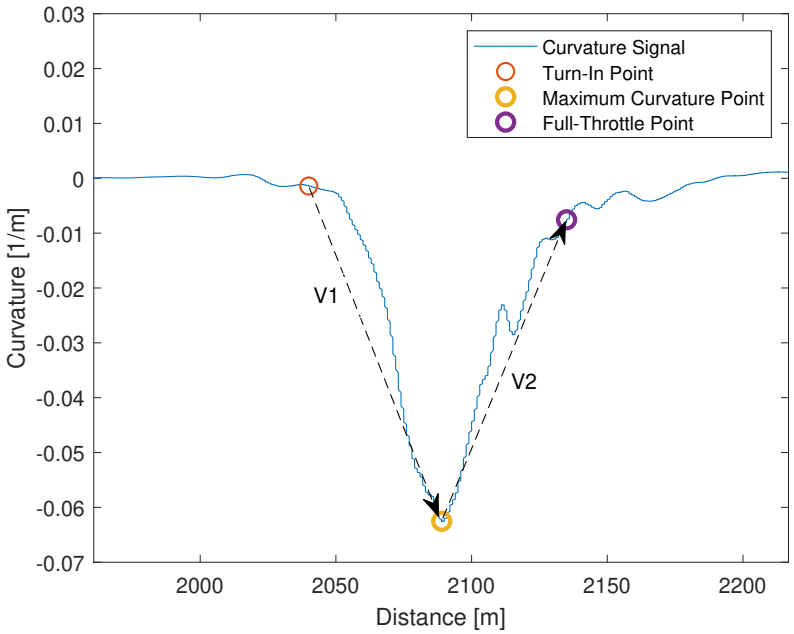


Figure 2.8.: Vectors V1 and V2 for the V-angle calculation; shown on the curvature V-angle

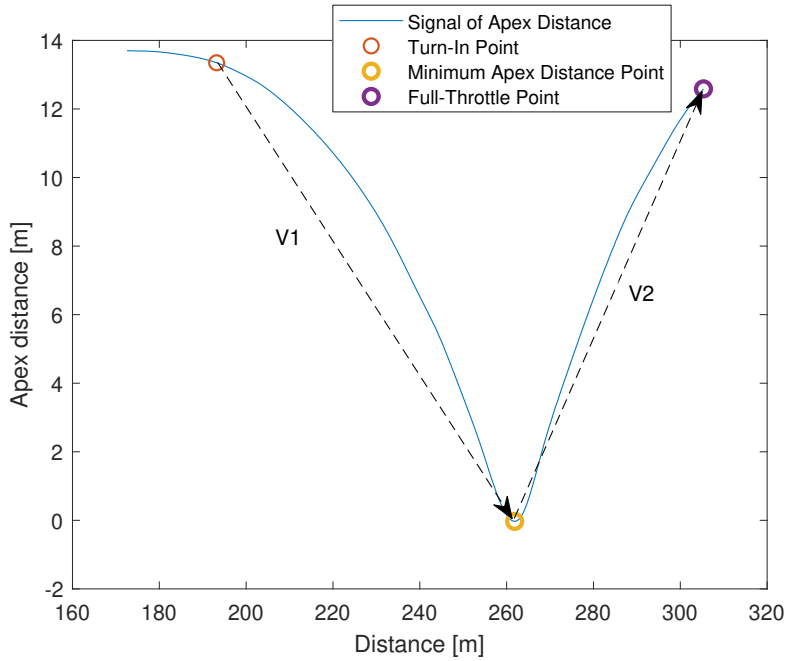


Figure 2.9.: Vectors V1 and V2 for the V-angle calculation; shown on the apex V-angle

2.2.3. Statistical Metrics

The steering behaviour of drivers contains useful information regarding driving styles. To capture as much information as possible metrics relying on statistical theory are defined, based on previous works on driving state detection which were used for road vehicles [HHRH11][SAK+14][FY10]. In contrast to the grip limited sector metrics (section 2.2.2), the statistical metrics are not calculated between specific throttle, brake or steering points, but within entire sectors as defined in 2.2.1.1. Table 2.2 lists the introduced statistical metrics where y acts as a placeholder for the steering wheel angle a_{Str}, the steering wheel angle rate d_{Str} or the road curvature filtered signals of both quantities a_{StrHp}, d_{StrHp}. The road curvature is filtered out of the steering wheel signal by applying a highpass filter. More precisely, a *finite impulse response* highpass filter is used with a stopband frequency of 0.3Hz, a passband frequency of 2Hz and a passband ripple of 0.1Hz. The highpass filter properties are defined empirically to remove the road curvature while still keeping the stabilization steering wheel inputs of the driver. Figure 2.10 shows the unfiltered and the filtered signal, respectively. The pronounced extremum of the unfiltered curvature signal shows the process of steering through a corner where in the filtered signal only oscillations derived from the stabilization inputs of the driver can be noticed.

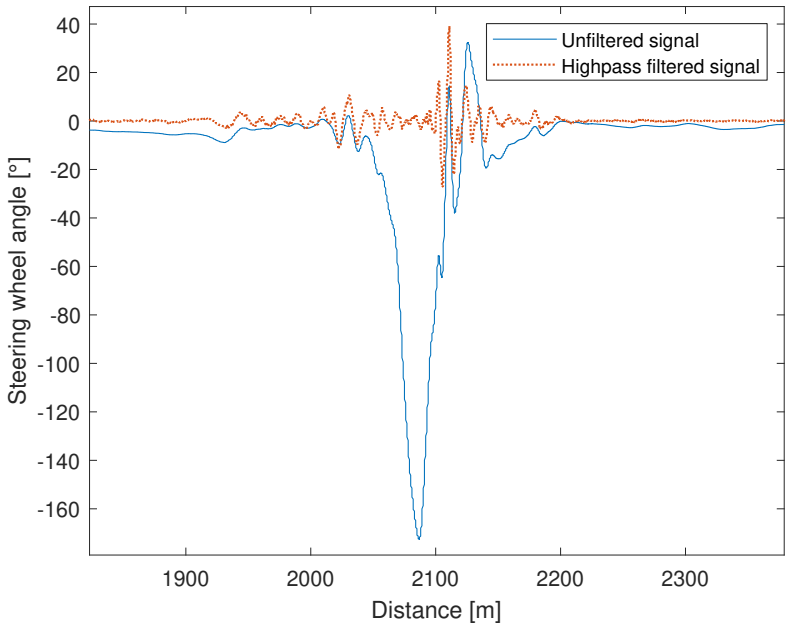


Figure 2.10.: Highpass filtered steering wheel angle signal compared to the raw signal

Table 2.2.: Statistical metrics

Metric		Description
	Time domain	
yAvg	Average	Mean value of signal
yRange	Range	Difference between minimum and maximum of signal
yStd	Standard deviation	Dispersion of the data around mean value
yEnergy	Energy	Sum of the square of signal magnitude
yZCR	Zero Crossing Rate	Number of signal direction changes per second
y1Quart	First Quartile	25^{th} quantile of signal
yMed	Median (second quartile) of signal	50^{th} quantile of signal
y3Quart	Third quartile of signal	75^{th} quantile of signal
yKFD	Katz Fractal Dimension	Index for characterizing fractal patterns or sets by quantifying their complexity as a ratio of the change in detail to the change in scale
ySkwn	Skewness of signal	Measure for signal similarity
yKurt	Kurtosis of signal	Measure of tailedness of the probability distribution of a random variable
yAppEnt	Approximate entropy	Complexity of signal in time domain based on distance in embedding dimension
ySamEnt	Sample entropy	Conceptually similar to Approximate entropy, but less dependend on the datasize
yShannEn	Shannon Entropy	Complexity of signal in time domain based on probability function
yRenEn	Renyi entropy	Generalization of Shannon entropy
yMob	Mobility parameter	Proportion of standard deviation of the power spectrum
yComp	Complexity parameter	Represents the change in frequency
	Frequency domain	
yFreqVar	Frequency variability	Variance of the frequency in the defined frequency band
ySpecEn	Spectral entropy	Complexity of signal in frequency domain
yCogF	Spectral center of gravity	Spectral centroid of the signal
yDomFreq	Dominating frequency	The frequency that has maximum value of the Power Spectral Density (PSD)
yPSDMean	Average PSD	Mean value of PSD
yPSDMin	Min PSD	Minimum value of PSD
yPSDMax	Max PSD	Maximum value of PSD

2.2.4. Metrics Defined for a Full Lap

Similar to the statistical metrics, the metrics defined by Segers et al. [Seg14] do not correspond to specific throttle, brake or steering points. Originally, they were calculated in the intervall of a whole lap but can easily be applied on the grip limited sector intervals. The metric definitions are summed up in table 2.3, where r_{Thr}, p_{Brk} and a_{Str} represent the throttle position, the combined brake pressure of the front and rear axle as well as the steering wheel angle, respectively. Some metrics share the same intention as the metrics of Wörle et al. but are defined for different intervals, namely $drThrSp$, $dpBrkSp$, $tCoast$, $tCrOver$. Further metrics are calculated as the means of only the positive or negative derivative values ($dpBrkAgg$, $dpBrkRls$) and generally many definitions rely on thresholds which have to be satisfied for the activation of the metric calculation. This approach may lead to different results as calculating the metrics between specific pedal points. For example the coast time metric $tCoast$ shares a similar intention as $sRollRel$, but gets activated at throttle values below ten percent of the 99^{th} quantile of the signal. Thus, even when the driver is still slightly applying the pedal, $tCoast$ is active where the definition of *coasting* would imply no pedal application. In contrary, the relative rolling distance $sRollRel$ is calculated as the distance between the no-brake and the on-throttle point relative to the sector length. Based on the definition of these points this interval assures no pedal application of the drivers. However, some drivers tend to rest their foot in coasting phases on the brake pedal which results in missing information of the $sRollRel$ metric. The end of a grip limited sector is defined with the full-throttle state. Thereupon the time spent at full-throttle $tThrFull$ will be always zero in such intervals and the full-throttle time is only used for the comparison of different feature extraction methods of section 2.6 and not considered for further applications in grip limited sectors.

Table 2.3.: Metrics defined by applying Segers et al.

Metric	Description
drThrSp	$\text{mean}\left(\frac{dr_{\text{Thr}}}{dt}\right)$
drThrAgg	$\text{mean}\left(\frac{dr_{\text{Thr}}}{dt}\right) \quad \forall\ r_{\text{Thr}} > 25\%$ and $a_{\text{x}} > 0.2\text{g}$
rThrAccept	$\frac{a_{\text{yExit}}}{\max(a_{\text{y}})}$, with $a_{\text{yExit}} = a_{\text{y}}(r_{\text{Thr}} = 100\%)$
rThrMean	$\text{mean}(r_{\text{Thr}})$
gThrEff	$\text{mean}\left(\frac{a_{\text{x}}}{r_{Thr}}\right), \quad \forall\ a_{\text{x}} > 0.2\text{g}$
tThrFull	$\sum_{t=t_0}^{t_1} \frac{1}{f_{\text{s}}} \cdot \frac{1}{t_{\text{Lap}}}, \quad \forall\ y > 95\%$
rThrInt	$\sum_{t=t_0}^{t_1} r_{\text{Thr}} \cdot \frac{1}{f_{\text{s}}}$
gBrkMax	$\min(a_{\text{x}})$
pBrkMax	$\max(p_{\text{Brk}})$
dpBrkSp	$\text{mean}\left(\frac{dp_{\text{Brk}}}{dt}\right)$
dpBrkAgg	$\text{mean}\left(\frac{dp_{\text{Brk}}}{dt}\right), \quad \forall\ \frac{dp_{\text{Brk}}}{dt} > 0$
pBrkInt	$\sum_{t=t_0}^{t_1} p_{\text{Brk}} \cdot \frac{1}{f_{\text{s}}}$
gBrkEff	$\text{mean}\left(\frac{a_{\text{x}}}{p_{\text{Brk}}}\right), \quad \forall\ a_{\text{x}} < -0.2\text{g}$
dpBrkRls	$\text{mean}\left(\frac{dp_{\text{Brk}}}{dt}\right), \quad \forall\ \frac{dp_{\text{Brk}}}{dt} < 0$
tBrkOn	$\sum_{t=t_0}^{t_1} \frac{1}{f_{\text{s}}} \cdot \frac{1}{t_{\text{sector}}}, \quad \forall\ p_{\text{Brk}} > 5\text{bar}$
nStrSmooth	$\text{mean}\left(\frac{da_{\text{Str}}}{dt}\right)$
aStrMax	$\max(\lvert a_{\text{Str}} \rvert)$
tStr	$\sum_{t=t_0}^{t_1} \frac{1}{f_{\text{s}}} \cdot \frac{1}{t_{\text{sector}}}, \quad \forall\ a_{\text{Str}} > 10^\circ$
aStrInt	$\sum_{t=t_0}^{t_1} \lvert a_{\text{Str}} \rvert \frac{1}{f_{\text{s}}} \cdot \frac{1}{t_{\text{sector}}}$
cMean	$\text{mean}(\lvert \kappa \rvert)$
tCoast	$\sum_{t=t_0}^{t_1} \frac{1}{f_{\text{s}}} \cdot \frac{1}{t_{\text{sector}}}, \quad \forall\ r_{\text{Thr}} < 10\%$ and $p_{\text{Brk}} < 5\text{bar}$
tCrOver	$\sum_{t=t_0}^{t_1} \frac{1}{f_{\text{s}}} \cdot \frac{1}{t_{\text{sector}}}, \quad \forall\ r_{\text{Thr}} > 5\%$ and $p_{\text{Brk}} > 5\text{bar}$

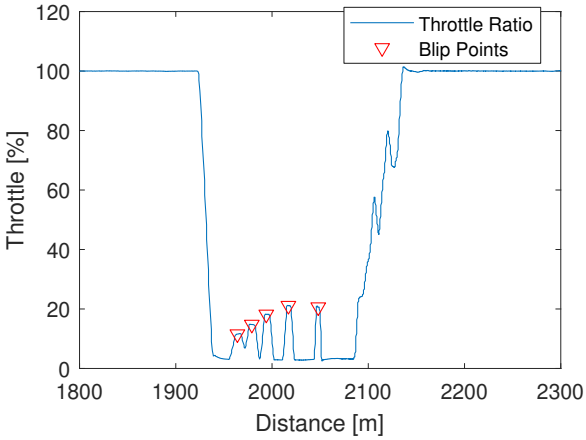
2.2.5. Shiftmetrics

For racing Series A, a set of metrics listed in table 2.4 can be introduced which rely on gear shift applications. The first category $rdShft$ is based on the absolute and relative engine speed when the driver executes a downshift. In order to not compare engine mappings with different engine speed limiters, the relative measure is calculated by dividing the absolute value by the maximum engine speed per lap. The maximum engine speed driven in a lap may not always correspond to the engine speed limiter value but is assumed as a suitable reference. Time based shift metrics $tdShft$ are defined as the absolute and relative time between two downshifts. The reference is the time spent in the respective sector. To ease the downshifting process for the powertrain, throttle blips[1] are applied automatically during the gear downshifting phase. With this technique the engine speed is matched better to the shifted lower gear which results in a smoother engine drag torque curve thus, a more predictible behaviour of the vehicle for the drivers. Without applying throttle blips the rotational inertia of the powertrain parts lead to sudden increases of the drag torque which may influence the stability of the vehicle. To analyze the brake pedal behaviour at downshift throttle blips, the gradient of the front brake pressure is observed at peaks in the throttle signal between the on-brake and the off-brake points named as $dpBrkBlip$. Figure 2.11 shows a grip limited sector with applied throttle blips on the downshifts. Racing series B does not use gear shifts, therefore the metrics introduced in this chapter cannot be applied on racing series B.

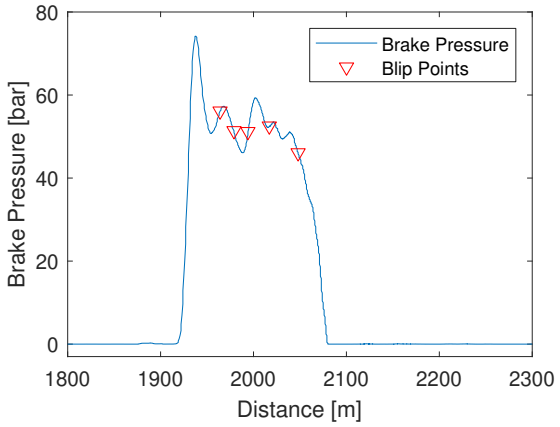
[1] Short throttle actuations during downshifting to synchronize the engine speed with the gearbox speed.

Table 2.4.: Shiftmetrics based on racing series A

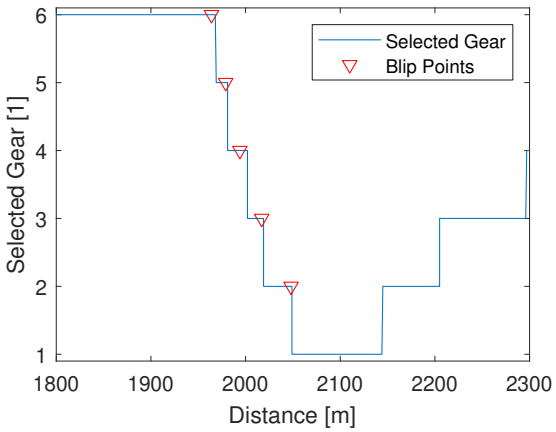
Metric	Description
rdShft$i$Abs	Engine revolutions per minute (rpm) at the point when the driver applies the downshift paddle in order to shift to the indicated gear; with $i = 1...5$
rdShftAbs	Mean of the downshift engine rpm of every gear
rdShft$i$Rel	Engine rpm at the point when the driver applies the downshift paddle in order to shift to the indicated gear relative to the maximum engine rpm of one lap; with $i = 1...5$
rdShftRel	Mean of the relative downshift engine rpm of every gear
tdShft$i$Abs	Time spent in the indicated gear between two downshifts; with $i = 2...5$
tdShftAbs	Mean of the downshift time for every gear
tdShft$i$Rel	Time spent in the indicated gear between two downshifts relative to the time spent in the sector; with $i = 2...5$
tdShftRel	Mean of the relative downshift time for every gear
dpBrkBlip	Mean of brake pressure gradient at throttle blip points



(a) Throttle blip points in the throttle signal



(b) Brake pressure with marked throttle blip points



(c) Selected gear with marked throttle blip points

Figure 2.11.: Downshifting process with marked throttle blip points

2.2.6. Energy Management Metrics

In racing series B, energy saving while still maintaining fast laptimes is a key factor for performance. *Lift and coast* is a technique where the driver does not brake immediately when he lifts off the throttle approaching a corner, but lets the car coast until the brake point is reached. Applying lift and coast at the end of long straights, generally is the most efficient way to save energy while still maintaining low laptimes, as the time loss is smaller if a high vehicle velocity is already present. For further energy saving, the drivers have the ability to apply a recuperation pedal to decelerate the car while recuperating the kinetic energy. The metrics presented in table 2.5 are defined as the absolute and relative time differences of the of the recuperation pedal or the off-throttle application points compared to all the laps. These driver metrics cannot be applied on racing series A.

Table 2.5.: Coast and recuperation metrics for racing series B

Metric	Description
tRgnDlyAbs	Absolute time differences between recuperation pedal activation points
tRgnDlyRel	Relative time differences between recuperation pedal activation points with the sector time as reference
tCstDlyAbs	Absolute time differences between off-throttle points
tCstDlyRel	Relative time differences between off-throttle points with the sector time as reference

2.3. Grip Optimization Algorithm

In this chapter, a method is introduced to investigate driving style characteristics on the stabilization level of the three level driving task approach [Don82]. The goal is to get information of the acceleration margin to the maximum capability of the vehicle. Driver metrics are developed with this method and additionally, correlations to the other metrics are investigated to observe the performance relevance of each metric.

The acceleration is defined as the second derivative of the driven distance by the time. If a driver is capable of extracting more acceleration out of the car in the same direction than another driver on the same spatial path and initial velocity, he will necessarily achieve faster laptimes. Therefore the acceleration margin left can be seen as a direct performance indicator.

Dal Bianco et al. [DLG18] formulated a laptime simulation based on a vehicle from the *GP2* racing series with a very sophisticated vehicle model including highly dynamical effects. Besides the longitudinal, lateral and yaw Degrees of Freedom (DOF) used for the spatial information of the vehicle, tire vertical, chassis pitch, roll and heave DOF are included as well for the calculation of tire loads with high precision. The driving force is modelled with a single control variable imitating the throttle and brake pedal actuation by the driver which determines the torque distribution of the tires. Respecting the dynamic angular behaviour of the wheels, the traction forces can be calculated from the tire torques hence, additional four DOF were introduced. Summarized, 14 DOF were included into the vehicle model, seven for the dynamic tire loads, four for the wheel angular motion and three for the spatial information of the vehicle. Additionally, the kinematic relationships between the unsuspended and the suspended mass induced by the suspension is implemented as well as an aerodynamic map containing the force distribution generated by aerodynamic effects depending on the front and rear rideheights. The laptime simulation is formulated as an *indirect optimal control problem* and solved with the *Pins* software which generally is faster than nonlinear programming based solvers [BBB16].

Kehrle et al. [KJKS11] developed a method to optimize the gear selection of the vehicle in real time to provide gear recommendations for the driver in a driving simulator. The method is based on *Nonlinear Model Predictive Control* (NMPC) which relies on optimal control theory principles. An optimal control problem is formulated within a limited time horizon to keep the computational effort relatively small which is necessary to perform it in real time. The vehicle is modelled with three DoF describing the spatial properties of the system. This method considers the influence of the gear selection at the driving trajectory as the steering wheel angular velocity, the brake pedal, the throttle pedal and the gear selection are defined as control inputs.

A laptime simulation is not suitable for the scope of developing driver metrics based on the exploitation of the vehicle performance by the drivers, as all the spatial discretization points are solved in a combined manner for a minimum laptime. This leads to a mathematically best solution without considering the driving characteristics of the drivers. NMPC is considering the driver behaviour up to a certain point, as the simulation starting point is based on the driver data derived from DIL simulator sessions, but the simulation optimization space is considering the near future of the driver without optimizing the initial state. The method introduced in this chapter focusses on the optimization of data, choosing the data of the drivers as starting point and treating every sample as an independent optimization problem. Constraints are used to assure the driver intentions and to not loose driving style related information, as for example the driving trajectory is considered in the optimization.

2.3.1. Algorithm

A two track vehicle model is used to calculate the longitudinal and lateral accelerations of the vehicle. The input state vector $\mathbf{x}_{\text{in}}$ of equation (2.19) for the vehicle model comes from simulator sessions where every needed physical quantity is available. The steering angle δ, bodyslip angle β and the slip ratio κ_{as} (suffix $a \in r, f$ stands for rear (r) or front (f) axle, while the suffix $s \in r, l$ stands for right (r) or left (l) side) of the four tires define the solver vector $\mathbf{x}_{\text{solv}}$ presented in equation (2.3.1). These variables are varied by an optimizer on every spatial sample of the recorded data with the objective to maximize the norm of the lateral and longitudinal accelerations. Figure 2.12 illustrates the sequence of the algorithm. The direction of the resulting acceleration vector and the angular momentum are held constant during the optimization in order to restrict the search space of the solver and to assure the ability of the vehicle to follow the given trajectory curvature. For simplicity reasons the slip ratios are optimized directly, instead of modeling the powertrain and defining the throttle input as optimization variable. The input state vector contains the following signals

$$\begin{aligned}\mathbf{x}_{\text{in}} = [&\dot{\psi}\ v\ Fz_{\text{fl}}\ Fz_{\text{fr}}\ Fz_{\text{rl}}\ Fz_{\text{rr}}\ \alpha_{\text{fl}_{\text{n}-1}}\ \alpha_{\text{fr}_{\text{n}-1}}\ \alpha_{\text{rl}_{\text{n}-1}}\ \alpha_{\text{rr}_{\text{n}-1}} \\ &\mu_{\text{fl}}\ \mu_{\text{fr}}\ \mu_{\text{rl}}\ \mu_{\text{rr}}\ b_{\text{fl}}\ b_{\text{fr}}\ b_{\text{rl}}\ b_{\text{rr}}\ \gamma_{\text{fl}}\ \gamma_{\text{fr}}\ \gamma_{\text{rl}}\ \gamma_{\text{rr}}\ Dx\ Dy \\ &r_{\text{fl}}\ r_{\text{fr}}\ r_{\text{rl}}\ r_{\text{rr}}\ n_{\text{Engine}}\ i_{\text{Tot}}\ \delta_0\ \beta_0\ \kappa_{\text{fl}_0}\ \kappa_{\text{fr}_0}\ \kappa_{\text{rl}_0}\ \kappa_{\text{rr}_0}]^{\text{T}}\ ,\end{aligned} \tag{2.19}$$

$$\mathbf{x}_{\text{solv}} = [\delta\ \beta\ \kappa_{\text{fl}}\ \kappa_{\text{fr}}\ \kappa_{\text{rl}}\ \kappa_{\text{rr}}]^{\text{T}}\ ,$$

where $\dot{\psi}$ is the angular velocity, v the resultant velocity of the vehicle, Fz_{as} the tire loads, $\alpha_{\text{as}_{\text{n}-1}}$ the side slip angles of the previous sample $n-1$, μ_{as} the dynamic toe angles, b_{as} the half tracks, γ_{as} the camber angles, Dx and Dy the aerodynamic drag forces, r_{as} the tire radii, n_{Engine} the angular velocity of the engine, i_{Tot} the overall gearing from engine to tire, δ_0 the initial steering wheel angle, β_0 the initial bodyslip angle and $\kappa_{\text{as}0}$ are the initial slip ratios of the tires.

The optimization process can be illustrated in an exemplary Miliken moment diagram [MM95] shown in figure 2.13. Miliken moment diagrams are created by letting the chassis side slip angle constant and calculating the yaw moment and the lateral acceleration resulting from the tire forces for varying steering wheel angles. Similarly the lines for a constant steering wheel angle and varying chassis side slip angles are calculated. The diagram is valid for a constant longitudinal acceleration. In the optimization process the lateral acceleration moves from an initial state driven in the simulator to the boundary of the miliken moment diagram on a horizontal line (due to the constant yaw angular acceleration), while increasing the norm of the longitudinal acceleration (due to the constant acceleration ratio) by varying the slip ratios and thus, the longitudinal tire forces.

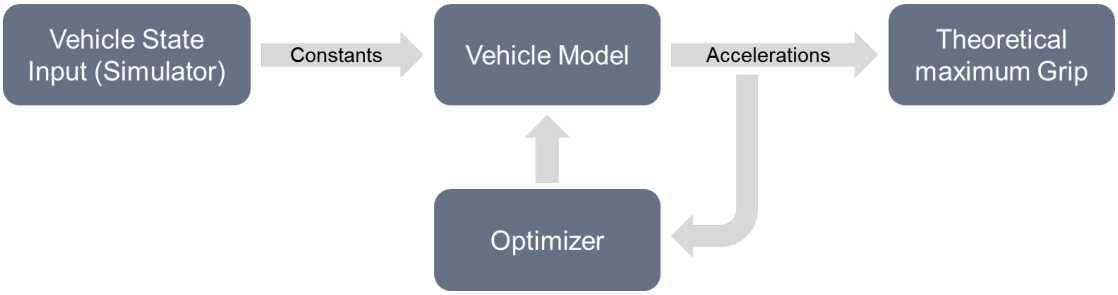


Figure 2.12.: Flowchart of the grip optimization method

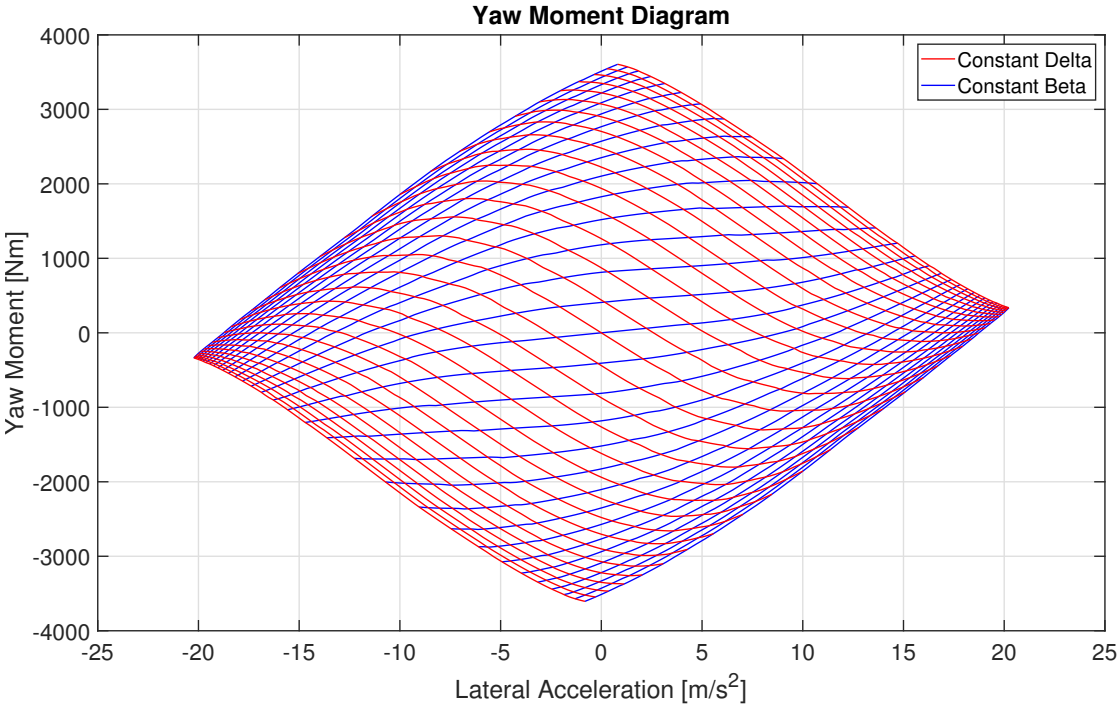


Figure 2.13.: Miliken moment diagram of a Formula Student racing vehicle with constant steering wheel angle (*delta*) and constant chassis side slip angle (*beta*) lines [KL19]

2.3.2. Vehicle model

A two track vehicle model as presented in figure 2.14 with non linearized kinematic tire relations in combination with a *Paceijka Magic Formula* tire model [PB92][PB12] is used by the grip optimization method. At first the kinematic tire velocities are calculated from the input state vector $\mathbf{x}_{\text{in}}$ which are needed to calculate the side slip angles of the tires according to equation (2.20) and (2.21). In order to consider the dynamic behaviour of the tires, a first order lag element for the side slip angle is introduced in (2.22). The

dynamic side slip angle $\alpha_{\mathrm{as_n}}$ can be calculated by discretizing the dynamic equation with a backward differencing scheme, which contrary to a central differencing scheme assures that the solver has an influence on the dynamic side slip angle at the sample point n. The longitudinal slip ratios get solved by the optimizer and longitudinal dynamic tire effects are neglected in the model as they are typically much smaller than lateral dynamic tire effects. Eventually the slip quantities are needed as input for the tire model as well as the vertical tire loads and the tire camber angle to obtain the lateral and longitudinal tire forces Fx_{Tas}, Fy_{Tas} and the corresponding tire aligning torque Mz_{Tas}. The external forces are calculated in (2.24) and (2.25). With the known tire forces, aligning torques and external forces, finally the lateral, longitudinal and angular momentum equations can be defined (2.26 to 2.28).

$$\begin{aligned}
{}_{\mathrm{I}}vx_{\mathrm{Tfl}} &= vx - b_{\mathrm{fl}} \cdot \dot{\psi},\ {}_{\mathrm{I}}vy_{\mathrm{Tfl}} = vy + l_{\mathrm{F}} \cdot \dot{\psi},\\
{}_{\mathrm{I}}vx_{\mathrm{Tfr}} &= vx + b_{\mathrm{fr}} \cdot \dot{\psi},\ {}_{\mathrm{I}}vy_{\mathrm{Tfr}} = vy + l_{\mathrm{F}} \cdot \dot{\psi},\\
{}_{\mathrm{I}}vx_{\mathrm{Trl}} &= vx - b_{\mathrm{rl}} \cdot \dot{\psi},\ {}_{\mathrm{I}}vy_{\mathrm{Trl}} = vy - l_{\mathrm{R}} \cdot \dot{\psi},\\
{}_{\mathrm{I}}vx_{\mathrm{Trr}} &= vx + b_{\mathrm{rr}} \cdot \dot{\psi},\ {}_{\mathrm{I}}vy_{\mathrm{Trr}} = vy - l_{\mathrm{R}} \cdot \dot{\psi},
\end{aligned} \tag{2.20}$$

$$\text{with } vx = v \cdot \cos(\beta) \text{ and } vy = v \cdot \sin(\beta),$$

$$\begin{aligned}
{}_{\mathrm{I}}\mathbf{v}_{\mathrm{Tas}} &= \begin{bmatrix} {}_{\mathrm{I}}vx_{\mathrm{Tas}} \\ {}_{\mathrm{I}}vy_{\mathrm{Tas}} \end{bmatrix},\\
{}_{\mathrm{T}}\mathbf{v}_{\mathrm{Tas}} &= \mathbf{T}_{\mathrm{TIas}} \ \cdot\ {}_{\mathrm{I}}\mathbf{v}_{\mathrm{Tas}}\ ,
\end{aligned} \tag{2.21}$$

$$\text{with } \mathbf{T}_{\mathrm{TIas}} = \begin{bmatrix} \cos\delta_{\mathrm{Tas}} & -\sin\delta_{\mathrm{Tas}} & 0 \\ \sin\delta_{\mathrm{Tas}} & \cos\delta_{\mathrm{Tas}} & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\frac{\sigma_{\mathrm{y}}}{{}_{\mathrm{T}}vx_{\mathrm{Tas}}}\ \dot{\alpha}_{\mathrm{as}} + \alpha_{\mathrm{as}} = \alpha_{\mathrm{stat_{as}}}, \tag{2.22}$$

$$\begin{aligned}
\frac{\sigma_{\mathrm{y}}}{{}_{\mathrm{T}}vx_{\mathrm{Tas}}} \cdot \frac{\alpha_{\mathrm{as_n}} - \alpha_{\mathrm{as_{n-1}}}}{\Delta t} + \alpha_{\mathrm{as_n}} &= \alpha_{\mathrm{stat_{as}}},\\
\text{with } \alpha_{\mathrm{stat_{as}}} &= \frac{{}_{\mathrm{T}}vy_{\mathrm{Tas}}}{{}_{\mathrm{T}}vx_{\mathrm{Tas}}},
\end{aligned} \tag{2.23}$$

$$Fx_{\mathrm{Ext}} \ = \ -D_{\mathrm{x}}, \tag{2.24}$$

$$Fy_{\mathrm{Ext}} \ = \ D_{\mathrm{y}}, \tag{2.25}$$

$$\begin{aligned} m \cdot a_{\mathrm{x}} \; = \; & Fx_{\mathrm{Trl}} \cdot \cos\delta_{\mathrm{Trl}} + Fx_{\mathrm{Trr}} \cdot \cos\delta_{\mathrm{Trr}} - Fy_{\mathrm{Trl}} \cdot \sin\delta_{\mathrm{Trl}} - Fy_{\mathrm{Trr}} \cdot \sin\delta_{\mathrm{Trr}} + \\ & Fx_{\mathrm{Tfl}} \cdot \cos\delta_{\mathrm{Tfl}} + Fx_{\mathrm{Tfr}} \cdot \cos\delta_{\mathrm{Tfr}} - Fy_{\mathrm{Tfl}} \cdot \sin\delta_{\mathrm{Tfl}} - Fy_{\mathrm{Tfr}} \cdot \sin\delta_{\mathrm{Tfr}} + \\ & Fx_{\mathrm{Ext}}, \end{aligned} \tag{2.26}$$

$$\begin{aligned} m \cdot a_{\mathrm{y}} \; = \; & Fx_{\mathrm{Trl}} \cdot \sin\delta_{\mathrm{Trl}} + Fx_{\mathrm{Trr}} \cdot \sin\delta_{\mathrm{Trr}} + Fy_{\mathrm{Trl}} \cdot \cos\delta_{\mathrm{Trl}} + Fy_{\mathrm{Trr}} \cdot \cos\delta_{\mathrm{Trr}} + \\ & Fx_{\mathrm{Tfl}} \cdot \sin\delta_{\mathrm{Tfl}} + Fx_{\mathrm{Tfr}} \cdot \sin\delta_{\mathrm{Tfr}} + Fy_{\mathrm{Tfl}} \cdot \cos\delta_{\mathrm{Tfl}} + Fy_{\mathrm{Tfr}} \cdot \cos\delta_{\mathrm{Tfr}} +, \\ & Fy_{\mathrm{Ext}} \end{aligned} \tag{2.27}$$

$$\begin{aligned} J_{\mathrm{zz}} \cdot \ddot{\psi} \; = \; & l_{\mathrm{F}} \cdot (Fx_{\mathrm{Tfl}} \cdot \sin\delta_{\mathrm{Tfl}} + Fx_{\mathrm{Tfr}} \cdot \sin\delta_{\mathrm{Tfr}} + Fy_{\mathrm{Tfl}} \cdot \cos\delta_{\mathrm{Tfl}} + Fy_{\mathrm{Tfr}} \cdot \cos\delta_{\mathrm{Tfr}}) - \\ & l_{\mathrm{R}} \cdot (Fx_{\mathrm{Trl}} \cdot \sin\delta_{\mathrm{Trl}} + Fx_{\mathrm{Trr}} \cdot \sin\delta_{\mathrm{Trr}} + Fy_{\mathrm{Trl}} \cdot \cos\delta_{\mathrm{Trl}} + Fy_{\mathrm{Trr}} \cdot \cos\delta_{\mathrm{Trr}}) + \\ & b_{\mathrm{fr}} \cdot (Fx_{\mathrm{Tfr}} \cdot \cos\delta_{\mathrm{Tfr}} - Fy_{\mathrm{Tfr}} \cdot \sin\delta_{\mathrm{Tfr}}) + b_{\mathrm{fl}} \cdot (-Fx_{\mathrm{Tfl}} \cdot \cos\delta_{\mathrm{Tfl}} + Fy_{\mathrm{Tfl}} \cdot \sin\delta_{\mathrm{Tfl}}) + , \\ & b_{\mathrm{rr}} \cdot (Fx_{\mathrm{Trr}} \cdot \cos\delta_{\mathrm{Trr}} - Fy_{\mathrm{Trr}} \cdot \sin\delta_{\mathrm{Trr}}) + b_{\mathrm{rl}} \cdot (-Fx_{\mathrm{Trl}} \cdot \cos\delta_{\mathrm{Trl}} + Fy_{\mathrm{Trl}} \cdot \sin\delta_{\mathrm{Trl}}) + \\ & Mz_{\mathrm{Tfl}} + Mz_{\mathrm{Tfr}} + Mz_{\mathrm{Trl}} + Mz_{\mathrm{Trr}} \end{aligned} \tag{2.28}$$

with m as the vehicle mass, J_{zz} the vehicle angular inertia, a_{x}, a_{y} and $\ddot{\psi}$ the longitudinal, lateral and angular vehicle acceleration, l_{F} and l_{R} the front and rear lever from the respective axle to the COG and δ_{Tas} the tire steering angles.

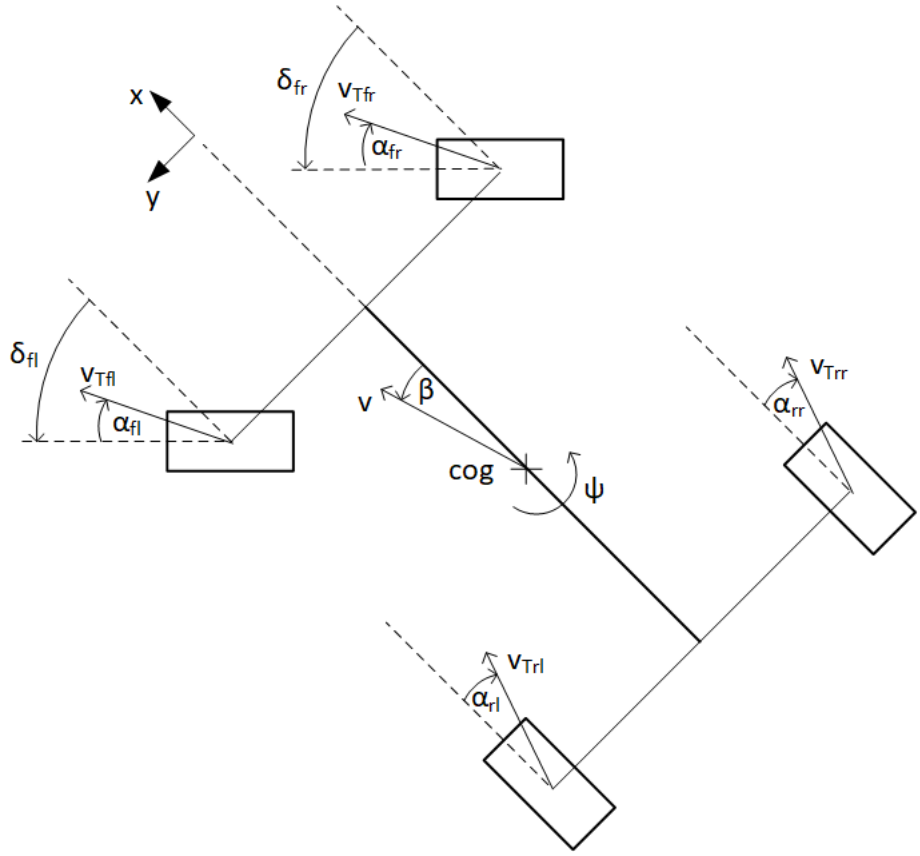


Figure 2.14.: Two track vehicle simplification representing kinematic relationships

2.3.3. Constraints

To ensure that the vehicle can still follow the same trajectory as the driver has intended in the DIL simulator, the yaw angular acceleration $\ddot{\psi}$ has to satisfy equation (2.30).

$$\begin{aligned} \omega_{\mathrm{tr}} &= \sqrt{\frac{a_{\mathrm{r}}}{R}}, \\ \dot{\omega}_{\mathrm{tr}} &= \frac{d}{dt}\left(\sqrt{a_{\mathrm{r}} \cdot \kappa}\right), \\ \dot{\omega}_{\mathrm{tr}} &= \frac{1}{2} \cdot \frac{\dot{a}_{\mathrm{r}} \cdot \kappa + \dot{\kappa} \cdot a_{\mathrm{r}}}{\sqrt{a_{\mathrm{r}} \cdot \kappa}}, \end{aligned} \tag{2.29}$$

$$\ddot{\psi} = \dot{\omega}_{\mathrm{tr}} + \ddot{\beta} \tag{2.30}$$

For simplicity the radial acceleration a_{r}, the angular acceleration of the side slip angle $\ddot{\beta}$ and thus also the yaw angular acceleration $\ddot{\psi}$ are kept constant within the constraint calculation process, leading to constraint (2.32). The constraint (2.31) consists in setting

the angle of the acceleration vector to be constant throughout the optimization process, in order to indicate the optimization direction and limit the search space of the solver. Limiting the search space leads to increased robustness and decreases calculation time. As already mentioned in section 2.3.1 the driving forces are modeled by defining the slip ratios κ_{as} directly as optimization variables. To realistically model the powertrain properties, constraints for the slip ratios have to be introduced accordingly: the tire braking torque difference between the left and the right tire has to be forced constant (2.33); the relation of the tire braking torque of the front axle to the rear axle has to be constant as it is guided by the braking distribution which the driver sets (2.35) and lastly the driving torque has to be distributed between the left and the right tire according to (2.34). Equation (2.36) is the approximated continuously differentiable braking distribution constraint with the cutoff property to set the front axle torque to the offset parameter ϵ_2 if the torque of the rear axle becomes positive, as can be seen in figure 2.15.

$$\frac{a_{\mathrm{x}}}{a_{\mathrm{y}}} - \frac{a_{\mathrm{x0}}}{a_{\mathrm{y0}}} \;=\; 0, \tag{2.31}$$

$$J_{\mathrm{zz}} \cdot (\ddot{\psi} - \ddot{\psi}_0) \;=\; 0, \tag{2.32}$$

$$T_{\mathrm{fl}} - T_{\mathrm{fr}} - (T_{\mathrm{fl0}} - T_{\mathrm{fr0}}) \;=\; 0, \tag{2.33}$$

$$T_{\mathrm{rl}} - T_{\mathrm{rr}} - (T_{\mathrm{rl0}} - T_{\mathrm{rr0}}) \;=\; 0, \tag{2.34}$$

$$\begin{aligned} T_{\mathrm{f}} - (T_{\mathrm{f}} + T_{\mathrm{r}}) \cdot br_{\mathrm{dist}} \;&=\; 0, \\ \text{with } T_{\mathrm{f}} \;=\; T_{\mathrm{fl}} + T_{\mathrm{fr}},& \text{ and } T_{\mathrm{r}} \;=\; T_{\mathrm{rl}} + T_{\mathrm{rr}}, \end{aligned} \tag{2.35}$$

$$\approx\; T_{\mathrm{f}} - (T_{\mathrm{f}} + T_{\mathrm{r}}) \cdot br_{\mathrm{dist}} \cdot \left(\frac{\arctan(-\epsilon_1 \cdot (T_{\mathrm{f}} - \epsilon_2))}{\pi} + \frac{1}{2} \right) \;=\; 0, \tag{2.36}$$

where the quantities with the suffix 0 represent the variables at the starting point of the optimization, T_{as} are the tire driving and braking torques, br_{dist} is the braking distribution and ϵ_1 and ϵ_2 are the smoothness and the offset parameters for the approximation.

The normalization of the nonlinear inequality constraints has a big impact on the robustness of the optimization. A successful normalization yields to a solver search space with equally distributed gradients which increases the probability in finding the point with the minimum objective function value. The nonlinear equality constraints are normalized by the standard deviation of the following simulator signals within a lap corresponding to the order of the constraint equations (2.31) to (2.35): longitudinal and lateral acceleration, angular acceleration, torque of the front left tire, driving torque difference of the rear tires, torque of the front tires. Additionally, different weighting factors w_{lat} and w_{T} for the brake distribution torque constraint are introduced in (2.37) and (2.38) which are activated at different vehicle states. At largely lateral cornering states with the absolute acceleration ratio $|\frac{a_{\mathrm{x}}}{a_{\mathrm{y}}}| < 0.28$ and the longitudinal acceleration $a_{\mathrm{x}} < 0$, w_{lat} is multiplied to the constraint. For every other maneuver the torque constraint gets multiplied with

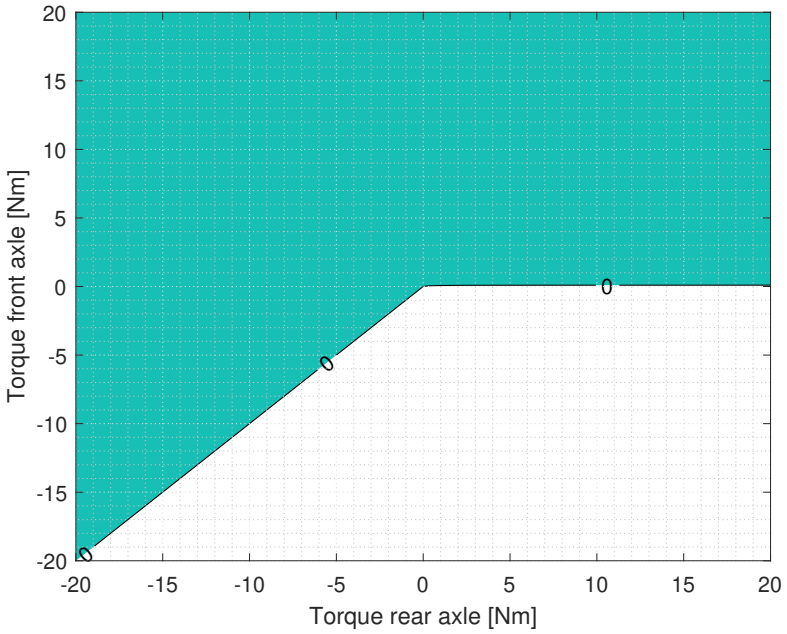


Figure 2.15.: Braking distribution constraint violation with $br_{\text{dist}} = 0.5$ and $\epsilon_1 = 10^4$, $\epsilon_2 = 0$

the factor w_{T}. The weighting values and activation conditions are chosen empirically and improve the convergence probability significantly.

$$w_{\text{lat}} = 10^{-1} \tag{2.37}$$

$$w_{\text{T}} = 6 \tag{2.38}$$

To limit the tire operating window inequality constraints for the tire side slip angles and the slip ratios are set accordingly in (2.39) and (2.40). Equation (2.41) assures that the driving torque does not exceed the maximum torque available from the engine.

$$\left(\frac{\alpha_{\text{as}}}{\alpha_{\text{max}}}\right)^2 - 1 \leq 0, \tag{2.39}$$

$$\left(\frac{\kappa_{\text{as}}}{\kappa_{\text{max}}}\right)^2 - 1 \leq 0, \tag{2.40}$$

$$T_{\text{r}} - T_{\text{Engine}_{\text{max}}} \cdot i_{\text{Tot}} \leq 0, \tag{2.41}$$

2.3.4. Solver

The optimization is run with the *fmincon* solver of the *MathWorks® MATLAB optimization toolbox* package [MAT18] as it accepts nonlinear constraints for the optimization

problem. The derivative information of the constraint and objective function is supplied to the solver using the automatic differentiation tool *ADiGator* [WR17] in the MATLAB environment. A *sequential quadratic programming* [BT95] solver algorithm is choosed, as it is established for the usage of nonlinear constraints with the settings of table 2.6 to reach a fast and stable convergence rate. The stepsize tolerance of the solver vector is set with f_{tol}. The optimization will be stopped at a possible local minimum if the stepsize goes below the value of the stepsize tolerance, regardless whether the first order optimality condition is satisfied or not. The satisfaction of the first order optimality tolerance opt_{tol} indicates the presence of a local minimum. Both stopping criteria are interpreted as a successful result, as typically a local minimum is reached if the stepsize gets smaller than the defined stepsize tolerance. If the relative constraint violation is below its respective tolerance con_{tol} (related to the constraint violation of the first iteration), the solution is declared as feasible. To abort the optimization process if no optimal solution can be found, $f_{\max}$ delimits the maximum number of iterations.

Table 2.6.: Solver Settings for the optimization

Setting	Value
f_{tol}	$1 \cdot 10^{-4}$
$f_{\max}$	$2 \cdot 10^{3}$
con_{tol}	$1 \cdot 10^{-1}$
opt_{tol}	$1 \cdot 10^{-4}$

2.3.5. Performance Relevance

From the kinematic relations of a point mass travelling on a given trajectory in a polar coordinate system the radial acceleration a_{r} can be expressed as function of the tangential velocity v and the given radius R in (2.42).

$$a_{\text{r}} = \frac{v}{R} \tag{2.42}$$

Assuming this relation constant the velocity can be seen as the time derivative of the distance s hence, integrating the velocity with respect to the time t leads to (2.43).

$$\begin{aligned} v &= \frac{ds}{dt}, \\ \int_0^{\tau} v \; dt &= \int_0^{S} ds, \\ \tau &= \frac{S}{a_{\text{r}} \cdot R}, \end{aligned} \tag{2.43}$$

with S and τ, as the distance travelled and the time spent with the latter mentioned assumptions. Eventually the approximation of small side slip angles, as already explained

in section 2.1.1, yields to a relation between the lateral acceleration a_y of the vehicle and the time as performance measure (2.44).

$$\tau = \frac{S}{a_\mathrm{y} \cdot R} \tag{2.44}$$

Similarly a relation between the longitudinal vehicle acceleration and the time can be elaborated with the definition of the longitudinal acceleration a_x as time derivative of the velocity (2.45).

$$\begin{aligned} a_\mathrm{x} &= \frac{dv_\mathrm{x}}{dt}, \\ \int_0^t a_\mathrm{x}\, dt &= \int_0^{v_\mathrm{x}} dv_\mathrm{x}, \\ a_\mathrm{x} \cdot t &= v_\mathrm{x}, \\ \text{with } \frac{ds_\mathrm{x}}{dt} &= v_\mathrm{x}, \\ \int_0^\tau a_\mathrm{x} \cdot t\, dt &= \int_0^S ds_\mathrm{x}, \\ \tau &= \sqrt{\frac{2\ S}{a_\mathrm{x}}}, \end{aligned} \tag{2.45}$$

with s_x as the longitudinal distance variable.

The introduced relations state that the time is inversely proportional to the lateral acceleration and is inversely proportional to the square root of the longitudinal acceleration. Hence, acceleration improvements should be prioritized for the lateral acceleration. Of course these relations are not valid throughout an entire lap but only for a small time interval τ where the accelerations and the travelled distance S remain constant. Nevertheless, it proves the performance relevance of variables which represent the vehicle acceleration exploitation by the driver and therefore gives the motivation to develop driver metrics upon these measures.

2.3.6. Metrics derived from the Grip Optimization Method

To use the grip optimization algorithm for the driving style characterization, the resulting time series of the theoretical maximum acceleration of the vehicle $a_\mathrm{GO}(t)$ has to be transformed to driver metrics. The ratio of the unused grip potential to the actual acceleration of the vehicle $ra_\mathrm{GO}(t)$ is defined in (2.46). Equation (2.47) defines the absolute difference between the theoretical maximum grip and the actual grip of the vehicle $\Delta a_{\mathrm{GO}_\mathrm{dir}}(t)$, where the suffix $dir \in \{x,\ y\}$ stands for longitudinal (x) or lateral (y) related to the acceleration direction

$$ra_\mathrm{GO}(t) = \frac{a_\mathrm{GO}(t)}{a(t)} - 1, \tag{2.46}$$

$$\Delta a_{\mathrm{GO}_{\mathrm{dir}}}(t) \;=\; |a_{\mathrm{GO}_{\mathrm{dir}}}(t) - a_{\mathrm{dir}}(t)|. \tag{2.47}$$

In table 2.8, metrics are developed on the absolute and the relative integrated value of the introduced measures. The reference time for the relative quantities is the time spent in the sector. The metrics are split for the corresponding corner phase entry, mid or exit as summarized in table 2.7. The values t_0 and t_1 represent the beginning of the corner phase and the end of the corner phase for the definition of the integration intervals. The acceleration signal follows the driver inputs with a certain lag, due to the dynamic behaviour of load transfer transition phases. Load transfer transition phases usually occur in the beginning of the braking application at corner entries or the beginning of the throttle application at corner exits. In addition with the fact that the grip optimization algorithm operates in the direction of the measured initial acceleration, acceleration margins may be defined in the wrong directions regarding the drivers intention. For example at the beginning of the braking phase the longitudinal acceleration is still positive even if the drivers are already applying the brakes with the aim of decelerating the vehicle. Now the grip optimizer searches acceleration margins which would result in the positive longitudinal direction hence, accelerating the vehicle which is not in accordance with the drivers. This behaviour is avoided by introducing longitudinal acceleration conditions during the metric definition: $a_{\mathrm{x}} < 0$ for the corner entry and $a_{\mathrm{x}} > 0$ for the corner exit phase. Thereupon, only data points with the correct sign in respect to the cornering phase are considered for the calculation of the metrics.

Table 2.7.: Corner phase split points

Corner phase	Split points	
	t_0	t_1
Entry	on-brake	off-brake
Mid	off-brake	on-throttle
Exit	on-throttle	full-throttle

Table 2.8.: Metrics based on the grip optimization algorithm

Metric	Description
rGOEntryAbs rGOMidAbs rGOExitAbs	$\sum_{t=t_0}^{t_1} ra_{\mathrm{GO}}(t) \cdot \frac{1}{f_\mathrm{s}}$
rGOEntryRel rGOMidRel rGOExitRel	$\frac{1}{t_\mathrm{sector}} \cdot \sum_{t=t_0}^{t_1} ra_{\mathrm{GO}}(t) \cdot \frac{1}{f_\mathrm{s}}$
dGOLongEntryAbs dGOLongMidAbs dGOLongExitAbs	$\sum_{t=t_0}^{t_1} \Delta a_{\mathrm{GO}_\mathrm{Long}}(t) \cdot \frac{1}{f_\mathrm{s}}$
dGOLongEntryRel dGOLongMidRel dGOLongExitRel	$\frac{1}{t_\mathrm{sector}} \cdot \sum_{t=t_0}^{t_1} \Delta a_{\mathrm{GO}_\mathrm{Long}}(t) \cdot \frac{1}{f_\mathrm{s}}$
dGOLatEntryAbs dGOLatMidAbs dGOLatExitAbs	$\sum_{t=t_0}^{t_1} \Delta a_{\mathrm{GO}_\mathrm{Lat}}(t) \cdot \frac{1}{f_\mathrm{s}}$
dGOLatEntryRel dGOLatMidRel dGOLatExitRel	$\frac{1}{t_\mathrm{sector}} \cdot \sum_{t=t_0}^{t_1} \Delta a_{\mathrm{GO}_\mathrm{Lat}}(t) \cdot \frac{1}{f_\mathrm{s}}$

2.3.7. Limitations

The input state vector $\mathbf{x}_\mathrm{in}$ is loaded from the simulator data and remains constant within the optimization process and the calculation for the yaw angular acceleration constraint is based on the initial state, instead of being a function of the radial acceleration and the side slip angular acceleration. This is a simplification of the model, as these quantities would be a function of the solver vector entries $\mathbf{x}_\mathrm{solv}$. However, on the other hand it has also a positive influence on the numerical effort. Instead of modeling the entire powertrain and giving the pedal application as solver input, the slip ratios κ_as of the four tires are supplied to the solver directly. Therefore the torque constraints mentioned in section 2.3.3 have to be introduced, to replicate the properties of the powertrain. The constraints lead to difficulties in the solving process if a tire load gets close to zero, causing variations in the tire slip quantities to have only very small influences on the vehicle model. With this approach the angular tire dynamics are neglected completely which may also lead to inaccuracies of the method.

2.3.8. Validation and Error Estimation

The tire steering angles δ_{Tas} are calculated with a linear ratio from the steering wheel angle δ and the tire actual toe angles μ_{as} as an offset. In the real car the tire steering angles may not be proportional to the steering wheel angle, but a function of many parameters which determine the kinematic relationships of the suspension-tire system. This results in a modeling error of the initial state $e_{\delta_{\mathrm{Tas}}}$ which propagates through the tire model to the lateral, longitudinal and angular momentum equations and may have a not negligible impact on the vehicle model. The error propagation is analyzed by a taylor series expansion in equation (2.48), (2.49) and (2.50).

$$e_{\mathrm{a_x}} = \frac{\partial a_{\mathrm{x}}}{\delta_{\mathrm{Tfl}}} \cdot e_{\delta_{\mathrm{Tfl}}} + \frac{\partial a_{\mathrm{x}}}{\delta_{\mathrm{Tfr}}} \cdot e_{\delta_{\mathrm{Tfr}}}, \tag{2.48}$$

$$e_{\mathrm{a_y}} = \frac{\partial a_{\mathrm{y}}}{\delta_{\mathrm{Tfl}}} \cdot e_{\delta_{\mathrm{Tfl}}} + \frac{\partial a_{\mathrm{y}}}{\delta_{\mathrm{Tfr}}} \cdot e_{\delta_{\mathrm{Tfr}}}, \tag{2.49}$$

$$e_{\ddot{\psi}} = \frac{\partial \ddot{\psi}}{\delta_{\mathrm{Tfl}}} \cdot e_{\delta_{\mathrm{Tfl}}} + \frac{\partial \ddot{\psi}}{\delta_{\mathrm{Tfr}}} \cdot e_{\delta_{\mathrm{Tfr}}}, \tag{2.50}$$

with $e_{\mathrm{a_x}}$, $e_{\mathrm{a_y}}$, $e_{\ddot{\psi}}$ as the lateral, longitudinal and angular momentum error.

However, as not all modelling errors are known, the lateral, longitudinal and yaw angular acceleration calculated from the grip optimization vehicle model are directly compared to the vehicle model from the DIL simulator in section 3.2.3.

2.4. Metric Selection

The metrics with the most relevant information according to their ability to distinct driving styles have to be selected. It is important to know the relevance in combination with the meaning of the metrics when it comes to analyzing drivers for their performance. Furthermore, classification algorithms, as the one presented in section 2.4.3, generally show higher accuracies if they are not feeded with nonrelevant information. Thus, only the most important metrics in terms of driving style classification are used.

First, the amount of Not a Number (NaN) values in the driver metrics is determined. If a metric contains more than ten percent NaN it gets removed for the driver classification. Then, a correlation matrix is generated to avoid using metrics that have a high correlation with each other. High correlated metrics show a similar importance for driving style classification, while decreasing the overall importance of them. The F-values of an Analysis of Variance (ANOVA) are used to get information about the relative variance of the metrics and the importance factors of the driving style classification algorithm are determined. These methods are applied recursively as each feature selection measure is dependend on the set of features used in the method [GWBV02]. On the combined information of these methods finally the most relevant metrics are selected.

2.4.1. Correlation Analysis

In general, the *Pearson* correlation describes measures of the linear dependency of two random variables that are defined by using the mean μ_{A}, μ_{B} and standard deviation σ_{A}, σ_{B} of variable A and B accordingly to (2.51) [CD03].

$$\rho(A,B) \;=\; \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{A_{\mathrm{i}}-\mu_{\mathrm{A}}}{\sigma_{\mathrm{A}}}\right)\left(\frac{B_{\mathrm{i}}-\mu_{\mathrm{B}}}{\sigma_{\mathrm{B}}}\right) \tag{2.51}$$

For the *Spearman* correlation, the rank rg of the variables is used instead of the direct variable as in (2.52).

$$\rho_{\mathrm{s}}(A,B) \;=\; \frac{1}{N}\sum_{i=1}^{N}\left(\frac{rg(A_{\mathrm{i}})-\mu_{\mathrm{rg(A)}}}{\sigma_{\mathrm{rg(A)}}}\right)\left(\frac{rg(B_{\mathrm{i}})-\mu_{\mathrm{rg(B)}}}{\sigma_{\mathrm{rg(B)}}}\right) \tag{2.52}$$

As a result, the Spearman correlation determines the strength and the monotonic relationship between two variables, instead of the strength and direction of the linear relationship as the Pearson correlation does [HK11].

Given the correlation coefficients among the metrics, a correlation coefficient matrix $\mathbf{R} \in \mathbb{R}^{2\times 2}$ in (2.53) can be calculated, in order to get a better information representation with a correlation coefficient range from minus one (inversely correlated) over zero (not linearly correlated) to one (correlated).

$$\mathbf{R} \;=\; \begin{bmatrix}\rho_{\mathrm{s}}(A,A) & \rho_{\mathrm{s}}(A,B)\\ \rho_{\mathrm{s}}(B,A) & \rho_{\mathrm{s}}(B,B)\end{bmatrix} \tag{2.53}$$

If a metric can not be calculated in a specific sector the metric gets the value NaN. As these NaN values cannot be used for the correlation analysis the affected entire sectors get neglected. In addition, metrics with more than 50 percent NaN values are not considered for the analysis.

Metrics that show a correlation higher than an empirically defined critical correlation of $\rho_{\mathrm{crit}} > |0.85|$ are removed from the classification task. The metrics with the higher importance factor according to section 2.4.3 is kept.

2.4.2. F-Value in ANOVA

The F-value of ANOVA gives an indication if the variance between the means of two populations is significantly different [Arc94]. Generally it can be defined as the ratio of the variance between the class means to the mean of the within class variances, see equation (2.54), where in the case of driver classification the different drivers represent a class. This means that higher F-values in the driver class indicate a better suitability

for metrics to distinct drivers and thus driving styles.

$$\begin{aligned} F &= \frac{\text{between-group variability}}{\text{within-group variability}} \,, \\ F &= \frac{\sum_{i=1}^{K} n_{\mathrm{i}} (\mu_{\mathrm{y_i}} - \mu_{\mathrm{y}})^2 / (K-1)}{\sum_{i=1}^{K} \sum_{j=1}^{n_{\mathrm{i}}} (y_{\mathrm{ij}} - \mu_{\mathrm{y_i}})^2 / (N-K)} \,, \end{aligned} \tag{2.54}$$

where n_{i} is the amount of metric values in the i^{th} group, $\mu_{\mathrm{y_i}}$ is the mean in the i^{th} group, K denotes the number of groups, y_{ij} is the j^{th} value in the i^{th} group and N is the overall amount of data.

In addition, for further observations of track dependencies of the metrics, the F-values of the track class are calculated. This gives information of how a metric variates among different race tracks for all drivers. If metric values for a specific driver do not variate among tracks, they may be able to describe driving styles independently.

2.4.3. Driver Classification based on a Decision Tree Approach

Driving styles show extensive properties with patterns that are difficult to describe manually by mathematical relations. Hence, pattern recognition approaches may be used to handle the problem of driving style distinction. In addition, the categorization of driving styles is rather complex and investigations are mainly based on subjective measures. For the task to objectively distinguish driving styles it is assumed as a first benchmark that every driver has an individual driving style. Thus the problem can be formulated by classifying drivers. A *Machine Learning* algorithm is applied is used for recognizing distinctive patterns in the data derived by the driver inputs.

Convential programming has a simple goal, namely automation. A traditional programmer would provide the computer program in collaboration with an input to produce a certain output. This approach would only be successful if a mathematical formula could be derived in order to solve a given problem. *Machine Learning* is characterized by automating the process of automation and takes input and output data to create a program itself. It breaks the approach of conventional programmers and works even if it is not possible to formulate a mathematical principle, by repeatedly recognizing patterns and developing its own algorithms as the input data changes [SAS19].

The random forest method which is used in this work relies on principles of decision tree algorithms. Decision trees logically combine a sequence of simple tests comparing numeric attributes against threshold values in order to split data recursively. The split points are named *nodes*, where the topmost nodes are referenced as *root* and the terminal nodes as *leaves*. Each leave represents a classification label that is used to make the prediction. Such symbolic classifiers are relative comprehensible. The logical rules followed by a decision tree rely on simple, easy interpretable principles for the decision

makers and thus improve the understandability of the model [Kot13]. Since a decision tree can implement boolean functions, there is a danger of *overfitting*: If the training set is not sufficiently large, there will still be too many consistent functions to make useful predictions on the training set. Overfitted models show poor generalization performance, thus the model might be not able to make successfull predictions based on new patterns that were not used to build the decision tree [Nil96].

In order to build a decision tree model an optimization algorithm is used to maximise the decrease of an impurity measure $\Delta i_{\mathrm{T}}(s_{\mathrm{T}}, t_{\mathrm{T}})$ for each node t_{T} and split s_{T} in the tree T according to equation (2.55) [LWSG13]. As criteria for node impurity the Gini Index g_{i} is applied [Moi08]. Feature importance measures can be introduced as the *Mean decrease in Gini*, which is defined as the total decrease in node impurity weighted by the probability of reaching that node. The node probability is approximated by the proportion of samples reaching the corresponding node [LWSG13]. The mean decrease in gini of a feature is dependent on all the other features in the training set and may vary if different features are used for the model [GWBV02].

$$\begin{aligned} \Delta i_{\mathrm{T}}(s_{\mathrm{T}}, t_{\mathrm{T}}) \;&=\; i_{\mathrm{T}}(t_{\mathrm{T}}) - p_{\mathrm{L}}\; i_{\mathrm{T}}(t_{\mathrm{TL}}) - p_{\mathrm{R}}\; i_{\mathrm{T}}(t_{\mathrm{TR}}), \\ \text{with } i_{\mathrm{T}}(t_{\mathrm{T}}) \;&=\; g_{\mathrm{i}}(t_{\mathrm{T}}), \\ \text{and } p_{\mathrm{L}} \;=\; \frac{N_{\mathrm{t_{TL}}}}{N_{\mathrm{t_T}}},\; p_{\mathrm{R}} \;&=\; \frac{N_{\mathrm{t_{TR}}}}{N_{\mathrm{t_T}}}, \end{aligned} \tag{2.55}$$

where $N_{\mathrm{t_T}}$ represents the sample size at the node t_{T} and $N_{\mathrm{t_{TR}}}$, $N_{\mathrm{t_{TL}}}$ the sample size of the splitted partition at the node.

The Mean decrease in Gini is primarily used in the construction phase of the Decision Tree but may also be utilized as an approximation for the importance of the metric for a classification task.

Random forests are a type of ensemble method, where each tree is trained in isolation and the predictions of several decision trees are combined using *bagging*. Bagging is a method for generating multiple versions of a predictor and using these to get an aggregated predictor. The aggregated predictor is given by averaging the base versions and in order to predict a class, a plurality vote is done [Bre96]. Due to the controlled injection of randomness and the strong law of large numbers, random forests always converge so that overfitting is not a problem in contrary to Isolate Decision Trees [Bre01]. The application of the method is done with the *ranger* package of the programming language *R* [R C19] and *R Studio Integrated Development Environment* [RSt18]. It is a fast implementation of random forests for high dimensional data and the package is already studied and validated with success by Wright et al. [WZ17].

2.5. Metric Transferability

2.5.1. Transferability from real Car to Simulator

In order to use simulator data for the methodology of this work, the application and specifically the transferability of the metrics from real car to DIL simulator data is studied. Histogram probability distributions, F-values and decision tree importance factors are compared among the data of both racing series.

2.5.2. Transferability between Racing Series

For the application of the driver metrics on more racing series with generally different vehicle classes, differences in the histogram distributions, F-values and decision tree importance factors are analyzed. In table 2.9 a short overview of the vehicle properties of the two racing series is given. Vehicle class A is characterized by high power to weight and high downforce properties with slick tires made for driving on wide racing circuits. Whereas vehicle class B is pronounced by a relatively low power to weight ratio, a chassis with a low downforce low drag concept and grooved tires. Racing series B is mainly competing at tight street circuits with corners of high curvatures. Energy saving plays a big role in racing series B while in racing series A generally enough energy is available.

Table 2.9.: General relative vehicle concept differences for the corresponding racing series

Module	Racing Series	Vehicle properties
Powertrain	A	High power to weight ratio with a characteristic driving torque curve
	B	Low power to weight ratio with a more constant driving torque
Aerodynamics	A	High downforce concept with associated high drag
	B	Low downforce concept with low drag
Tires	A	High peak friction coefficient with high tire slip stiffness
	B	Lower peak friction coefficient with lower tire slip stiffness
Energy	A	Sufficient energy available, energy saving not needed primarily
	B	High priority on energy management

The consideration of the driver-vehicle-environment interaction as a closed loop system implies that the driving style strongly depends on the driven vehicle. To compare driver characteristics of racing series B with racing series A, the metrics must be transferred to the corresponding vehicle class. A transfer offset tf_{metric} is introduced for each metric as the median of the metric values from racing series A $\tilde{y}_{\text{A}}$ subtracted by the median of the metric values from racing series B $\tilde{y}_{\text{B}}$. The offset of (2.56) is then added to the metric values of racing series B where y represents a specific metric.

$$tf_{\text{metric}} = \tilde{y}_{\text{A}} - \tilde{y}_{\text{B}} \tag{2.56}$$

For the metrics of table 2.10 other definitions are used for the normalization across the racing series. The gradient of the linearized throttle activation is normalized by multiplying the maximum power output of the powertrain related to the vehicle mass, resulting in $drThrOnNorm$. As the brake pressure depends on the respective vehicle, the linearized brake activation and deactivation gradients are divided by the maximum brake pressure of the corresponding sector ($dpBrkOnNorm$, $dpBrkOffNorm$). The vehicle of racing series B has the possibility to recuperate energy instead of braking conventionally by dissipating energy with the mechanical braking system. In order to decelerate the vehicle with the focus on saving energy, drivers would practice lift and coast followed by the application of the recuperation pedal. Followingly, if more deceleration is necessary than the recuperation system can deliver, the mechanical brakes are actuated subsequently. The intention of the metric $sBrkRel$ and $tBrkDlyAbs$ is to capture the deceleration with the brake pressure signal that for racing series B is not necessarily correlated with the deceleration phase. Thus for the normalization of the latter metrics they are defined on the recuperation pedal actuation if a recuperation actuation is recognized, otherwise, the original definition with the on-brake point detection as activation condition is used.

Table 2.10.: Specific normalized metrics

Metric	Description
drThrOnNorm	$drThrOn \cdot P_{\text{Engine}_{\text{max}}}/m$
dpBrkOnNorm	$dpBrkOn \cdot 1/\max(p_{\text{Brk}})$
dpBrkOffNorm	$dpBrkOff \cdot 1/\max(p_{\text{Brk}})$
sBrkRgnRel	Relative distance between on- and off-brake points or if the recuperation is actuated relative distance between recuperation actuation and off-brake points
tBrkRgnDlyAbs	Time delay between off-throttle and on-brake points or if the recuperation is actuated time delay between off-throttle and recuperation actuation points

2.6. Comparison of Feature Extraction Methods

As machine learning methods are used for the driver prediction, it is important to provide the algorithm with only the most significant information measures (features) with respect to the required task (see section 2.4.3). Therefore, the feature calculation intervals must be chosen with care and three different feature extraction methods are compared with each other in this chapter.

Arefnezhad et al. [ASEN19] defines features using sliding windows with a constant width of three seconds and 50 percent overlap for identifying driver drowsiness states for road car drivers. The features of section 2.2.3 based on the steering wheel angle and the steering wheel angle rate are used. Before calculating the features from the steering wheel angle, the signal gets filtered from the road curvature by subtracting the mean of the signal within a sliding window of the raw signal. However, racing car drivers usually have a very high bandwidth for their control inputs compared to common road car drivers. This leads to a higher steering wheel activity and a higher information density in the signal. Thereupon, a highpass filter is applied for the filtering of the road curvature from the steering wheel angle, instead of subtracting the mean of the signal within a sliding window. The highpass filter properties (explained in section 2.2.3) are defined empirically to remove the road curvature while keeping the information of the stabilization effects which typically occur in braking maneuvers.

Segers et al. [Seg14] calculates features with a more practical meaning for driving styles in racing application and many of them are similar to the features defined by Wörle et al., presented in section 2.2. The features are calculated in the interval of entire laps.

The aforementioned feature extraction methods are directly compared to the method of Wörle et al. [WGE18]. He uses grip limited sectors as metric calculation intervals which is the foundation of the method elaborated in this work and further explained in section 2.2.1.1. For the comparison, the features originally used in the feature extraction methods of Arefnezhad et al. and Segers et al. are applied directly on grip limited sector intervals.

3. Results and Discussion

The following chapter contains the results of the methods elaborated in this work. First, information about the database on which the methods are applied is given, followed by the error propagation of the introduced approximations. Next, the results of the grip optimization algorithm are presented including the derived driver metrics. An in depth analysis of the metrics calculated from the different racing series is illustrated including the results from the driver classification problem. Furthermore, the transferability analyzes of the metrics among the racing series and the data sources (real car or DIL simulator) are represented. The chapter is concluded by the results of a comparison of different feature extraction methods.

3.1. Database

The biggest part of the data was gathered from racing series A. During three seasons 16840 laps were recorded of ten different drivers competing at 28 race weekends where each consists of three free practices, two qualifyings and two races (figure 3.1). The races were held at ten different race tracks. For racing series B 1096 laps are available, recorded during two seasons with 17 race weekends consisting of two free practices, one to two qualifyings and one race for four different drivers (figure 3.2). The least amount of data is derived from DIL simulator sessions for racing series A (figure 3.3). Twelve DIL events and four additional events for a dissertation project with five different drivers lead to further 1374 laps which are analyzed for this work. No race track information is available for DIL simulator sessions.

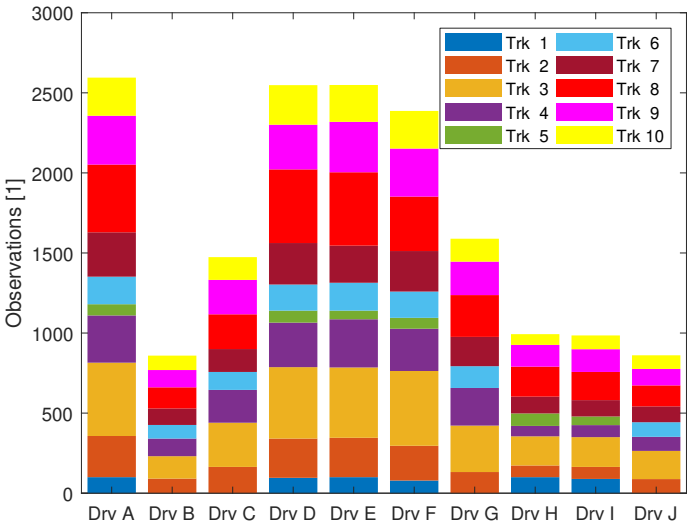


Figure 3.1.: Amount of laps used for each driver on the corresponding race tracks for racing series A

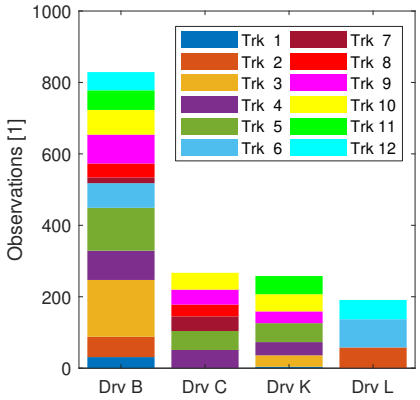


Figure 3.2.: Amount of laps used for each driver on the corresponding race tracks for racing series B

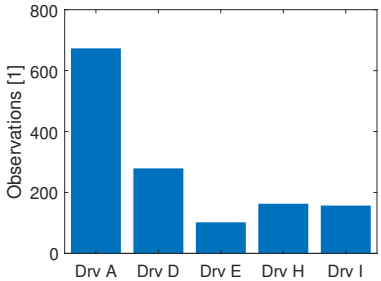
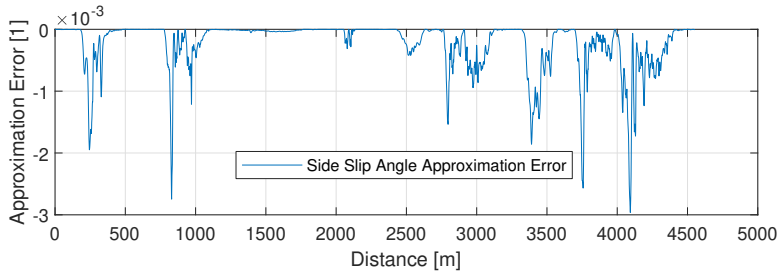


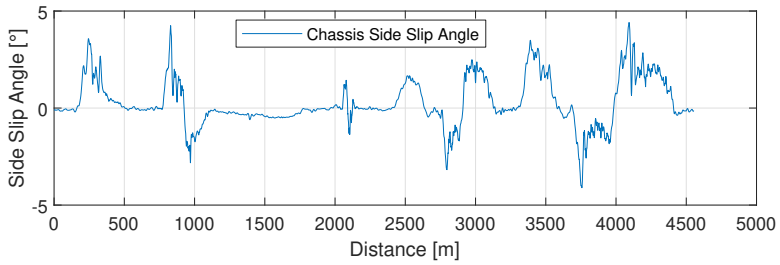
Figure 3.3.: Amount of laps driven in the DIL simulator for each driver of racing series A

3.1.1. Approximation of Missing signals

The approximations introduced in section 2.1.1 generate errors in the corresponding signals. For small chassis side slip angles the approximation $\cos\beta \approx 1$ is applied which results in an error with a maximum of $e_{\cos\beta-1} = -3\cdot10^{-3}$ at high chassis side slip angles, as can be seen in figure 3.4. Typical values of the mean and standard deviation of the chassis side slip angle during a lap are $\mu_\beta \approx 0.25°$ and $\sigma_\beta \approx 1.31°$. The approximation influences the calculation of the path curvature in figure 3.5 with a maximum error of $e_\kappa = 6\cdot10^{-6}[1/\mathrm{m}]$. The mean of the relative error between the two calculation methods is $\mu_{e_\kappa} \approx -0.006[1/\mathrm{m}]$ and the standard deviation equals $\sigma_{e_\kappa} \approx 0.099[1/\mathrm{m}]$. The yaw angular velocity is approximated by assuming the chassis side slip angle β to be constant wich leads to a chassis side slip angle rate $\dot{\beta}$ of zero. Hence, the yaw angular velocity is determined by the curvature of the driven trajectory. The yaw velocity error equals the chassis side slip rate (figure 3.6) and the mean of the relative error is $\mu_{e_{\dot\psi}} \approx -0.013°/s$ and the standard deviation equals $\sigma_{e_{\dot\psi}} \approx 0.599°/s$. It can be seen that all the error maxima lie in areas with high chassis side slip angles. The yaw angular velocity approximation has a relatively big maximum error of $e_{\dot\psi} = \pm20°/\mathrm{s}$ compared to the exact measure, but is also the quantity with the least amount of missing data.

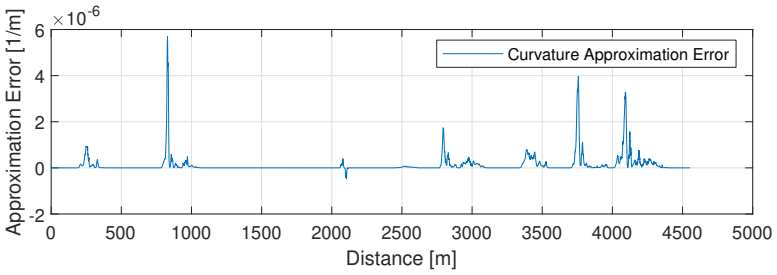


(a) Approximation error ($\cos\beta - 1$)

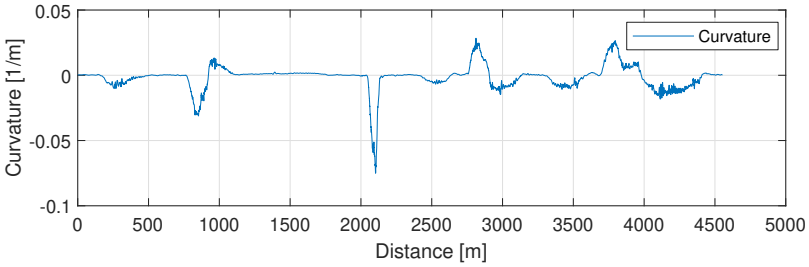


(b) Chassis side slip angle

Figure 3.4.: Approximation error for small chassis side slip angles according to (2.2)

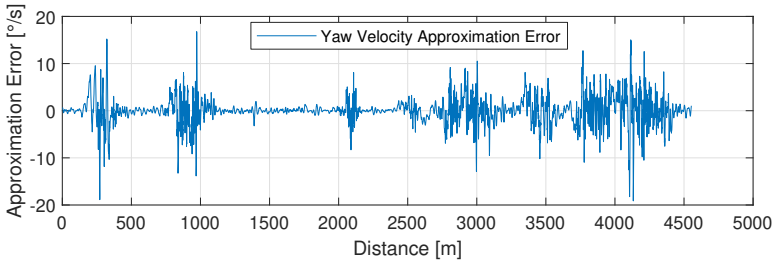


(a) Path curvature approximation error

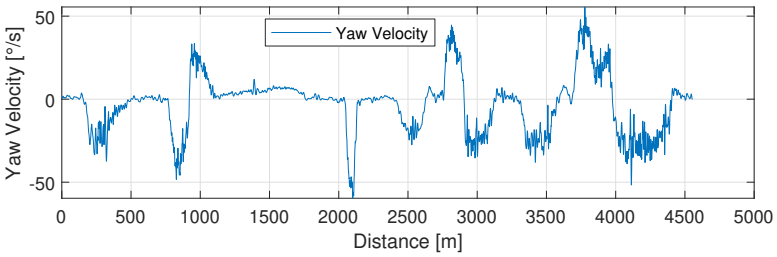


(b) Exact path curvature

Figure 3.5.: Error propagation of the approximation for small chassis side slip angles in the path curvature (2.3)



(a) Yaw velocity approximation error



(b) Exact yaw velocity

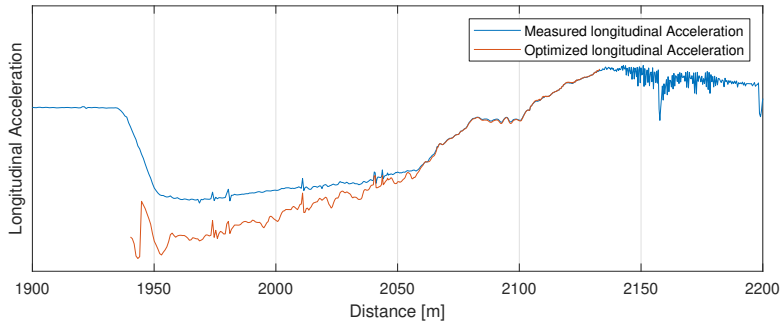
Figure 3.6.: Approximation error for neglecting the chassis side slip angle rate (2.6)

3.2. Grip Optimization Algorithm

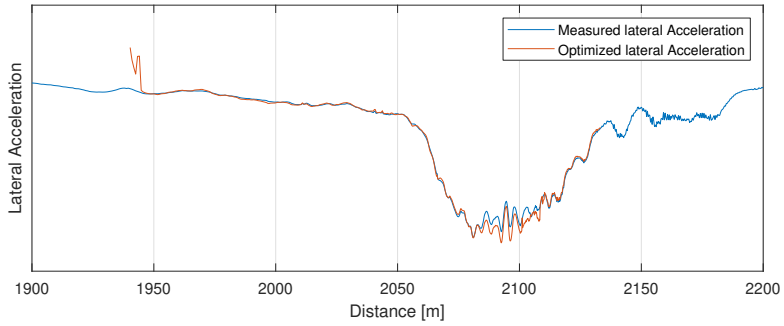
The results from this chapter are based on DIL simulator data of five drivers recorded during one season for racing series A. First the results of the grip optimization method (section 2.3) are presented, followed by metrics defined on these quantities. Next, a validation and error analysis is done, to evaluate the used vehicle model. The chapter is concluded with a discussion on the results.

3.2.1. Results

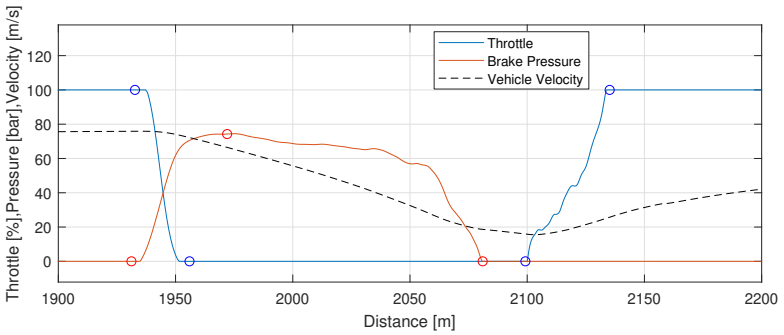
The grip optimization method yields the theoretical maximum vehicle performance in terms of acceleration. Figure 3.7 illustrates the resulting optimized acceleration in longitudinal and lateral direction, exemplary at a sector with a relatively large braking zone. The biggest difference between the measured longitudinal acceleration and the optimized longitudinal acceleration occurs from 1950 to 2050 meters at the beginning of the braking phase. In the lateral acceleration signal only small deviations between the measured and the optimized quantity are visible between 2080 and 2120 meters which corresponds to the rolling phase in the corner. Figure 3.8 is calculated at one sample and illustrates the optimization process of the longitudinal tire forces for all four tires. The blue cross represents the slip ratio and the corresponding longitudinal tire force of the initial optimization starting point, whereas the yellow circle shows the optimized tire force. Due to the characteristics of the used tire model which includes the modeling of the *friction circle*, the longitudinal tire force is influenced by the lateral side slip angle. Thereupon the blue dashed curve represents the longitudinal tire force distribution for the initial side slip angle at the optimization starting point and the red solid curve represents the distribution for the optimized side slip angle. Similarly, figure 3.9 illustrates the optimization process of the lateral tire forces for all tires. For the front right tire it can be clearly noticed how the lateral force curve changes due to the increased slip ratio. The constraints of the constant angle between the translatory accelerations and the constant yaw acceleration is shown in figure 3.10. Figure 3.11 exemplarily compares the measured and optimized longitudinal and lateral accelerations in dependence of the vehicle velocity. The biggest differences occur at negative longitudinal accelerations, corresponding to braking maneuvers.



(a) Grip optimization results in the longitudinal direction



(b) Grip optimization results in the lateral direction



(c) Reference vehicle velocity with throttle and brake pressure signal and their characteristic points

Figure 3.7.: Grip optimization results shown in a grip limited sector

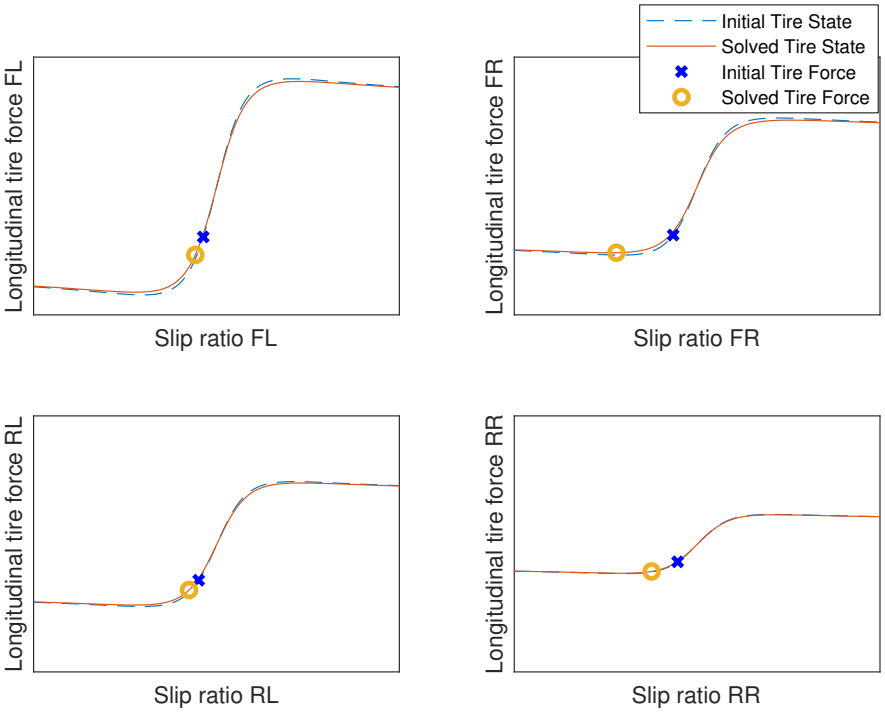


Figure 3.8.: Initial and optimized longitudinal tire states

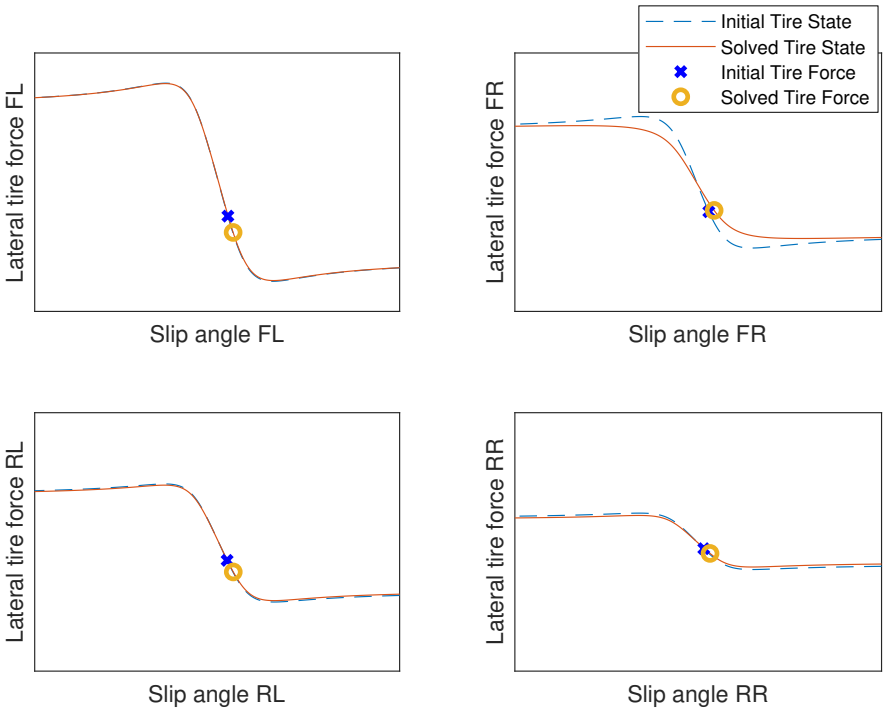
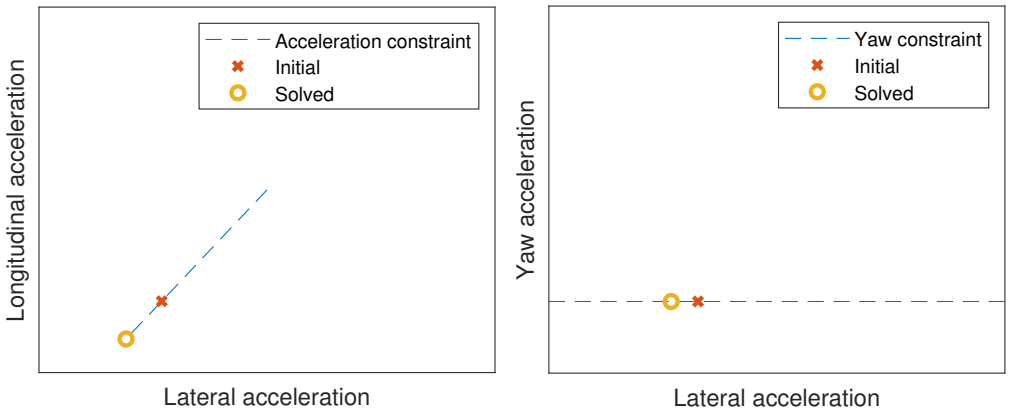


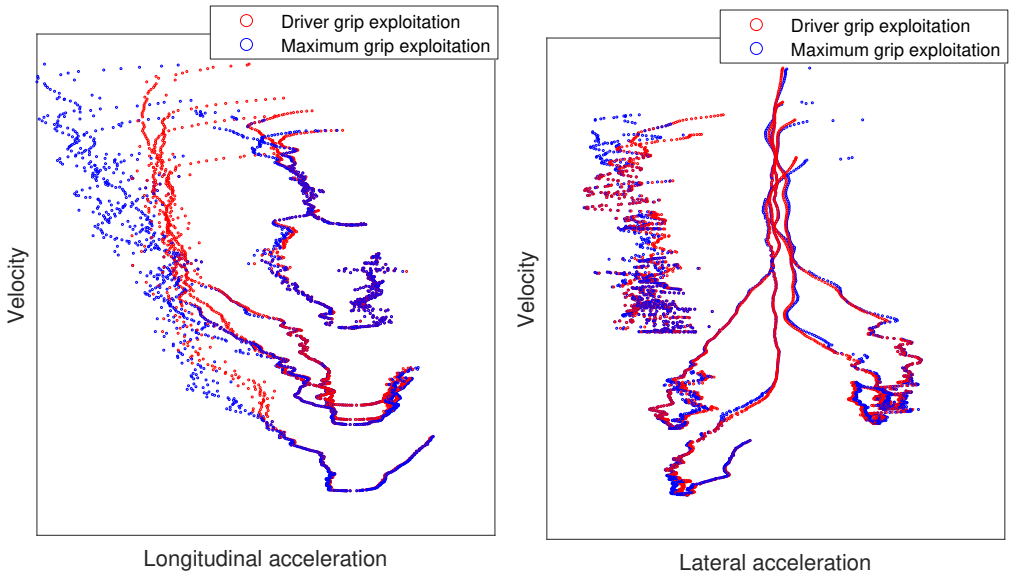
Figure 3.9.: Initial and optimized lateral tire states



(a) Acceleration direction constraint

(b) Yaw angular acceleration constraint

Figure 3.10.: Grip optimization acceleration constraints



(a) Longitudinal acceleration vs. vehicle velocity

(b) Lateral acceleration vs. vehicle velocity

Figure 3.11.: Grip optimized vehicle accelerations compared to measured vehicle accelerations

3.2.2. Grip Optimizer Metrics

Every metric with more than ten percent of NaN values is not considered for the driver classification. Figure 3.12 shows the amount of NaN in the corresponding metrics. To use a shift metric for the driver classification, the NaN values of $rdShftRel$ are replaced with a zero. As the apex distance is only available for a few DIL simulator sessions, the apex metrics $aVApex$, $xApexMin$ and $sApexRel$ contain 100% NaN values for real vehicle data of racing series A.

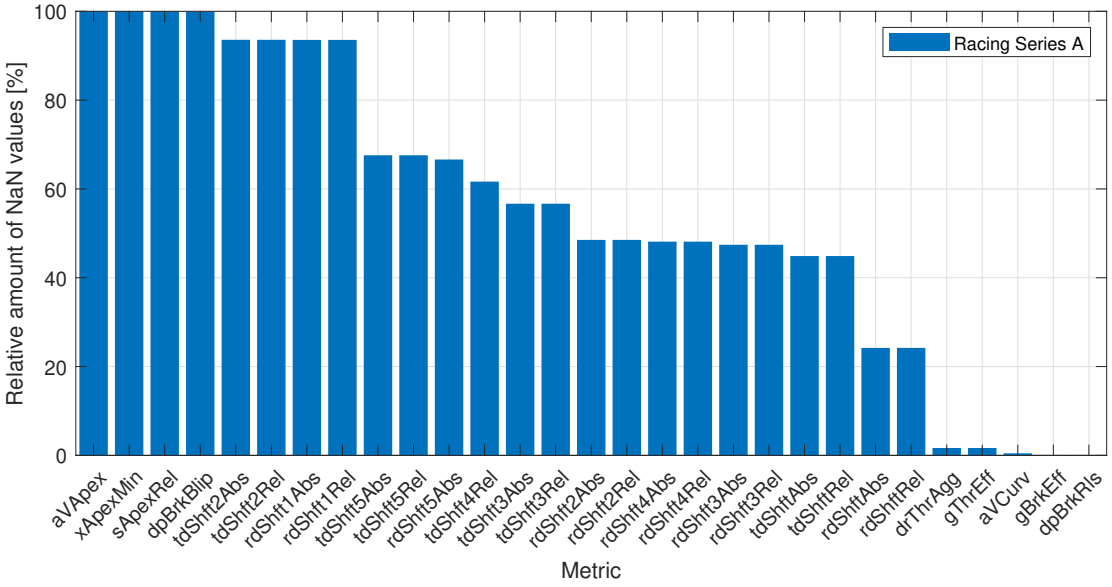


Figure 3.12.: Occurrences of NaN's in the metric values of simulator data in racing series A recorded during one season

Figure 3.13 represents the spearman correlation factors of the grip optimization metrics related to the rest of the metric set. Most of the brake, path curvature and steering metrics show relative high correlations to the entry grip optimization metrics. The rolling distance $sRollRel$ correlates with the grip optimization mid metrics and most of the throttle metrics show relative high correlation factors for grip optimization exit metrics. In figure 3.14, as expected, the grip optimization absolute metrics correlate with the corresponding relative measures. The correlations with the statistical metrics are represented in figure A.7 and A.8, where similar observations can be seen as for correlations to the throttle, brake and steering metrics. Driver metrics with performance correlations higher than $|\pm 60|\%$ are listed in table 3.1.

One grip optimization metric is within the first twenty driver metrics in terms of the decision tree importance factors, as shown in figure 3.15 with the importance factors of the residual grip optimization metrics. The ranking according to the importance factor is represented in table 3.2 where the relative entry metrics show the highest and the

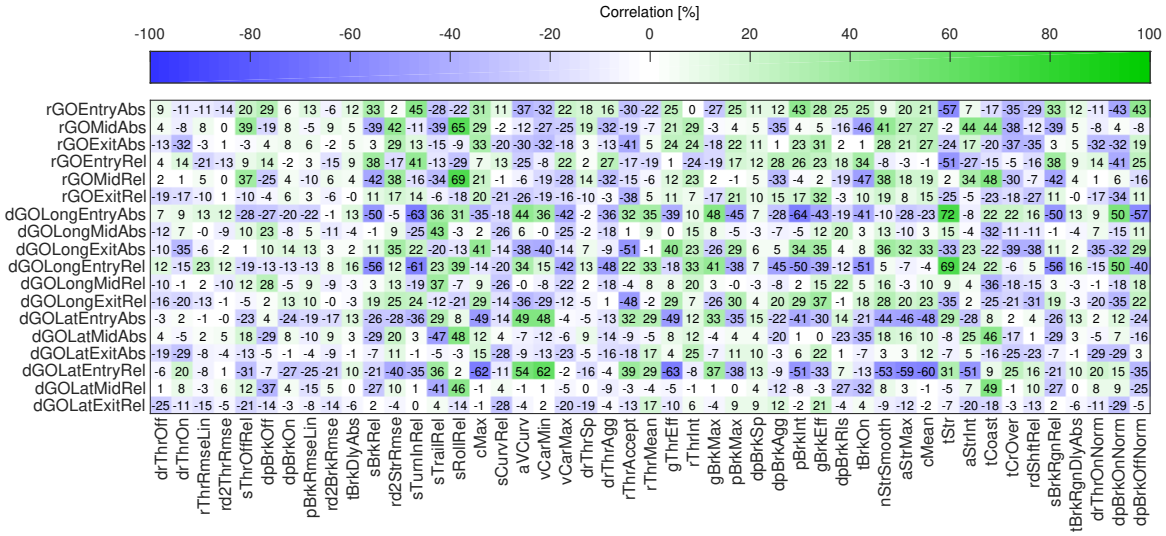


Figure 3.13.: Spearman correlation matrix of metrics valid for DIL simulator data of racing series A except the statisticals

Table 3.1.: Metric with absolute correlation factors higher than 60%

Metric	Correlations [%]	Correlated Performance Metric
sTurnInRel	-63	dGOLongEntryAbs
sRollRel	69	rGOMidRel
cMax	-62	dGOLatEntryRel
vCarMin	62	dGOLatEntryRel
gThrEff	-63	dGOLatEntryRel
pBrkInt	-64	dGOLongEntryAbs
tStr	69	dGOLongEntryAbs
aStrStd	-63	dGOLatEntryRel
aStrRange	-63	dGOLatEntryRel

absolute mid and exit metrics the lowest importances.

The grip optimization entry metrics, except $dGOLongEntryAbs/-Rel$, show the highest F-values in the variance analysis, illustrated in figure 3.16.

The probability distributions for the grip optimization metrics for four drivers are represented in figure 3.17. The number below the driver name indicates the amount of data available corresponding to the driver and the special colored bin represents the outliers with the binheight as the sum of the probabilities of the outliers. Upper outliers are defined as data above the third quartile plus three and a half times the interquartile range of metric values for one driver. Identically the lower outlier threshold is defined as the first quartile minus three and a half times the interquartile range.

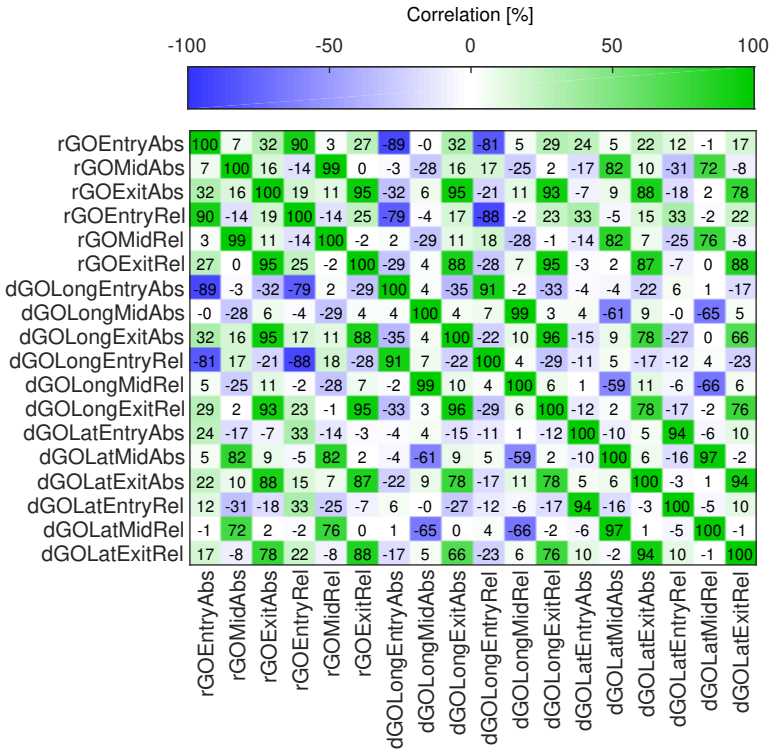


Figure 3.14.: Spearman correlation matrix of grip optimization metrics for DIL simulator data of racing series A

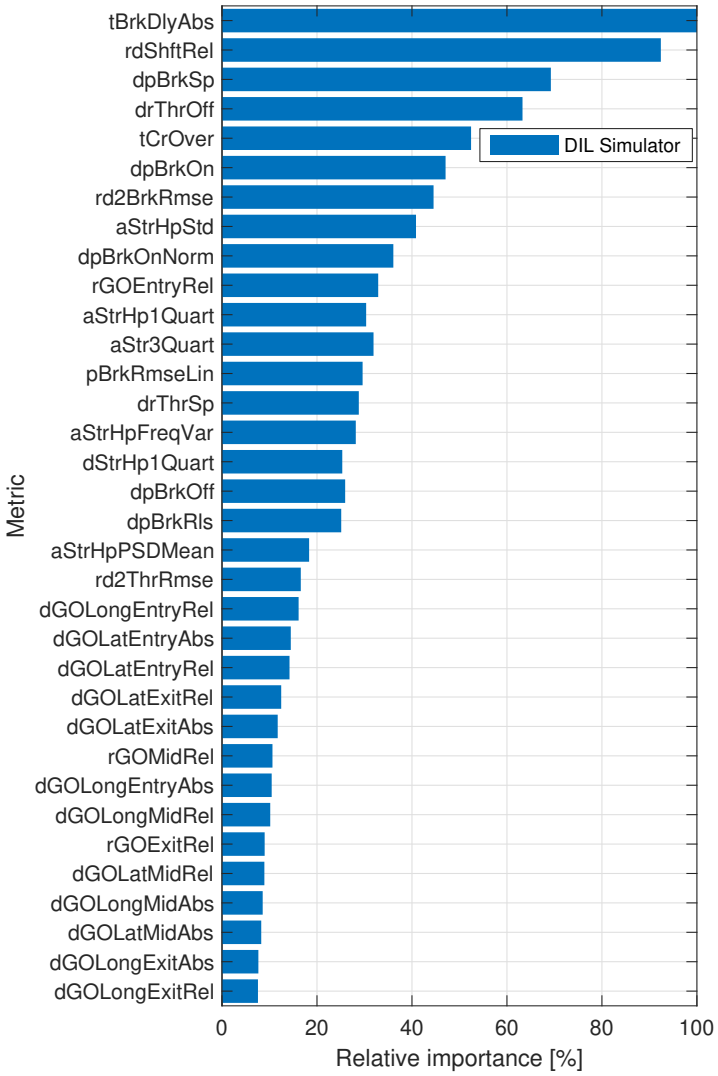


Figure 3.15.: Decision tree importance factors of simulator data of one season for racing series A

Table 3.2.: Decision tree importance factors for the metrics derived from the grip optimization method

#	Metric	Importance [%]
15	rGOEntryRel	20
21	dGOLongEntryRel	16
30	dGOLatEntryAbs	14
31	dGOLatEntryRel	14
38	dGOLatExitRel	12
43	dGOLatExitAbs	12
50	rGOMidRel	11
51	dGOLongEntryAbs	10
56	dGOLongMidRel	10
71	rGOExitRel	9
73	dGOLatMidRel	9
79	dGOLongMidAbs	9
83	dGOLatMidAbs	8
93	dGOLongExitAbs	8
94	dGOLongExitRel	8

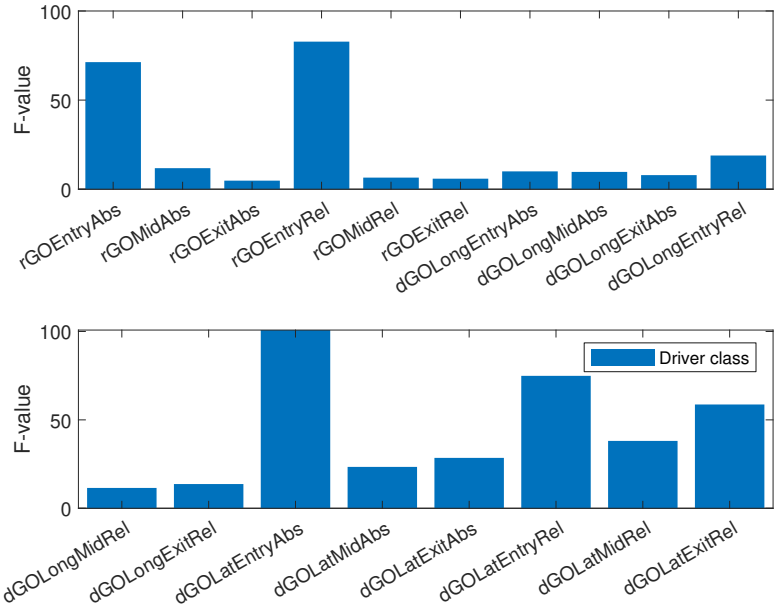
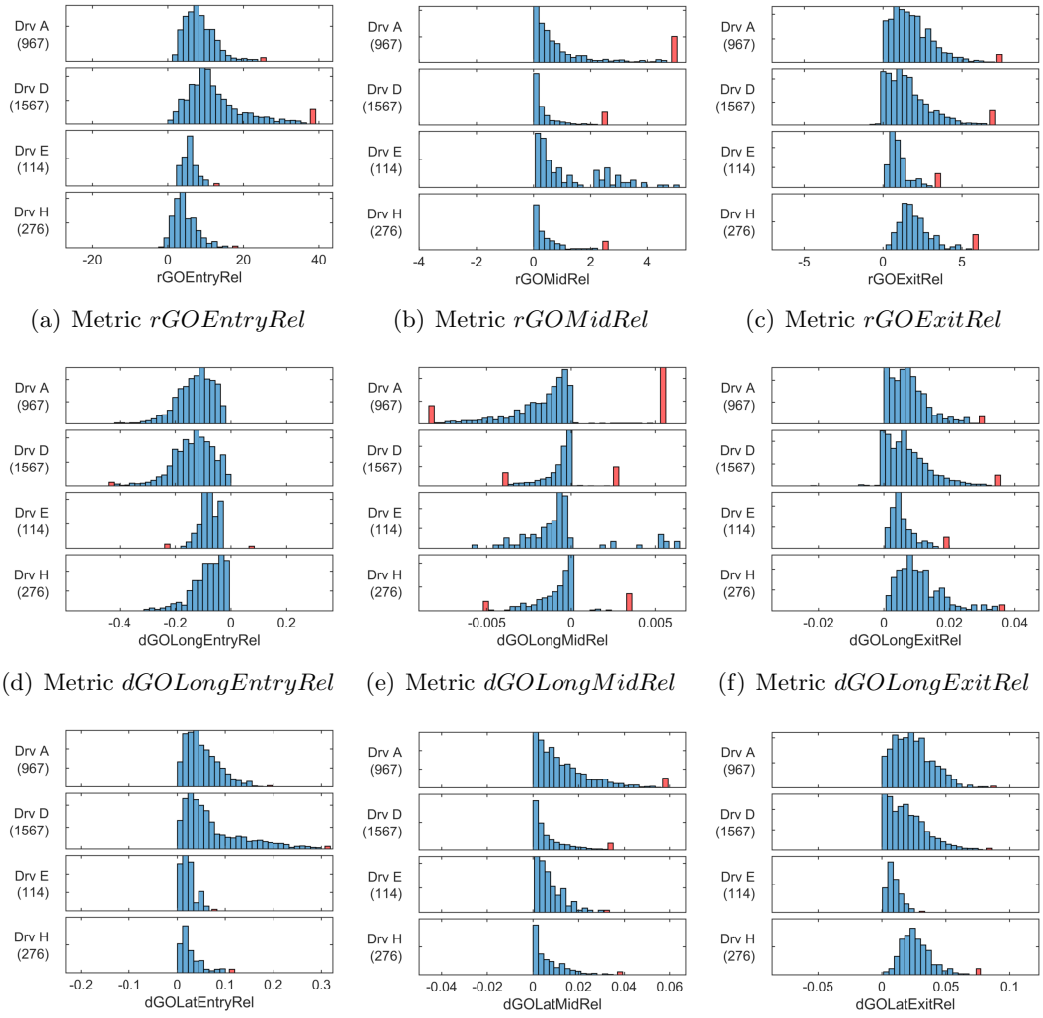


Figure 3.16.: F-values for the driver and track class of DIL simulator data of racing series A



(a) Metric *rGOEntryRel* (b) Metric *rGOMidRel* (c) Metric *rGOExitRel*

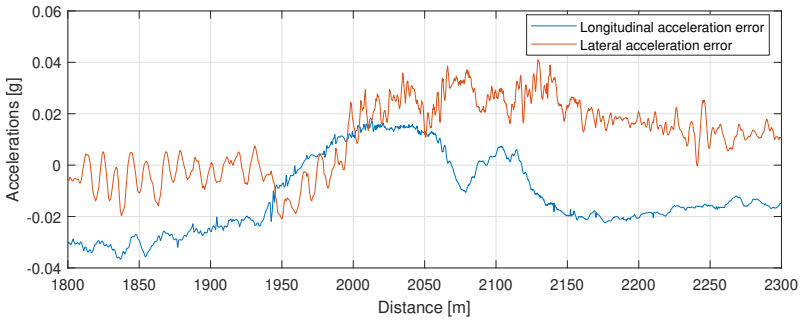
(d) Metric *dGOLongEntryRel* (e) Metric *dGOLongMidRel* (f) Metric *dGOLongExitRel*

(g) Metric *dGOLatEntryRel* (h) Metric *dGOLatMidRel* (i) Metric *dGOLatExitRel*

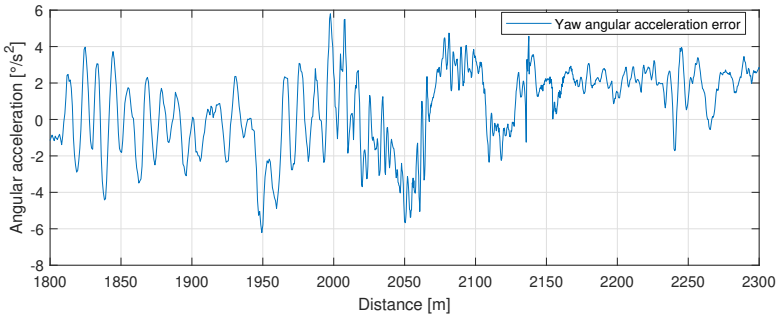
Figure 3.17.: Histogram probability distributions among the drivers of the relative grip optimization metrics for DIL simulator data of racing series A

3.2.3. Validation

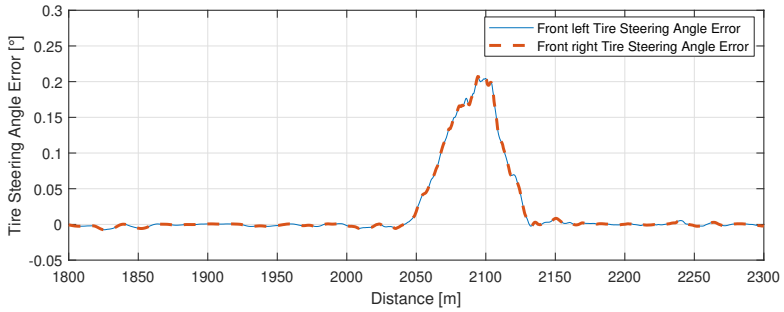
Figure 3.18 represents the differences of the accelerations resulting from the grip optimization vehicle model compared to the DIL simulator vehicle model during a braking and cornering maneuver. In the lateral and longitudinal acceleration signals maximum deviations of $e_{\mathrm{aGO}_{\mathrm{Lat}}} \approx 0.03\mathrm{g}$ and $e_{\mathrm{aGO}_{\mathrm{Long}}} \approx 0.03\mathrm{g}$ can be observed. No significant correlation between the tire steering angle error and the acceleration errors can be noticed.



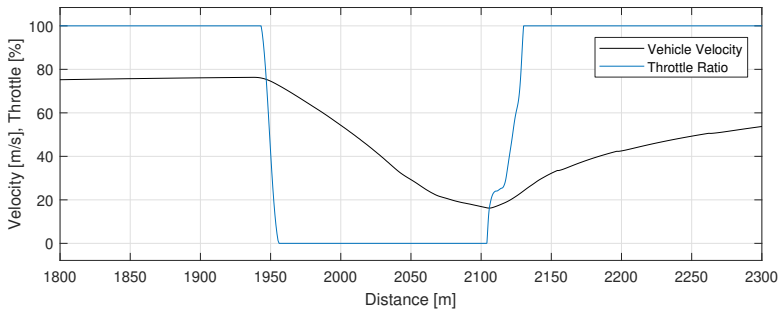
(a) Error of the accelerations calculated from the grip optimization vehicle model compared to the accelerations calculated in the DIL simulator



(b) Error of the yaw acceleration calculated from the grip optimization vehicle model compared to the yaw acceleration calculated in the DIL simulator



(c) Error of the tire steering angles calculated from the grip optimization vehicle model compared to the DIL simulator



(d) Reference vehicle velocity with throttle application

Figure 3.18.: Deviation between grip optimization and the DIL simulator

3.2.4. Error Analysis

Due to the deviation of the accelerations calculated by the vehicle model used for the grip optimization method and the vehicle model from the DIL simulator, the grip optimization metrics are afflicted with errors. In table 3.3 an exemplary error calculation is shown with the maximum acceleration deviations of section 3.2.3. The error of the metrics is depending on the duration of the cornering phase t_0, t_1, the occured accelerations a_x, a_y and on the duration time of the grip limited sector t_sector. The exemplary values in the error table are calculated with constant signal values of $a_\mathrm{x} = -2\mathrm{g}$, $a_\mathrm{y} = 0.5\mathrm{g}$, $t_1 - t_0 = 1\mathrm{s}$ and $t_\mathrm{sector} = 3\mathrm{s}$ which could represent an entry cornering phase. The resultant error $e_{\mathrm{a_{GO}}}$ and acceleration a is calculated according to (3.1) and (3.2):

$$e_{\mathrm{a_{GO}}} = \sqrt{e^2_{\mathrm{a_{GO_{Long}}}} + e^2_{\mathrm{a_{GO_{Lat}}}}}, \tag{3.1}$$

$$a = \sqrt{a_\mathrm{x}^2 + a_\mathrm{y}^2}. \tag{3.2}$$

Table 3.3.: Error analysis for grip optimization metrics (*Phs* is a placeholder for *entry*, *mid*, *exit*)

Metric	Error	Exemplary value
rGO*Phs*Abs	$\sum_{t=t_0}^{t_1} \frac{e_{\mathrm{a_{GO}}}}{a} \cdot \frac{100}{f_\mathrm{s}}$	$\pm 2.06\% \cdot s$
rGO*Phs*Rel	$\sum_{t=t_0}^{t_1} \frac{e_{\mathrm{a_{GO}}}}{a} \cdot \frac{100}{f_\mathrm{s}} \cdot \frac{1}{t_\mathrm{sector}}$	$\pm 0.69\%$
dGO*Phs*LongAbs	$e_{\mathrm{a_{GO_{Long}}}} (t_1 - t_0)$	$\pm 0.03\mathrm{g} \cdot s$
dGO*Phs*LongRel	$e_{\mathrm{a_{GO_{Long}}}} \frac{(t_1 - t_0)}{t_\mathrm{sector}}$	$\pm 0.01\mathrm{g}$
dGO*Phs*LatAbs	$e_{\mathrm{a_{GO_{Lat}}}} (t_1 - t_0)$	$\pm 0.03\mathrm{g} \cdot s$
dGO*Phs*LatRel	$e_{\mathrm{a_{GO_{Lat}}}} \frac{(t_1 - t_0)}{t_\mathrm{sector}}$	$\pm 0.01\mathrm{g}$

The exemplary errors are not values of a completely negligible size, but also not of a magnitude which distortes interpretable results. Furthermore, for the task of comparing driving styles of different drivers, tendencies of the grip optimization results are of a higher interest than absolute values.

3.2.5. Discussion

The metrics derived by the grip optimization algorithm seem to be not very suitable for the driver classification according to their decision tree importance factors. The long term intention of describing driving styles with objective metrics is to improve the performance by adapting the vehicle properties to suit the driver metrics. However,

the performance relevance of the metrics with the highest importances for the driver classification task is difficult to interpret. As already mentioned, grip optimization measures give information of the vehicle performance exploitation by the drivers, thereupon correlations to other metrics to obtain their performance relation are of a high interest. For example the downshift metric $rdShftRel$ has clearly a high importance for the driver classification, but it is difficult to interpret the metric regarding the performance relevance. Looking at the correlation matrix of the grip optimization metrics, the downshift metric has a correlation of -38% to the longitudinal acceleration corner exit metric $dGOLongExitAbs$ which indicates that if drivers have an aggressive downshift behaviour, they are able to extract more acceleration out of the vehicle during the corner exit phase. It seems that the biggest performance differences in driving styles mostly occur in the corner entry phase during braking maneuvers. The metrics with the highest correlations are influenced by the corner phase length and the trajectory chosen by the drivers which indicates that finding an optimal racing line has a high importance regarding the performance. Furthermore, it can be noticed that for example some metrics correlate negatively to the entry and show a positive correlation factor for the exit. Hence, for observations on the overall cornering performance all the cornering phases have to be taken into account and analyzed together.

3.3. Final Metric Selection and Evaluation

In this chapter the driver metrics are observed seperately based on three datasets: racing series A, racing series B and both racing series combined.

For the driver classification task metrics with more than ten percent NaN values are removed. Then, after the analysis with a spearman correlation, every metric with correlations higher than $|\rho_s| = 85\%$ is considered to be removed. Taking additionally the decision tree importance factors into account and analyzing subjectively the importances of the metrics for their ability to describe driving styles, a list of initially dropped metrics can be defined in table A.1. For the driver selection method only the first 20 metrics according to the decision tree importance factors of the residual metrics are used. Furthermore, observations on the F-values of ANOVA and the normalized probability distributions in histograms are presented. The special colored bins in the histogram plots represent the outliers with the amount of outliers as bin height. Outliers are defined identically as in section 3.2.2.

3.3.1. Racing Series A

The highest amount of NaN values are containing the shift metrics (figure 3.19). To still obtain information of at least one shift metric, the NaN values of $rdShftRel$ are replaced with a zero and the other shift metrics are removed. The metrics $gBrkEff$ to $dpBrkOffNorm$ contain less than 1% NaN values.

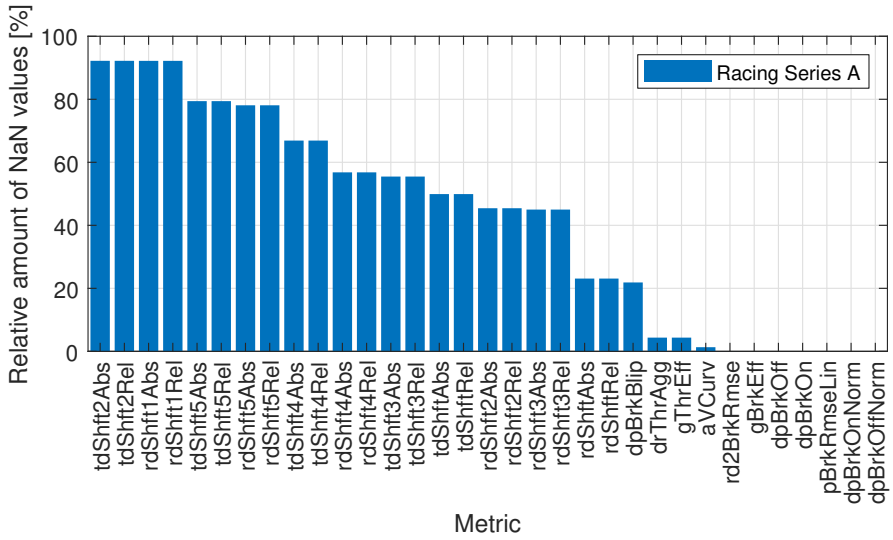


Figure 3.19.: Occurrences of NaN's in the metric values of racing series A

In figure 3.20 high correlations between the steering and the path curvature metrics can be seen. As expected, for racing series A the normalized metrics show a 100% correlation with the respecting initial metric: $drThrOn$ and $drThrOnNorm$, $tBrkDlyAbs$ and $tBrkRgnDlyAbs$, $sBrkRel$ and $sBrkRgnRel$. Only for racing series B these metrics differ from the initial definitions. The sector phases (entry, mid, exit) influence each other, therefore some brake metrics correlate negatively with steering metrics and positively with throttle metrics. For example the brake pressure integral $pBrkInt$ shows a correlation of -70% on the steering time $tStr$ and 57% on the positive throttle gradient $drThrAgg$. The average throttle gradient $drThrSp$ remains relatively uncorrelated because of the high amount of zero values if the throttle is applied constantly. The statistical metrics (figure A.1 and A.2) show partially relative high correlations with the other metric set. Most of the statistical metrics based on the steering wheel angle correlate with the respective measure based on the steering wheel angle rate.

The first 20 driver metrics according to their decision tree importance factors are illustrated in figure 3.21. The off-throttle gradient $drThrOff$ and the on-brake gradient normalized by the maximum brake pressure within a sector $dpBrkOnNorm$ show clearly the highest importances. The influence of the normalization can be noticed when comparing the importances of the normalized on-brake gradient (79%) with the value of the not normalized measure $dpBrkOn$ (46%). The brake metrics represent the highest importances with five of the first seven driver metrics being related to the brake pressure signal.

The confusion matrices in figure 3.22 represent the absolute and relative amount of predictions among the driver class. The highest accuracies are shown for driver A and D with 92% and 93%. The lowest relative accuracy is achieved for driver G with 59%.

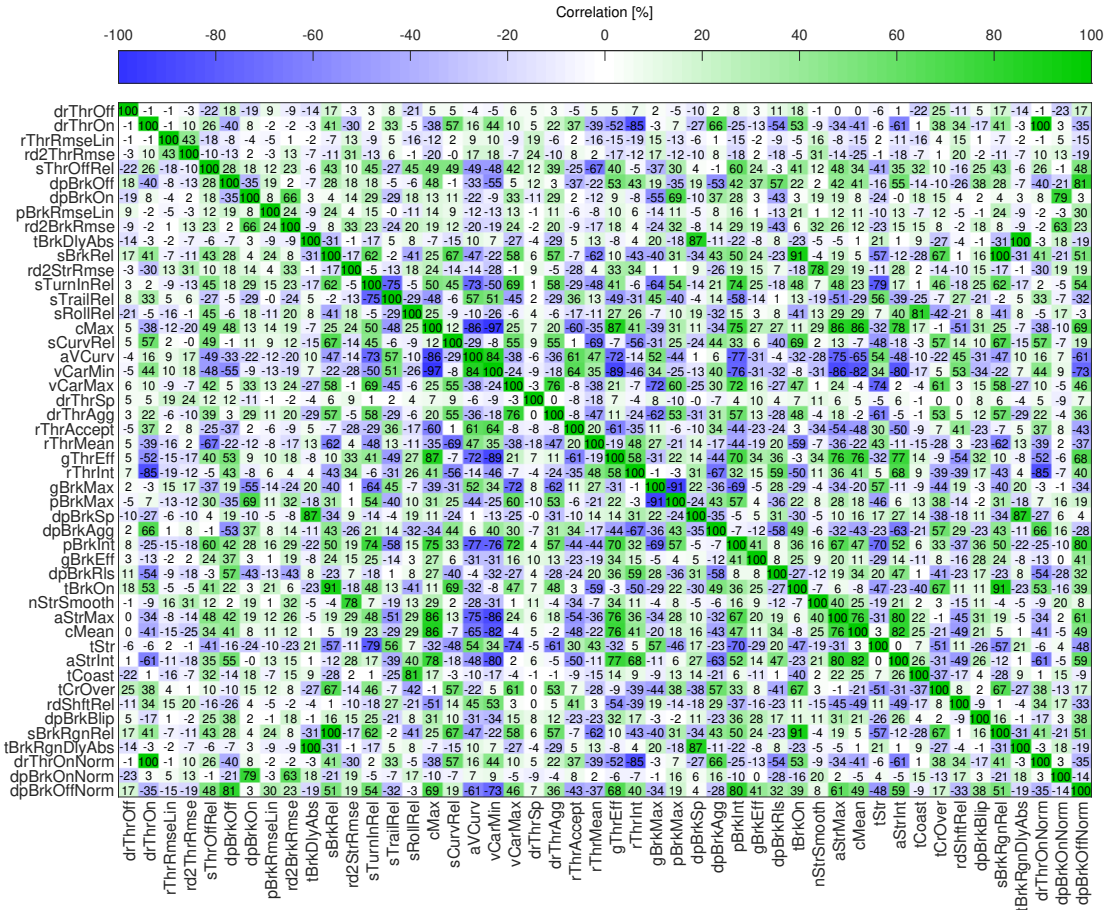


Figure 3.20.: Spearman correlation matrix of metrics valid for racing series A except the statisticals

The mean prediction accuracy among the drivers lies at 79%.

Figure 3.23 illustrates the F-values of the driver and the track class. The highest F-values in the driver class with relatively low values in the track class correspond to the metrics with the highest decision tree importance factors of figure 3.21.

The histograms in figure 3.24 represent the probability distributions of the metrics for the drivers of racing series A. Driver D and H have the slowest off-throttle gradient $drThrOff$. The normalized on-brake gradient $dpOnBrkNorm$ is the highest for driver A and I. Driver A has the least deviated brake delay time $tBrkDlyAbs$ whereas driver B shows the biggest range of values, indicating that driver B may slightly apply the brake pedal even before the braking point. Driver B and H have a distinct behaviour at the mean brake speed metric $dpBrkSp$. Figure 3.25 compares the metric values for a specific high speed sector of a track where the driving style differences can be noticed even clearer.

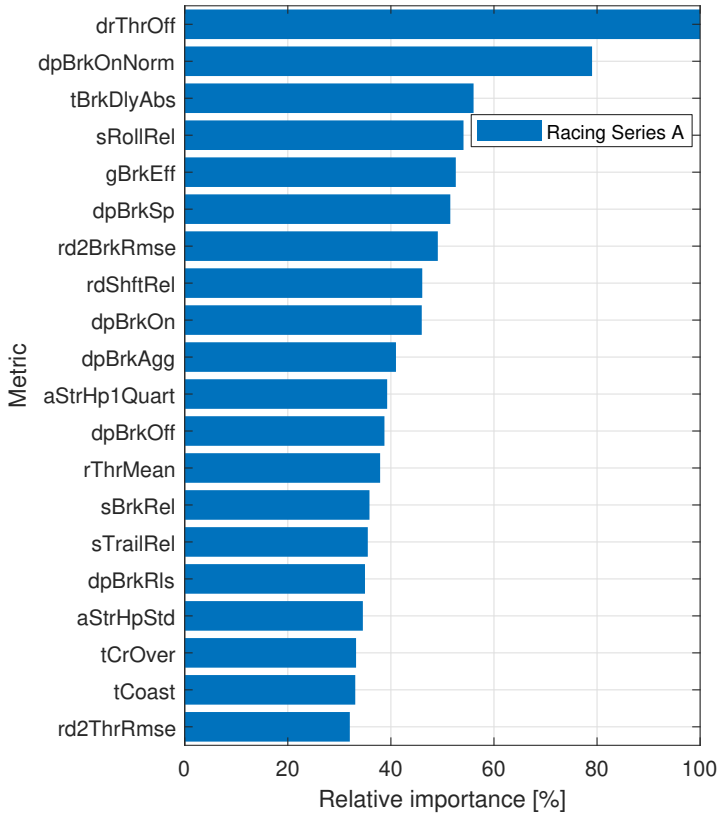
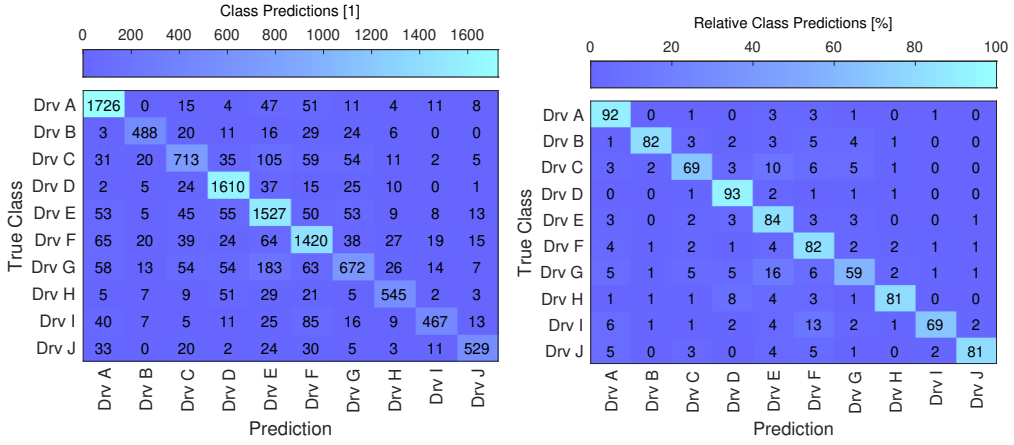


Figure 3.21.: Decision tree importance factors of racing series A



(a) Absolute predictions

(b) Relative predictions

Figure 3.22.: Driver class predictions of racing series A

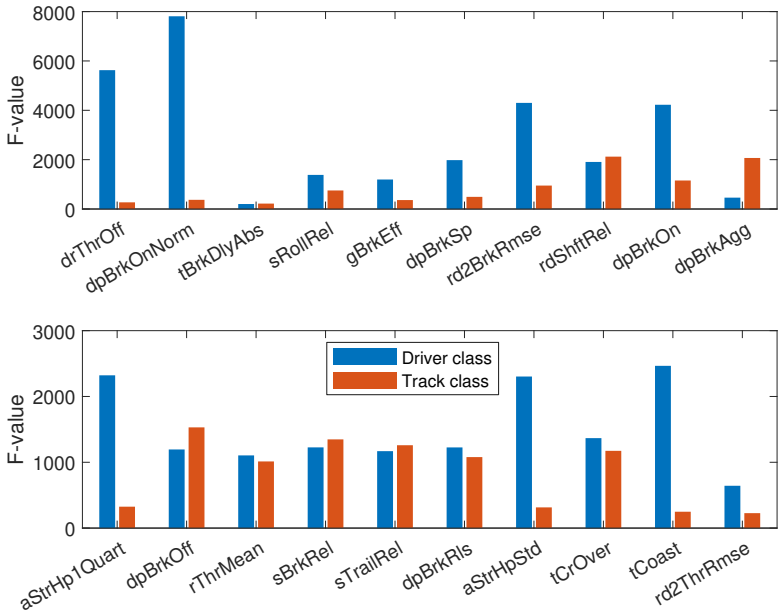
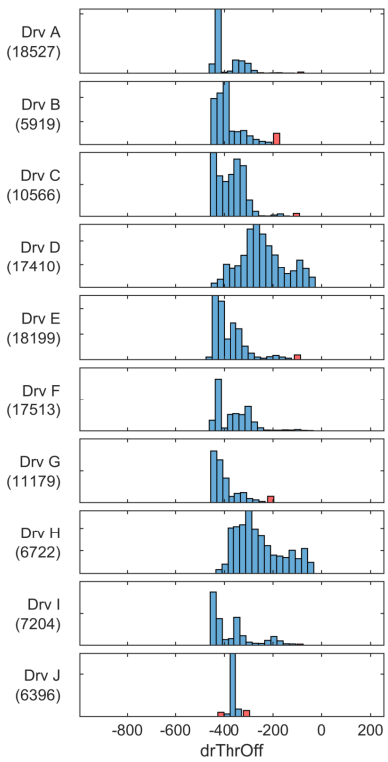
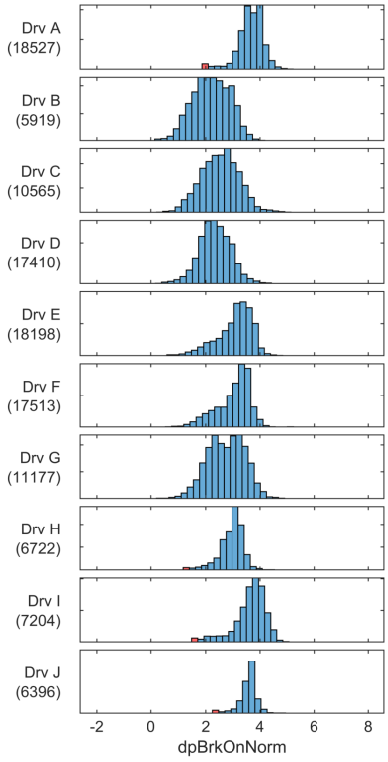


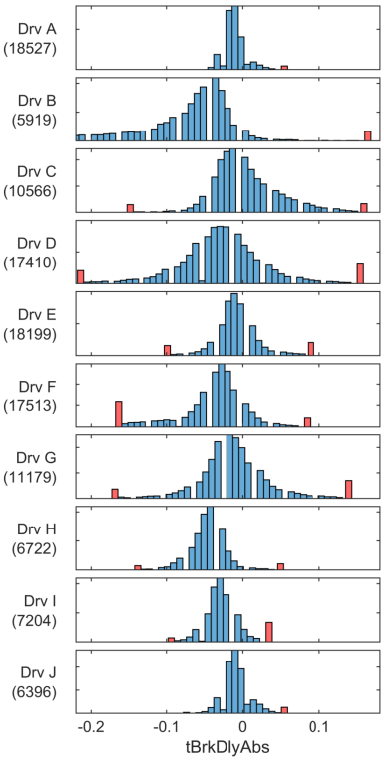
Figure 3.23.: F-values for the driver and track class of racing series A



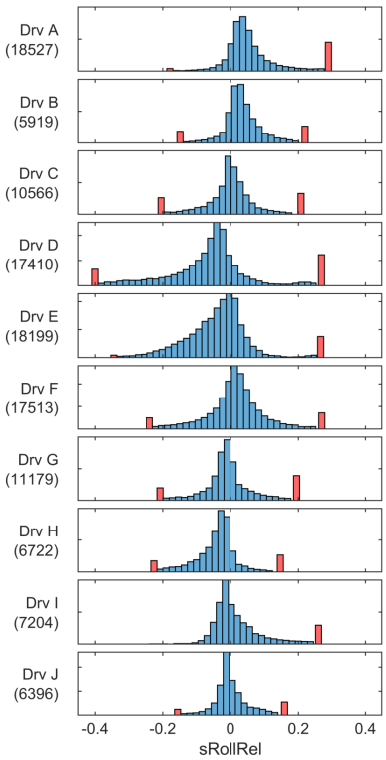
(a) Metric *drThrOff*



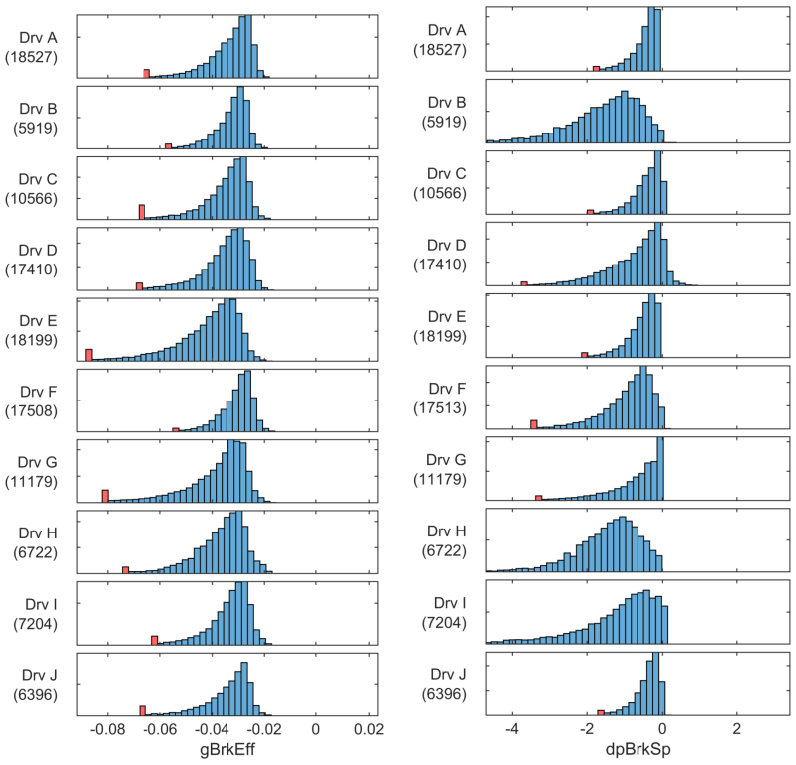
(b) Metric *dpBrkOnNorm*



(c) Metric *tBrkDlyAbs*

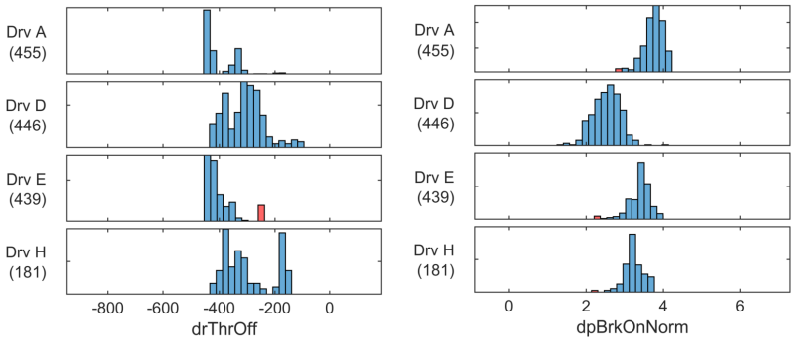


(d) Metric *sRollRel*



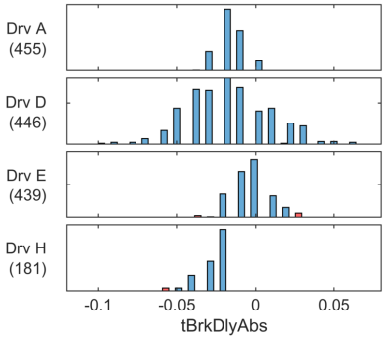
(e) Metric $gBrkEff$ (f) Metric $dpBrkSp$

Figure 3.24.: Histogram probability distributions among the drivers for racing series A

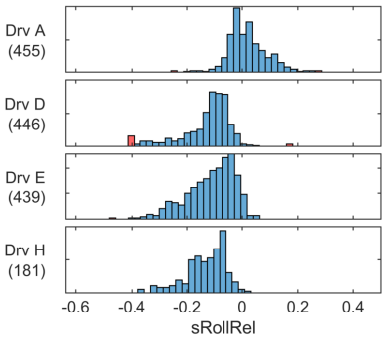


(a) Metric $drThrOff$

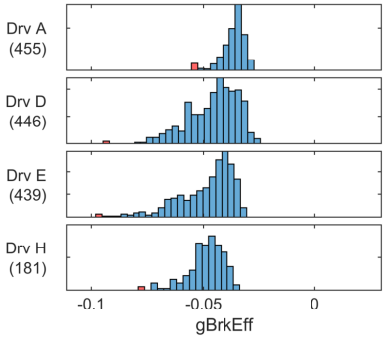
(b) Metric $dpBrkOnNorm$



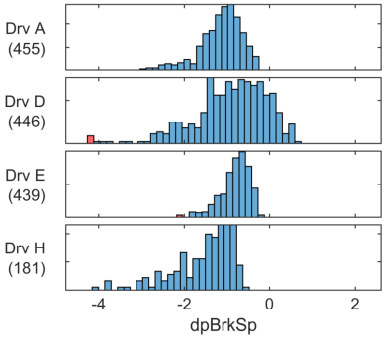
(c) Metric $tBrkDlyAbs$



(d) Metric $sRollRel$



(e) Metric $gBrkEff$



(f) Metric $dpBrkSp$

Figure 3.25.: Histogram probability distributions for racing series A at sector 1 of track 1

3.3.2. Racing Series B

Observing the NaN values of the metrics in figure 3.26 it can be seen that the recuperation and coast phase are not detected regularly. In the data preprocessing the data gets filtered according to their best laptimes. If the drivers apply the coasting and recuperation technique, they decelerate slower than if they would at braking with the mechanical brake system. Therefore, most of the laps with recuperation and coasting information may be filtered out. The metrics from $gBrkEff$ to $dpBrkOffNorm$ contain less than 1% of NaN values.

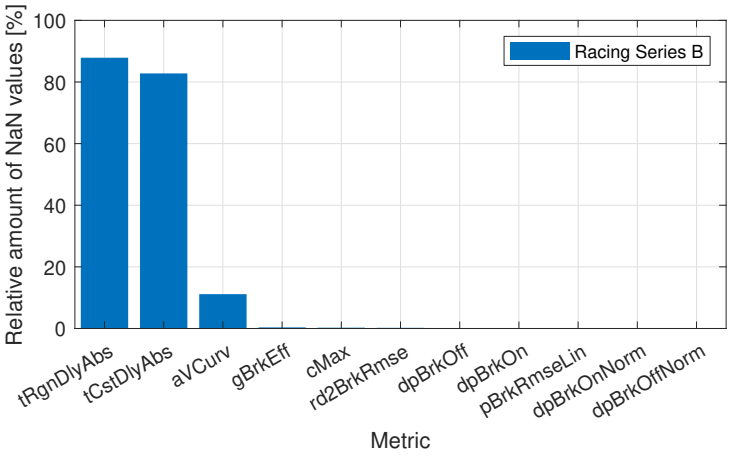


Figure 3.26.: Occurrences of NaN's in the metric values of racing series B

The spearman correlation matrix of figure 3.27 shows similar tendencies as the one of racing series A (figure 3.20) with noticeable differences in the brake metrics. Further differences to racing series A can be seen by observing the correlation information of the statistical driver metrics in figure A.1 and A.2 where most of the metrics from racing series B show slightly lower correlation values.

Figure 3.28 represents the decision tree importance factors for racing series B. The off-throttle gradient $drThrOff$ and the normalized on-brake gradient $dpBrkOnNorm$ importances have decreased significantly to 51% and 49% compared to 100% and 79% for racing series A. Again the driver metrics based on the brake pedal represent the group with the highest importances with three brake metrics among the first four metrics. The mean brake speed metric $dpBrkSp$ has the highest importance (100%) for the decision tree classification compared to 52% for racing series A. The second highest importance is represented by the brake delay $tBrkRgnDlyAbs$ with 51% (56% for racing series A). The big differences in the decision tree importance factors between the first two metrics indicates that the classifier detects one or more drivers by a specific behaviour.

In figure 3.29 the confusion matrices for the driver classification of racing series B are presented. Driver B was predicted with a relative accuracy of 94% and also contains the

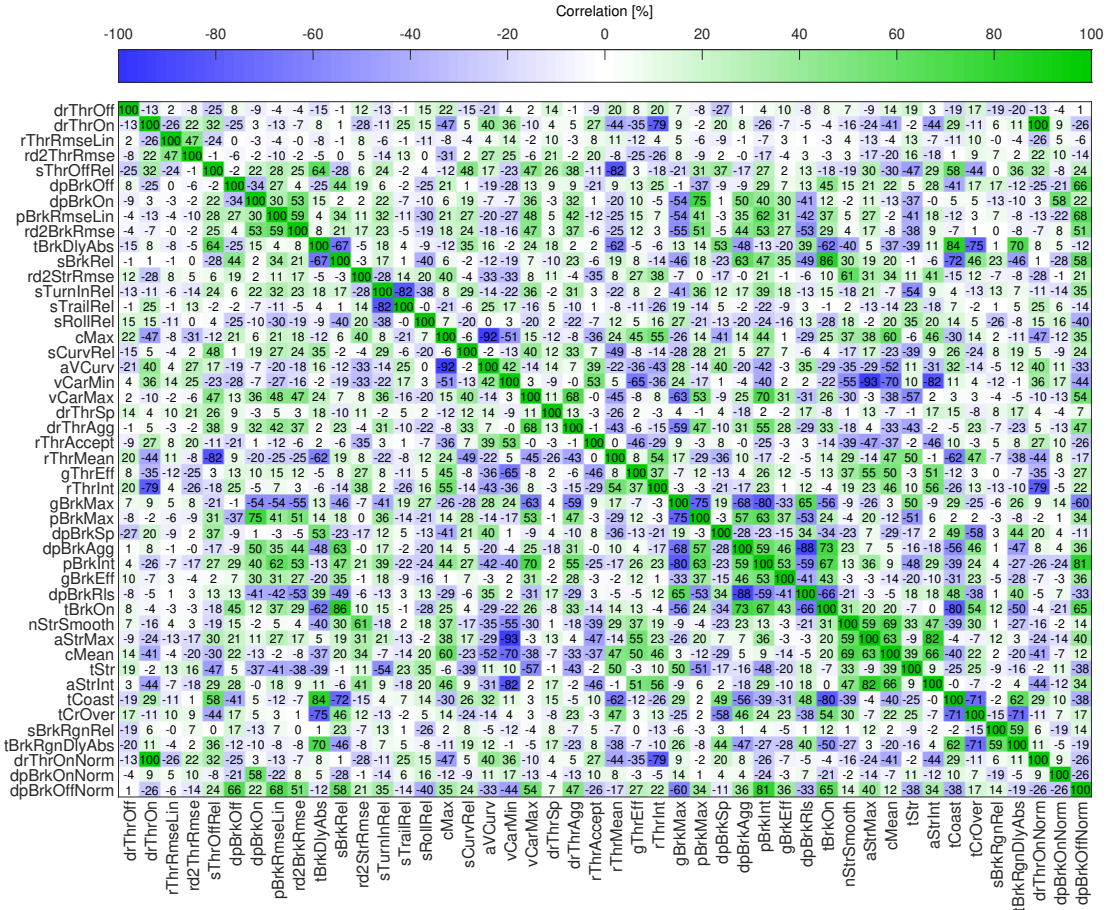


Figure 3.27.: Spearman correlation matrix of metrics valid for racing series B except the statisticals

highest amount of data. The lowest prediction accuracy occured for driver B with 76% and the mean accuracy for the dataset of racing series B equals to 82%.

The F-values of the *dpBrkSp* and $gBrkEff$ metric are high for the driver and low for the track class (figure 3.30). The throttle efficiency $gThrEff$ metric has the highest values for the driver class, but also relative high track class values. The relative rolling distance *sRollRel* and the first quartile of the steering wheel angle $aStr1Quart$ show relative high decision tree importance factors of 35% and 29% but have low F-values for the driver class and higher F-values for the track class. For racing series B, the F-value distribution does not correspond to the decision tree importance factors as well as for racing series A.

In the histogram probability distributions of figure 3.31 the biggest differences among the drivers can be seen in the *dpBrkSp* where driver K shows a distinct behaviour. Looking at the throttle metrics driver L represents slightly distinct properties compared to the

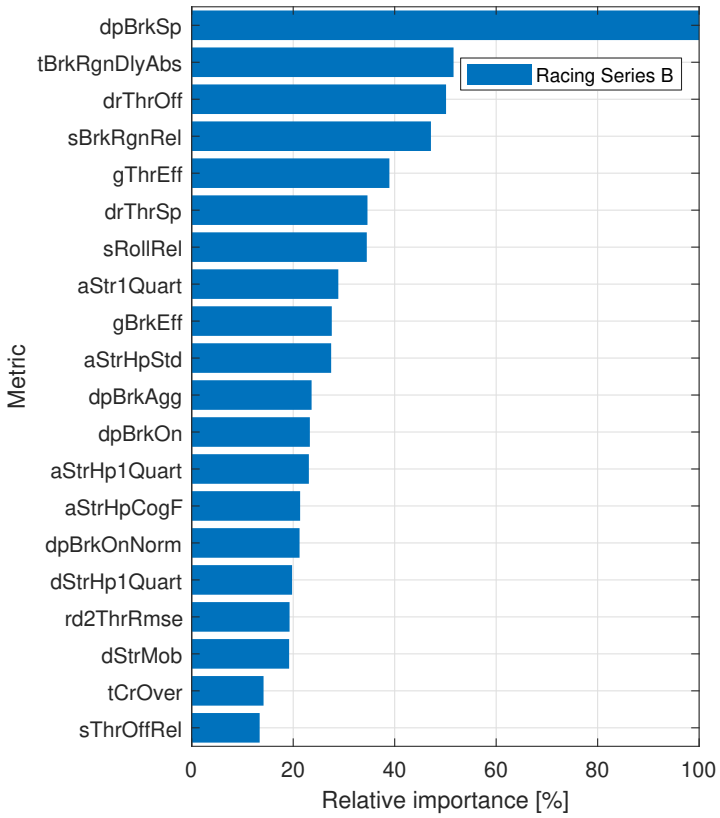
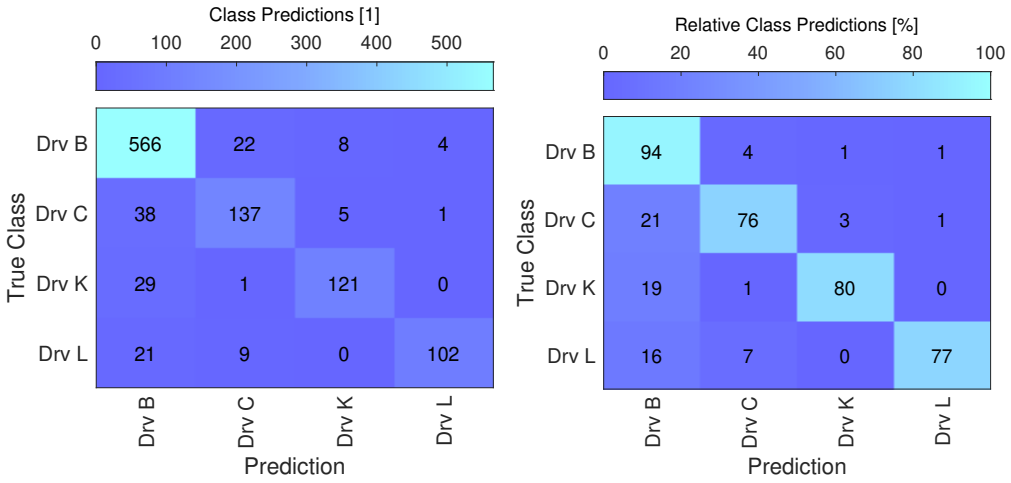


Figure 3.28.: Decision tree importance factors of racing series B

other drivers.



(a) Absolute predictions (b) Relative predictions

Figure 3.29.: Driver class predictions of racing series B

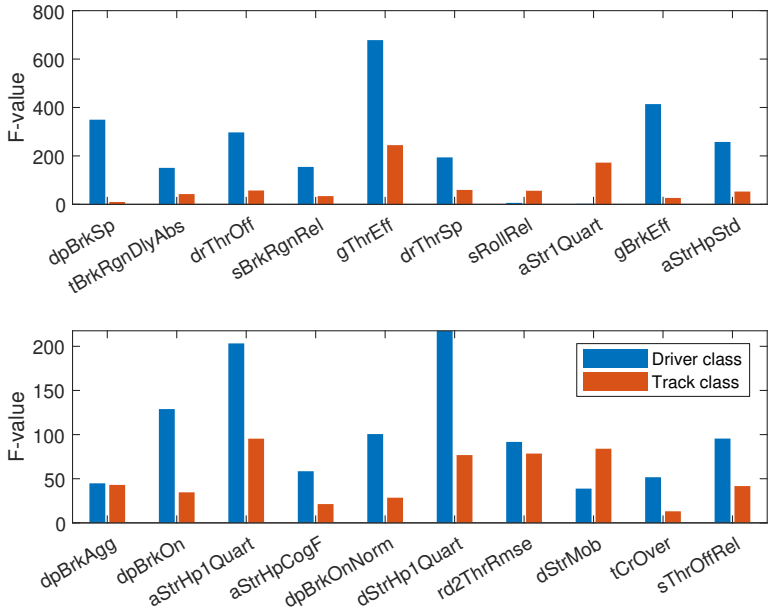
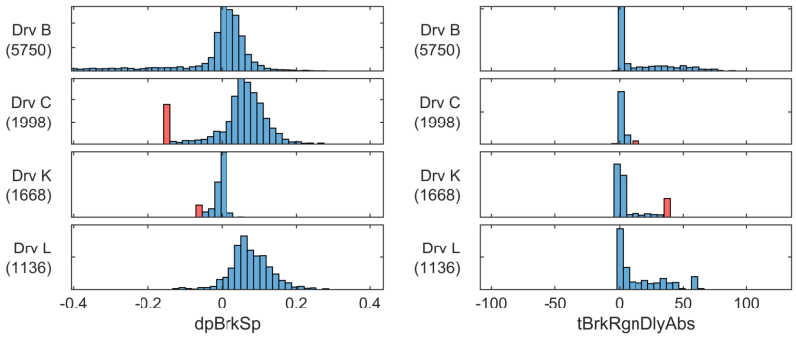
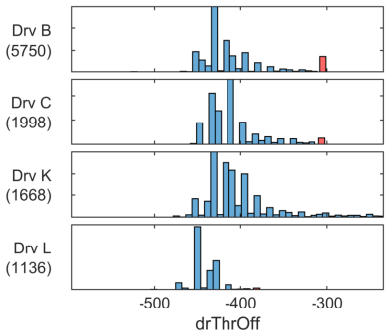


Figure 3.30.: F-values for the driver and track class of racing series B

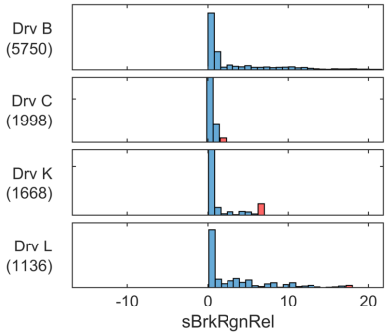


(a) Metric *dpBrkSp*

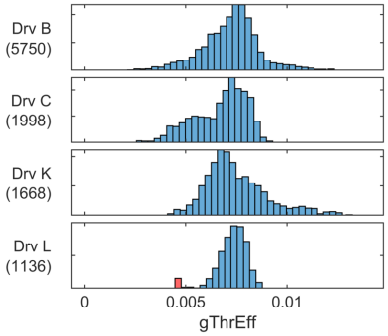
(b) Metric *tBrkRgnDlyAbs*



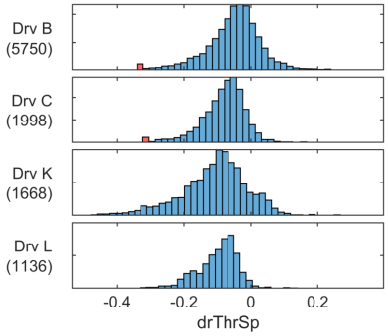
(c) Metric *drThrOff*



(d) Metric *sBrkRgnRel*

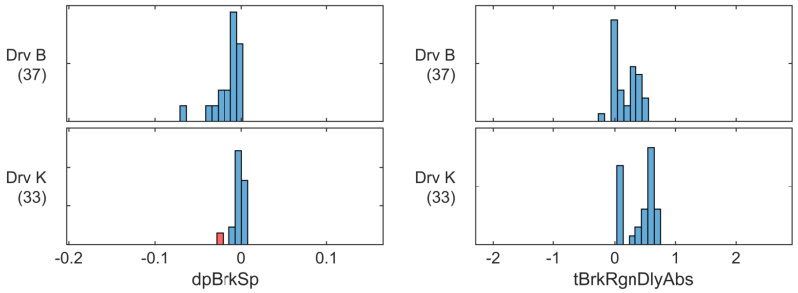


(e) Metric *gThrEff*



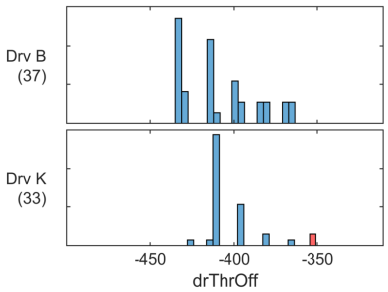
(f) Metric *drThrSp*

Figure 3.31.: Histogram probability distributions among the drivers for racing series B

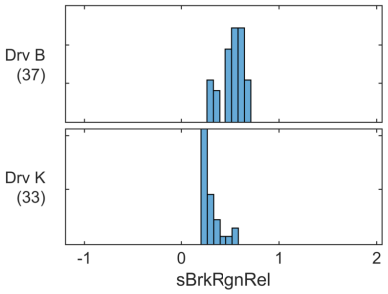


(a) Metric $dpBrkSp$

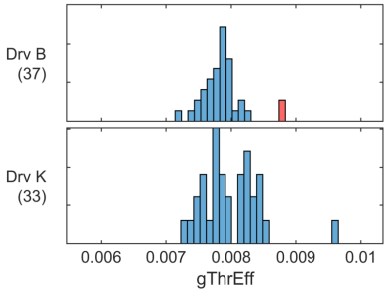
(b) Metric $tBrkRgnDlyAbs$



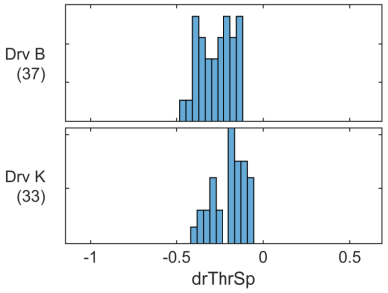
(c) Metric $drThrOff$



(d) Metric $sBrkRgnRel$



(e) Metric $gThrEff$



(f) Metric $drThrSp$

Figure 3.32.: Histogram probability distributions for racing series B at sector 1 of track 2

3.3.3. Results from both Racing Series

The results of this chapter combine those of racing series A and racing series B. Racing series A contains nearly 16 times of the amount of data that represents racing series B. Thereupon racing series A is weighted more and has a bigger influence on the results of the combined dataset, hence, figures 3.33 to 3.38 have similar tendencies as the ones of racing series A in section 3.3.1.

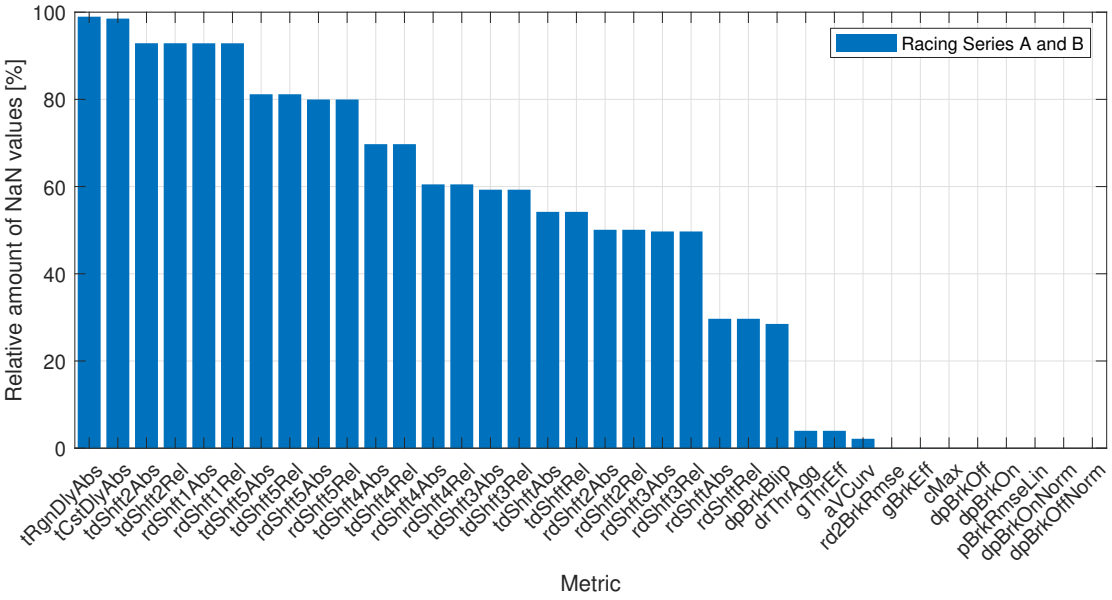


Figure 3.33.: Occurrences of NaN's in the metric values of both racing series

The decision tree importance factors (figure 3.35) are similar to the dataset of racing series A with the influence of racing series B as for example the mean brake gradient $dpBrkSp$ has an importance of 59% compared to 52% in racing series A. Compared to the seperate datasets, the importance of the normalized brake delay time $tBrkRgnDlyAbs$ increased at the combined dataset to 72%. Driver D still has the highest prediction accuracy of 93% in figure 3.36, where driver A is predicted with a slightly lower accuracy of 90% compared to 92% for racing series A. The accuracy of 87% of driver B has settled in between the dataset of racing series A (82%) and racing series B (94%). Driver G has with a relative prediction accuracy of 58% the lowest value, but was also the driver with the worst classification result for the dataset of racing series A. The combined mean prediction accuracy equals 78% compared to 82% in racing series B and 79% in racing series A. The histogram probability distributions (figure 3.38) show similar behaviours for the two drivers which competed in both racing series with slight deviations in the mean brake gradient $dpBrkSp$ and the relative rolling distance $sRollRel$.

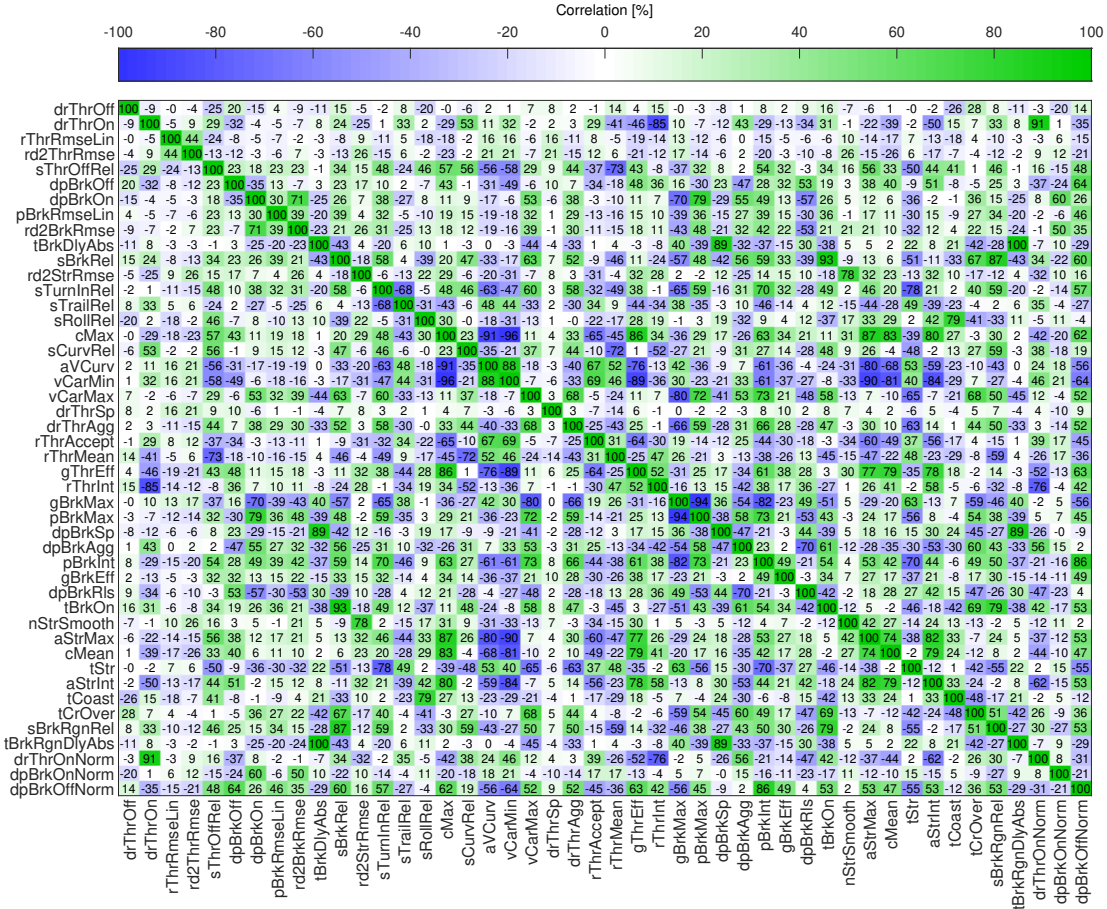


Figure 3.34.: Spearman correlation matrix of metrics valid for both racing series except the statisticals

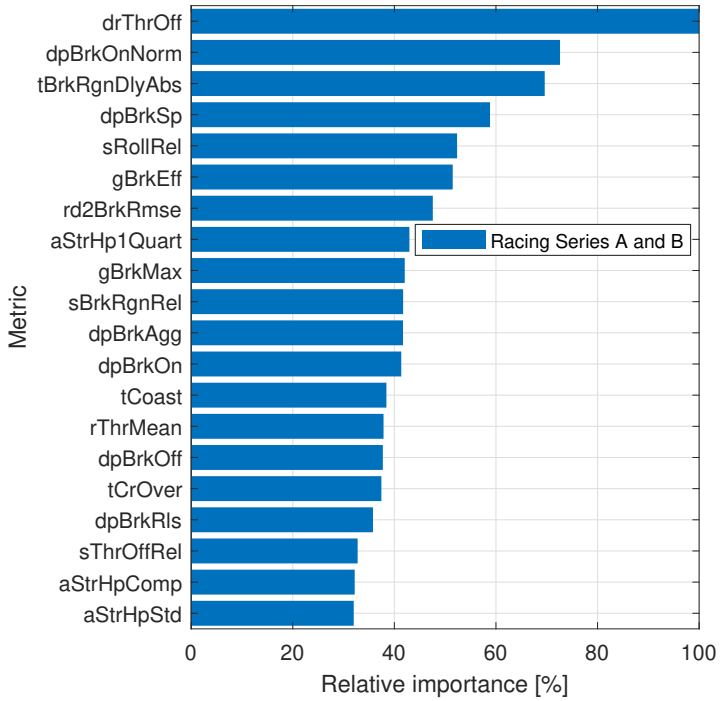
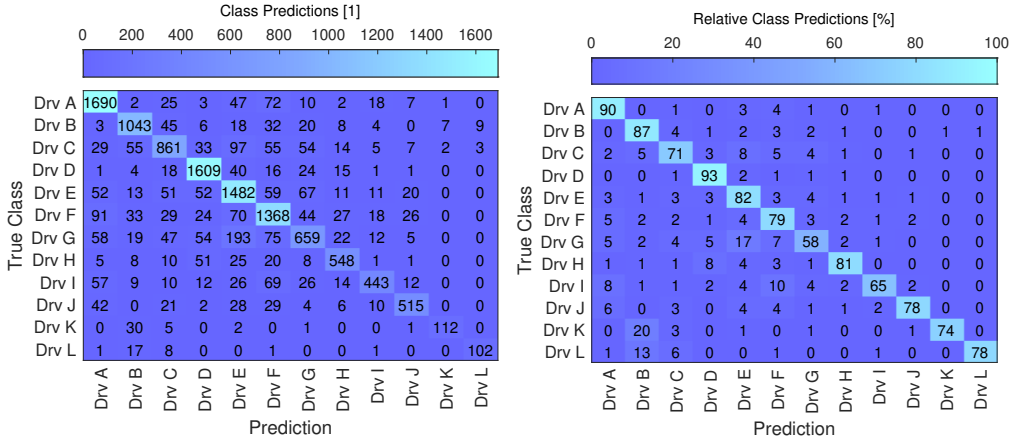


Figure 3.35.: Decision tree importance factors of both racing series combined



(a) Absolute predictions

(b) Relative predictions

Figure 3.36.: Driver class predictions of both racing series

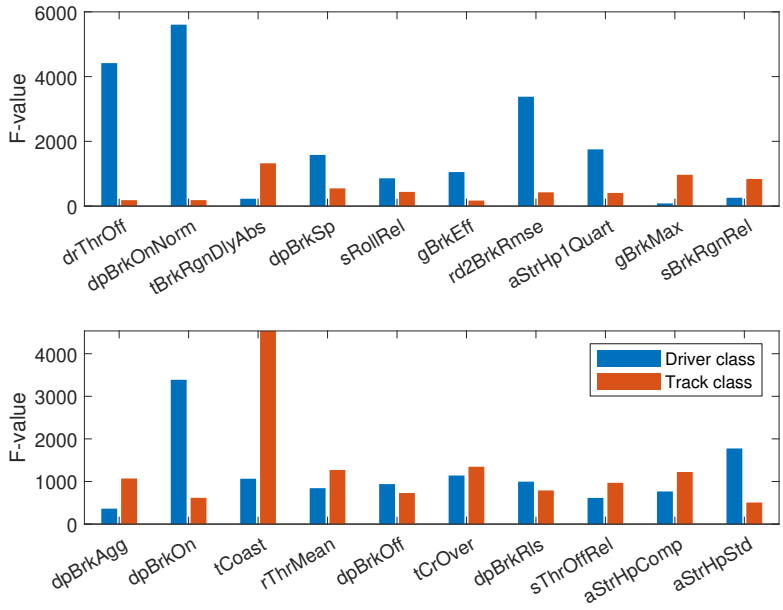
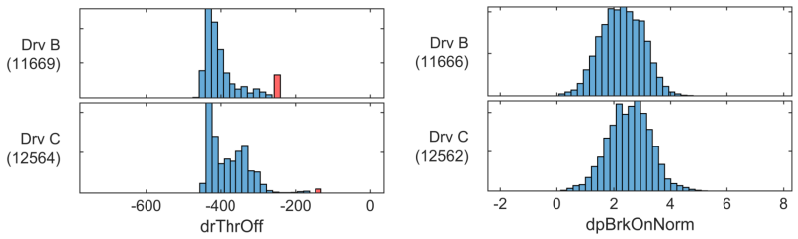
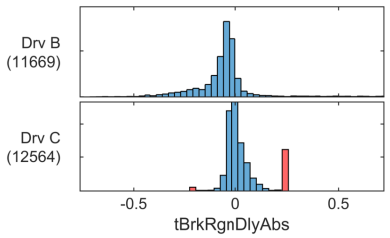


Figure 3.37.: F-values for the driver and track class of both racing series

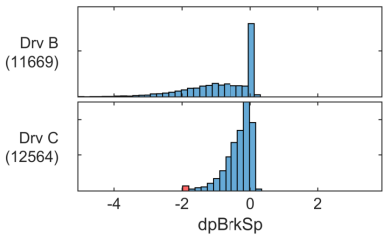


(a) Metric $drThrOff$

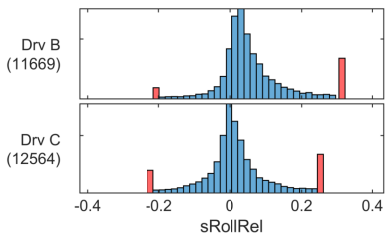
(b) Metric $dpBrkOnNorm$



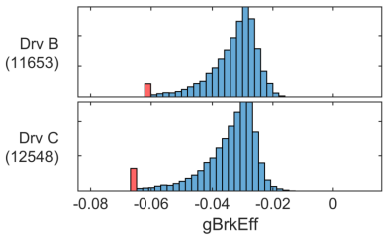
(c) Metric $tBrkRgnDlyAbs$



(d) Metric $dpBrkSp$



(e) Metric $sRollRel$



(f) Metric $gBrkEff$

Figure 3.38.: Histogram probability distributions among the drivers which competed in both racing series

3.3.4. Discussion

The F-value and histogram analyzes seem to correlate with the results from the driver classification. The metrics with the highest driver and low track class F-values are within the first metrics ranked to their decision tree importances factors. Qualitatively, the correlation is better for racing series A than for racing series B. Furthermore, these metrics show noticeable differences in the probability distributions among the drivers.

Interestingly, for racing series A the on-brake gradient metric shows a much better behaviour if it gets normalized by the maximum brake pressure of the corresponding sector ($dpOnBrkNorm$). For the dataset in racing series B the importance for the normalized metric is only slightly smaller than the initial metric $dpOnBrk$. Generally, the brake metrics represent the group with the highest decision tree importance factors for the classification algorithm. The mean brake gradient $dpBrkSp$ shows the highest importance for racing series B where for racing series A $dpOnBrkNorm$ has the highest brake metric importance. Compared to racing series A and racing series B the importance of the normalized brake delay time $tBrkRgnDlyAbs$ increased at the combined dataset.

Another point of interest lies in the observation of the longitudinal acceleration related to the throttle ratio $gThrEff$ and the longitudinal acceleration to the brake pressure $gBrkEff$. The acceleration is a result of the pedal application, hence, one would expect this measure to be relatively constant within a vehicle class and therefore not suitable for describing driving styles. Clearly, the results show different properties for these metrics, as they are within the first twenty metrics of racing series B according to their importance factors and for racing series A $gBrkEff$ is within the first twenty driver metrics.

The decision tree importance factors do not necessarily represent the metrics with the best ability to describe driving styles. For example if a metric shows an anomaly for one driver, the classifier can predict this driver with a relative high accuracy leading to a high importance factor. This could explain the extreme importance factor of the mean brake gradient $dpBrkSp$ for racing series B which is almost twice as high as the second ranked metric for the dataset.

3.4. Metric Transferability - Real Car to DIL Simulator

In this chapter driver metrics derived from simulator data are compared with metrics calculated from data of the real car. The data is based on racing series A and contains the respective set of metrics.

In figure 3.39 it can be noticed that no throttle blips are applied during the downshifting phase, therefore the brake gradient at throttle blips $dpBrkBlip$ contains nearly 100% NaN values. Again, as in section 3.3.1, the NaN values of the shift metric $rdShftRel$ are set to zero. For the entire DIL simulator dataset, not enough information of the apex distance metrics $aVApex$, $xApexMin$ and $sApexRel$ is available, therefore they are also not considered for the driver classification.

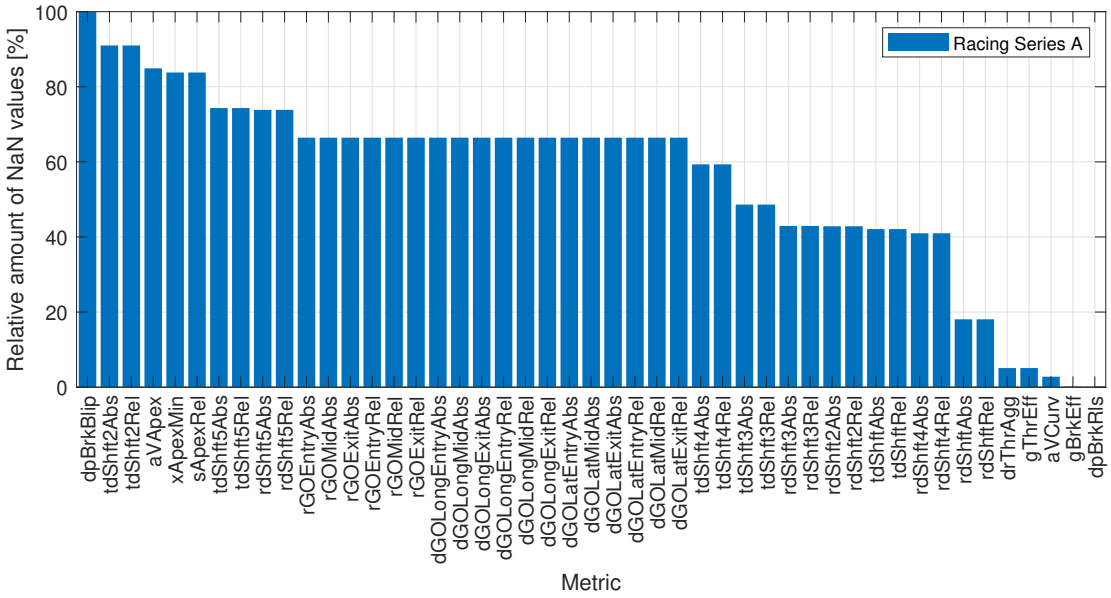


Figure 3.39.: Occurrences of NaN's in the metric values of simulator data in racing series A

The decision tree importance factors of figure 3.40 differ to those of figure 3.21 from real car data. The relative downshift engine speed $rdShftRel$ shows a much higher relative importance (100%) than for real car data (41%). Furthermore, the off-throttle gradient $drThrOff$ and the normalized on-brake gradient $dpBrkOnNorm$ importances have decreased, while the absolute brake delay $tBrkDlyAbs$ importance has increased.

Figure 3.41 represents the classification accuracies of the driver prediction. For driver A and D the relative accuracy correlates with real car data. Driver E, H and I show much better results for simulator DIL sessions with relative prediction accuracies of 90%, 91% and 92% compared to 84%, 81% and 69% of the real car dataset. The simulator driver classification represents the highest mean classification result of 92%.

The ANOVA in figure 3.42 illustrates the F-values of the driver class together with the F-values among the source class which stands for DIL simulator or real car data. The source class clearly shows higher values than the driver class which means that the variances among the drivers are smaller than the variances between real car and simulator data.

The histograms of figure 3.43 compare the probability distributions of the simulator data with the data of the real car for the drivers and metrics. Clear differences can be seen at the off-throttle gradient $drThrOff$, the brake delay time $tBrkDlyAbs$, the relative rolling distance $sRollRel$ and the mean brake gradient $dpBrkSp$. For the normalized on-brake gradient $dpBrkOnNorm$ only driver A shows clear deviations.

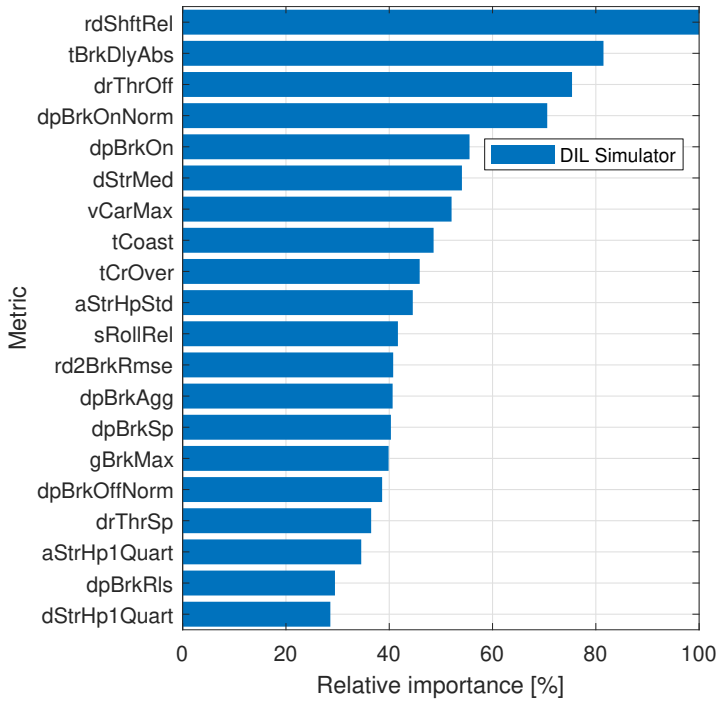
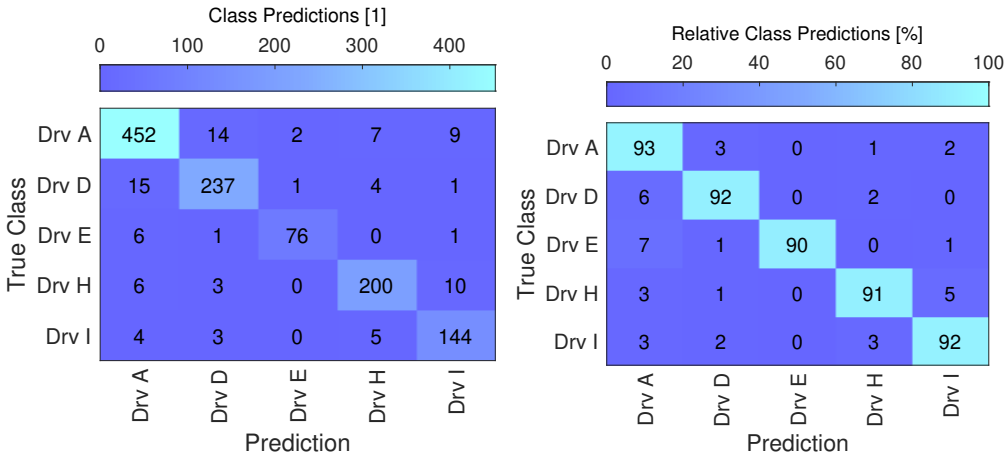


Figure 3.40.: Decision tree importance factors of simulator data of two seasons for racing series A



(a) Absolute predictions

(b) Relative predictions

Figure 3.41.: Driver class predictions of DIL simulator data of racing series A

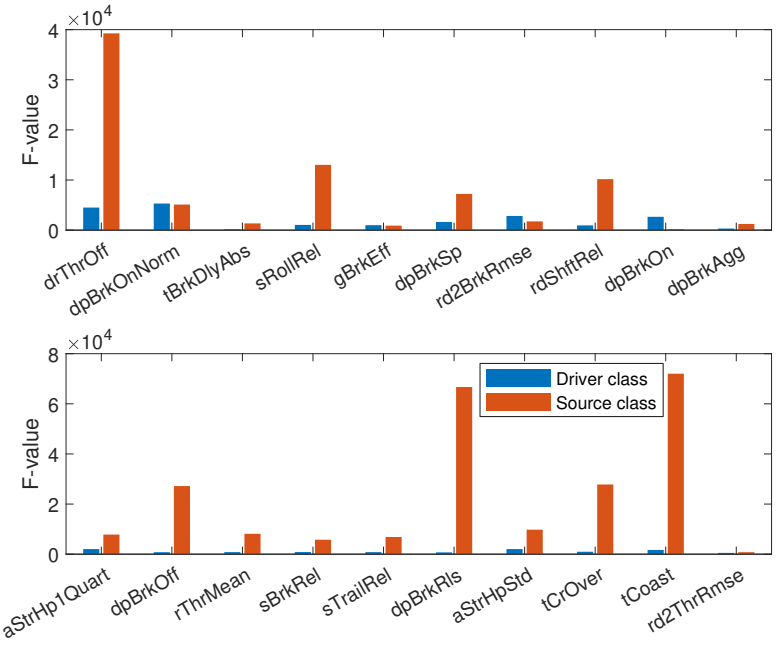
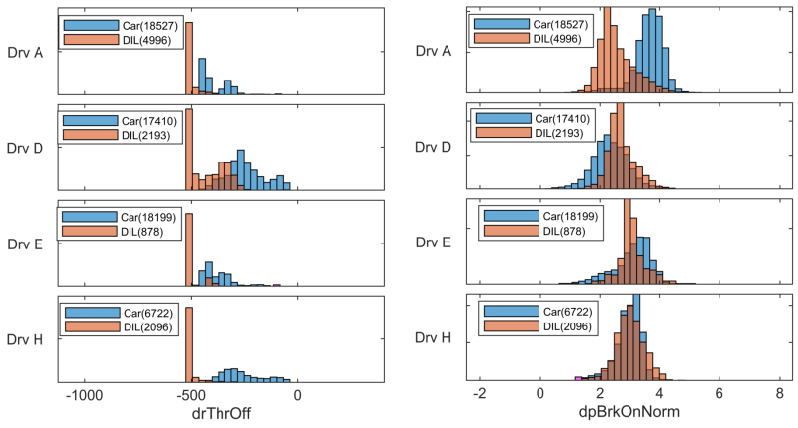
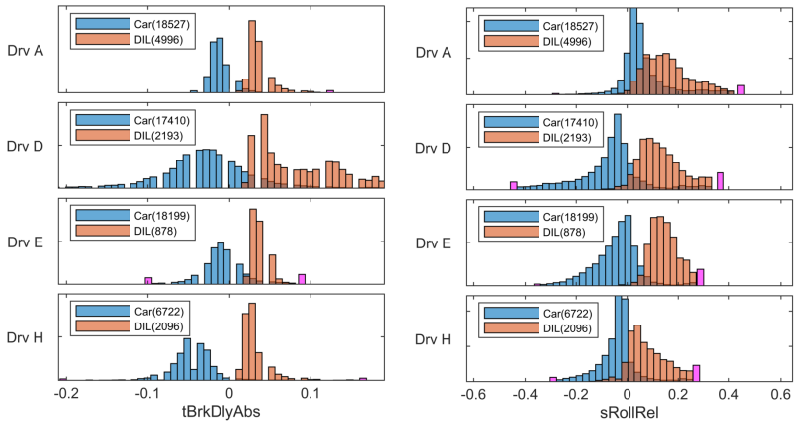


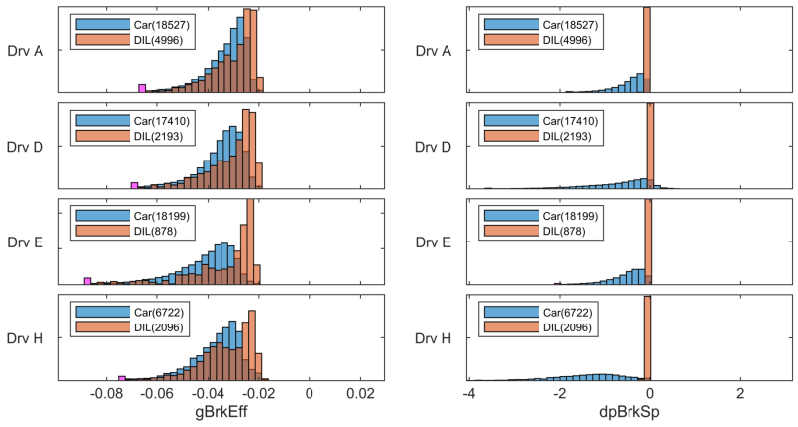
Figure 3.42.: Transferability comparison of the F-values for the driver and racing source class of racing series A



(a) Metric $drThrOff$ (b) Metric $dpBrkOnNorm$



(c) Metric $tBrkDlyAbs$ (d) Metric $sRollRel$



(e) Metric $gBrkEff$ (f) Metric $dpBrkSp$

Figure 3.43.: Histogram probability distributions among the drivers for racing series A of real car and DIL simulator data

3.4.1. Discussion

The classification accuracies show impressive results for the DIL simulator dataset, but this could also be related to the smaller amount of data compared to the real car. Similar as at the real car dataset, driver A and D have very high classification results, indicating a distinct driving behaviour which manifests at observing the histogram distributions. Mainly for the brake metrics $dpBrkOnNorm$ and $tBrkDlyAbs$ distinct driving behaviours can be noticed which again indicates the high importance of the braking metrics for the driving style classification.

Observing the histogram probability distributions of the metrics and drivers, big differences between real car and DIL simulator data can be noticed. The ANOVA F-values for the source class further prove this behaviour. Probably the pedal properties of the simulator are not correlating well with those of the real car which should be further investigated in future.

3.5. Transferability of Metrics - Among different Racing Series

In this chapter the transferability of the driver metrics among the racing series is analyzed.

In figure 3.44 the decision tree importance factors of the eleven metrics in common of both racing series from figures 3.21 and 3.28 are illustrated. The brake metrics show higher relative importances in racing series A with the exception of the mean brake gradient $dpBrkSp$ which has half the relative importance compared to racing series B. Similarly the throttle metrics have higher relative importance factors in racing series A, $drThrOff$ has even twice the importance as in racing series B. The roll distance driver metric $sRollRel$ also shows higher relative importances in racing series A compared to racing series B. However, without the high importance of the mean brake gradient $dpBrkSp$ in racing series B which could be caused by anomalies in the brake pressure signal for a specific driver, the relative importances of the other metrics would be more similar to the dataset of racing series A.

The variance analysis of figure 3.45 illustrates the F-values of the driver and racing series class for the initial metric set and for the offset transferred metrics according to section 2.5.2. The transferred dataset shows significantly lower F-values in the series class while the F-values in the drivers class remain relatively constant compared to the initial dataset with the exception for the on-brake gradients $dpBrkOn$ and $dpBrkOnNorm$.

However, observing the initial probability distributions in figure 3.46 and the offset transferred histograms in figure 3.47, the transferability indicated in the variance analysis is not proved. The initial dataset shows some metrics with a poor transferability as illustrated quantitatively in table 3.4, especially for the brake delay time $tBrkDlyAbs$ and the mean brake gradient $dpBrkSp$. Contrary, the normalized on-brake gradient

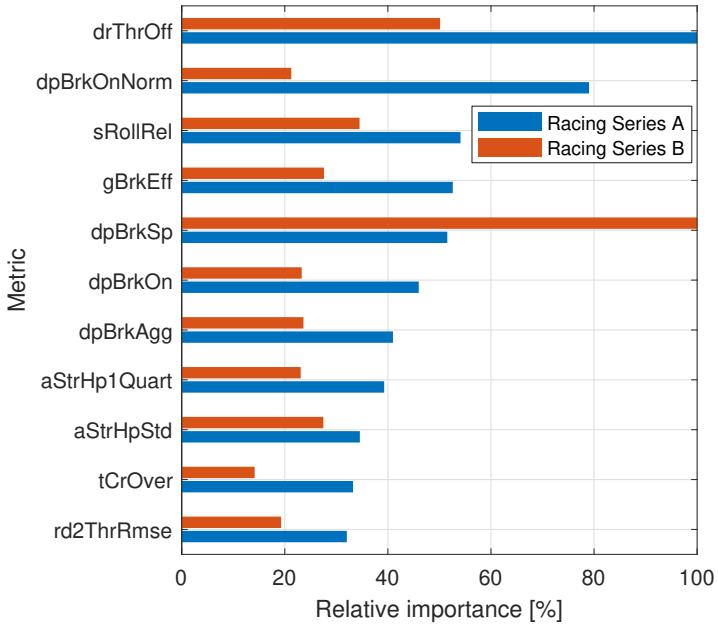
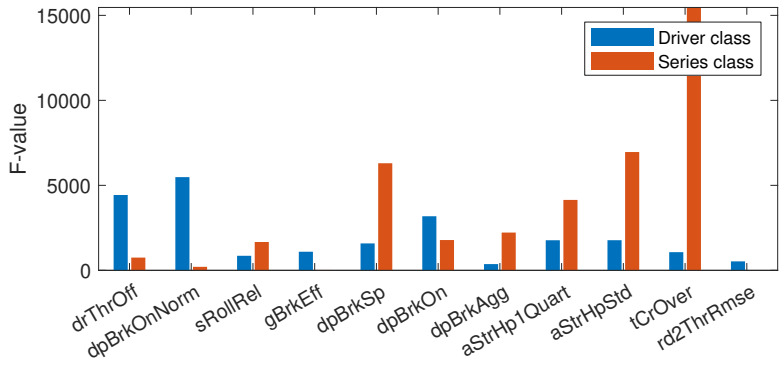


Figure 3.44.: Decision tree importance factors of racing series A compared to racing series B

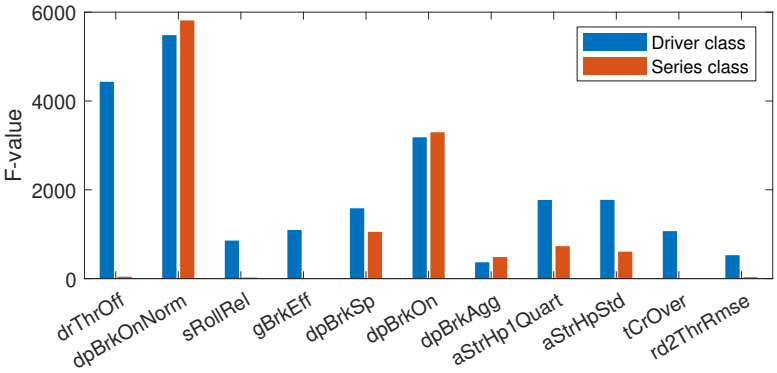
$dpBrkOnNorm$, the brake efficiency $gBrkEff$ and the relative rolling distance for driver B $sRollRel$ shows good correlations among the racing series.

The transferred metrics show significantly worse transferabilities for the on-brake gradients $dpBrkOn$ and $dpBrkOnNorm$ and the brake delay time $tBrkDlyAbs$, resulting in smaller histogram overlaps in table 3.5. For the transferred $tBrkDlyAbs$ metric the peak of the histogram distribution for racing series B is contained in the outliers. This behaviour may result due to a median (for the offset definition) that does not correlate with the probability peak in the initial $tBrkDlyAbs$ metric histogram. For the off-throttle gradient $drThrOff$, the relative rolling distance $sRollRel$ and the brake efficiency $gBrkEff$ the transferability is better than for the initial dataset.

Figure 3.48 shows adapted metric definitions (table 2.10) to improve the transferability among different racing series. The throttle and brake metric adaptions of $drThrOnNorm$ and $dpBrkOnNorm$ show better results which are quantitatively proved in table 3.6. The not adapted off-brake gradient $dpBrkOff$ has the highest probability overlap and decreases with the normalization. The adaptions for $tBrkDlyAbs$ and $sBrkRel$ seem to have less impact on the transferability which may be explained by the relative rare actuation of the recuperation pedal in the observed data.

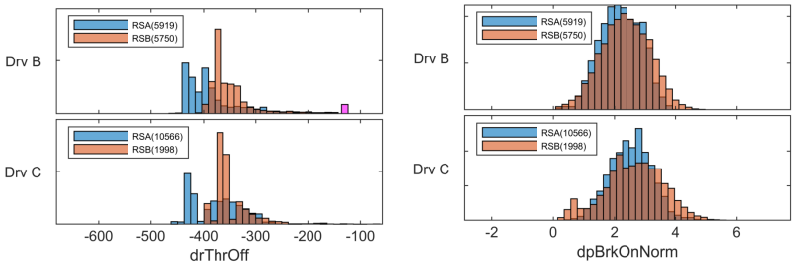


(a) Initial metrics



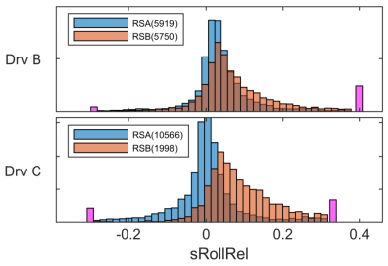
(b) Transferred metrics

Figure 3.45.: Transferability comparison of the F-values for the driver and racing series class of racing series A and B

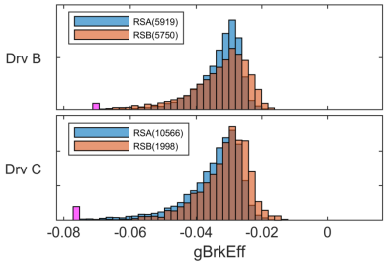


(a) Metric $drThrOff$

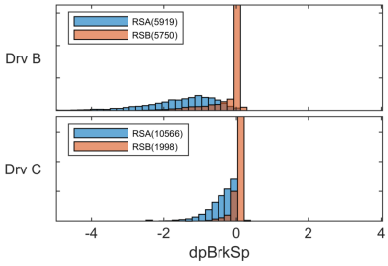
(b) Metric $dpBrkOnNorm$



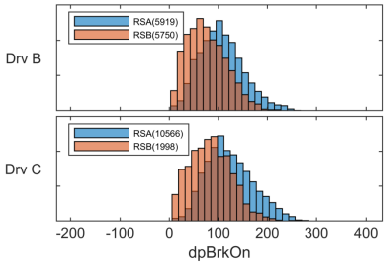
(c) Metric $sRollRel$



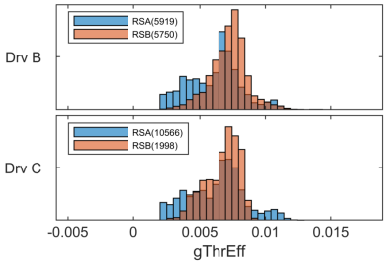
(d) Metric $gBrkEff$



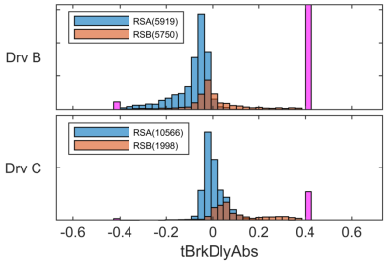
(e) Metric $dpBrkSp$



(f) Metric $dpBrkOn$

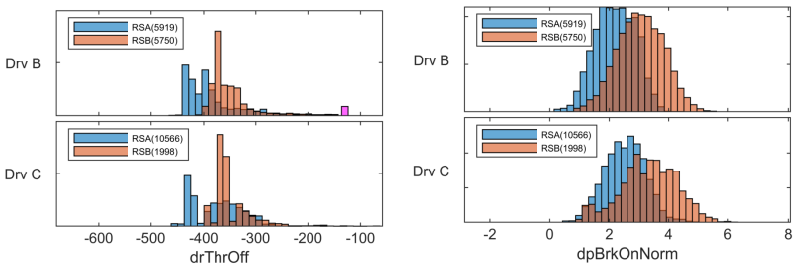


(g) Metric $gThrEff$



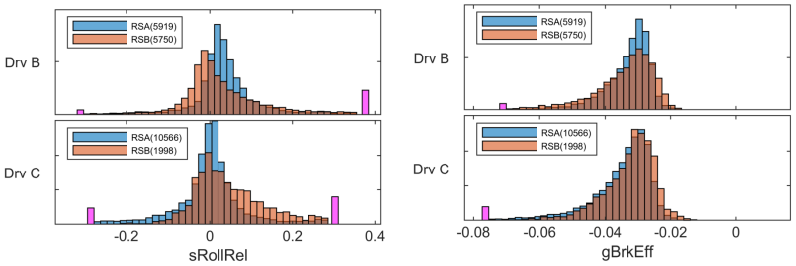
(h) Metric $tBrkDlyAbs$

Figure 3.46.: Histogram probability distributions among the drivers for both racing series



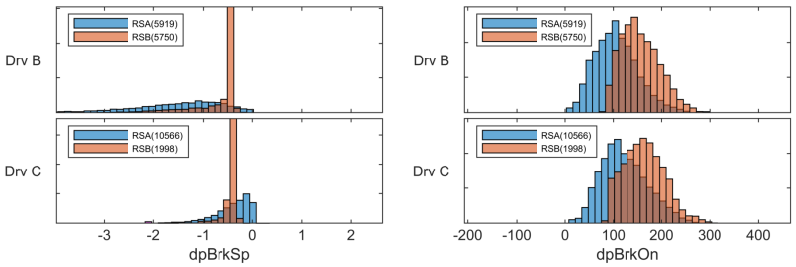
(a) Metric $drThrOff$

(b) Metric $dpBrkOnNorm$



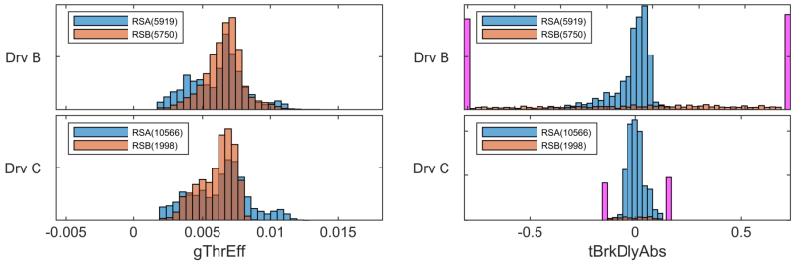
(c) Metric $sRollRel$

(d) Metric $gBrkEff$



(e) Metric $dpBrkSp$

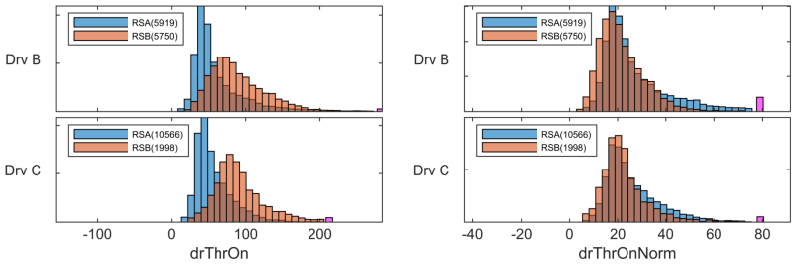
(f) Metric $dpBrkOn$



(g) Metric $gThrEff$

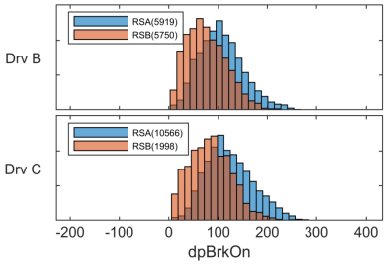
(h) Metric $tBrkDlyAbs$

Figure 3.47.: Histogram probability distributions among the drivers for both racing series with an applied racing series offset

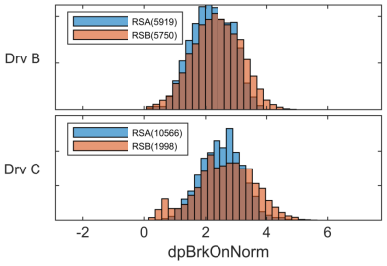


(a) Metric *drThrOn*

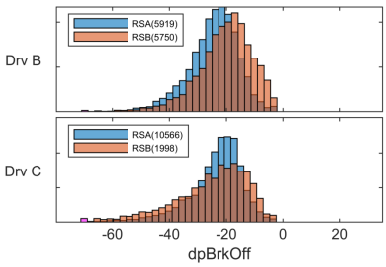
(b) Metric *drThrOnNorm*



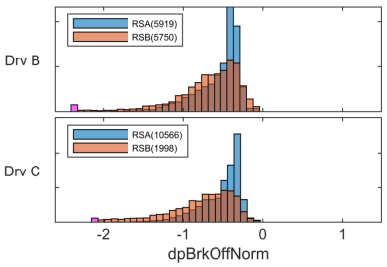
(c) Metric *dpBrkOn*



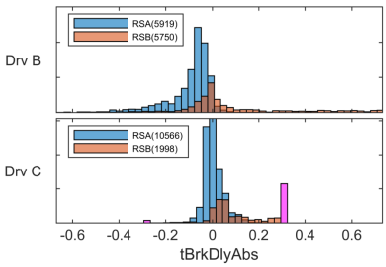
(d) Metric *dpBrkOnNorm*



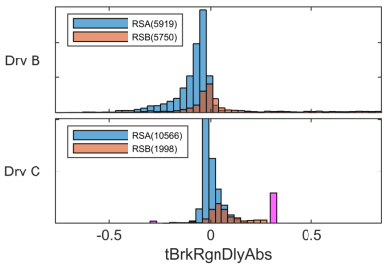
(e) Metric *dpBrkOff*



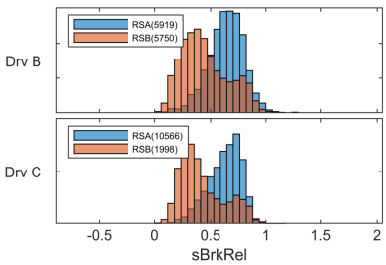
(f) Metric *dpBrkOffNorm*



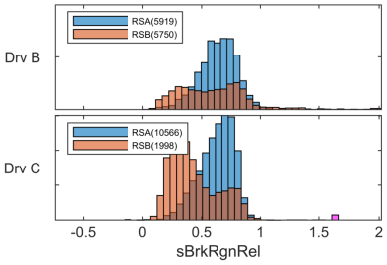
(g) Metric *tBrkDlyAbs*



(h) Metric *tBrkRgnDlyAbs*



(i) Metric *sBrkRel*



(j) Metric *sBrkRgnRel*

Figure 3.48.: Histogram probability distributions among the drivers for both racing series of the initial and adapted metrics

Table 3.4.: Relative overlap of the probability distributions from the two racing series

Metric	Driver B	Driver C
drThrOff	33%	40%
dpBrkOnNorm	58%	48%
sRollRel	63%	28%
gBrkEff	58%	50%
dpBrkSp	13%	9%
dpBrkOn	39%	36%
gThrEff	48%	37%
tBrkDlyAbs	6%	43%

Table 3.5.: Relative overlap of the probability distributions from the two racing series with an applied racing series offset

Metric	Driver B	Driver C
drThrOff	42%	35%
dpBrkOnNorm	25%	29%
sRollRel	66%	44%
gBrkEff	58%	58%
dpBrkSp	19%	22%
dpBrkOn	26%	32%
gThrEff	47%	36%
tBrkDlyAbs	13%	29%

Table 3.6.: Relative overlap of the probability distributions from the two racing series of the initial and adapted metrics

Metric	Driver B	Driver C
drThrOn	48%	25%
drThrOnNorm	48%	60%
dpBrkOn	39%	36%
dpBrkOnNorm	58%	48%
dpBrkOff	60%	60%
dpBrkOffNorm	50%	44%
tBrkDlyAbs	6%	43%
tBrkRgnDlyAbs	6%	45%
sBrkRel	35%	35%
sBrkRgnRel	24%	28%

3.5.1. Discussion

The brake pedal metrics seem to have a better ability to describe driving styles for racing series A than for racing series B, as the corresponding decision tree importance factors

differ between the racing series. Generally the importance factors are lower for racing series B which can also be seen in the worse mean prediction accuracy.

Without modifications the metrics show partially noticeable differences between the racing series. Driver C spends more time in the rolling phase at racing series B. The brake pressure efficiency $gBrkEff$ seems to correlate well which indicates that the brake pressure has the same influence on the longitudinal acceleration for both racing series. The throttle efficiency $gThrEff$ differs between the racing series which can be explained by the differences in the power output and vehicle mass of the two racing series. Both drivers tend to have a slightly higher brake delay time in racing series B which may be explained by the softer suspension setup of the low downforce vehicles of racing series B. With a softer suspension the needed time for the load transfer increases and the drivers adapt their brake delay.

Transferring racing series B with an offset shows partially better results, but also worse for some metrics. For example, the relative rolling distance $sRollRel$ overlap can be increased, while the brake delay time $tBrkDlyAbs$ and the on-brake gradients $dpBrkOn$ and $dpBrkOnNorm$ show a worse behaviour.

A better transferability of some metrics among the racing series is achieved by adapting specific metrics individually. The on-brake gradient $dpBrkOn$ indicates a faster brake application for racing series A. Contrary, the normalized measure $dpBrkOnNorm$ shows a better correlation. Thereupon, the differences in $dpBrkOn$ come from different maximum brake pressures of the vehicle classes. The off-brake gradient $dpBrkOff$ appears to be already correlating well without normalizing it by the maximum brake pressure which indicates that the drivers are releasing the brakes without considering the maximum brake pressure. Interestingly the on-throttle gradient seems to be transferable when it is normalized by an empiric factor depending on the vehicle power output divided by the vehicle mass.

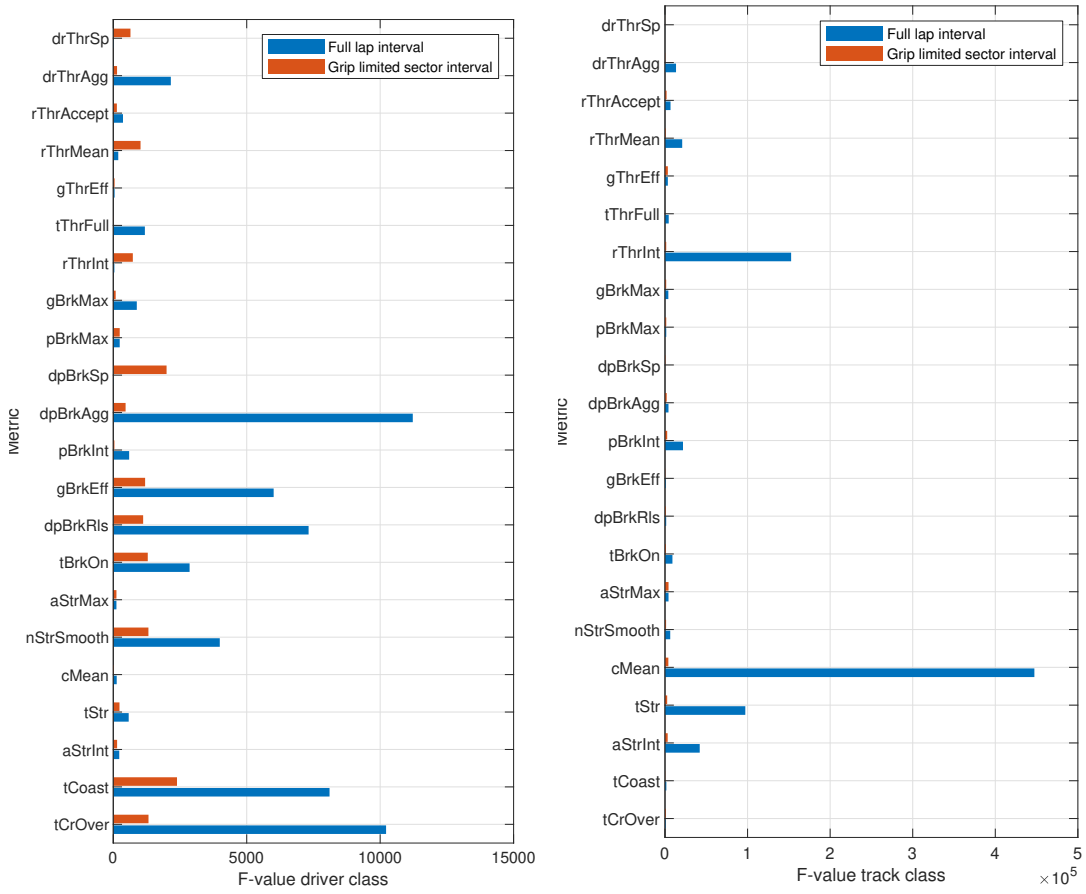
Training the classifier with data from racing series A and letting it classify the drivers from racing series B would produce meaningful statements on the transferability. On the other hand, only two drivers are competing in both racing series which is not enough to yield valuable information with this method.

3.6. Comparison of Feature Extraction Methods

For the feature extraction method comparison eleven drivers on ten different tracks of one racing series without DIL simulator data are analyzed. The data was recorded during three racing seasons with a sum of 16840 laps, an average of 5.4 sectors and 113.3 sliding windows per lap.

In figure 3.49 the F-values for the driver class on the left and the track class on the right are compared for the full lap feature extraction intervals and the grip limited sector intervals. For the comparison the metrics of section 2.2.4 are used, as these metrics were

initially defined for the driver comparison on entire laps.

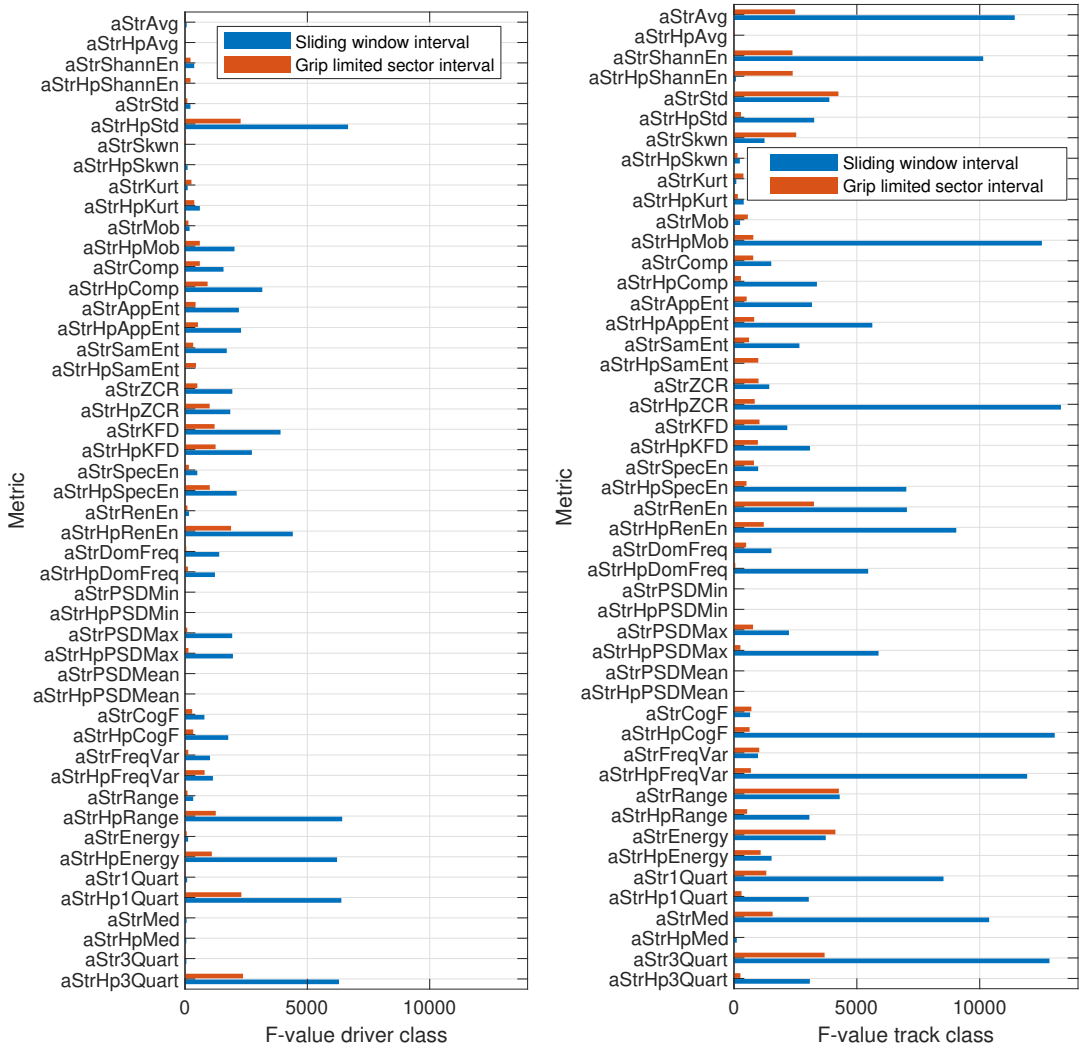


(a) Full lap calculation method

(b) Grip limited sector method

Figure 3.49.: F-value driver and track class comparison of full lap and grip limited sector feature extraction methods

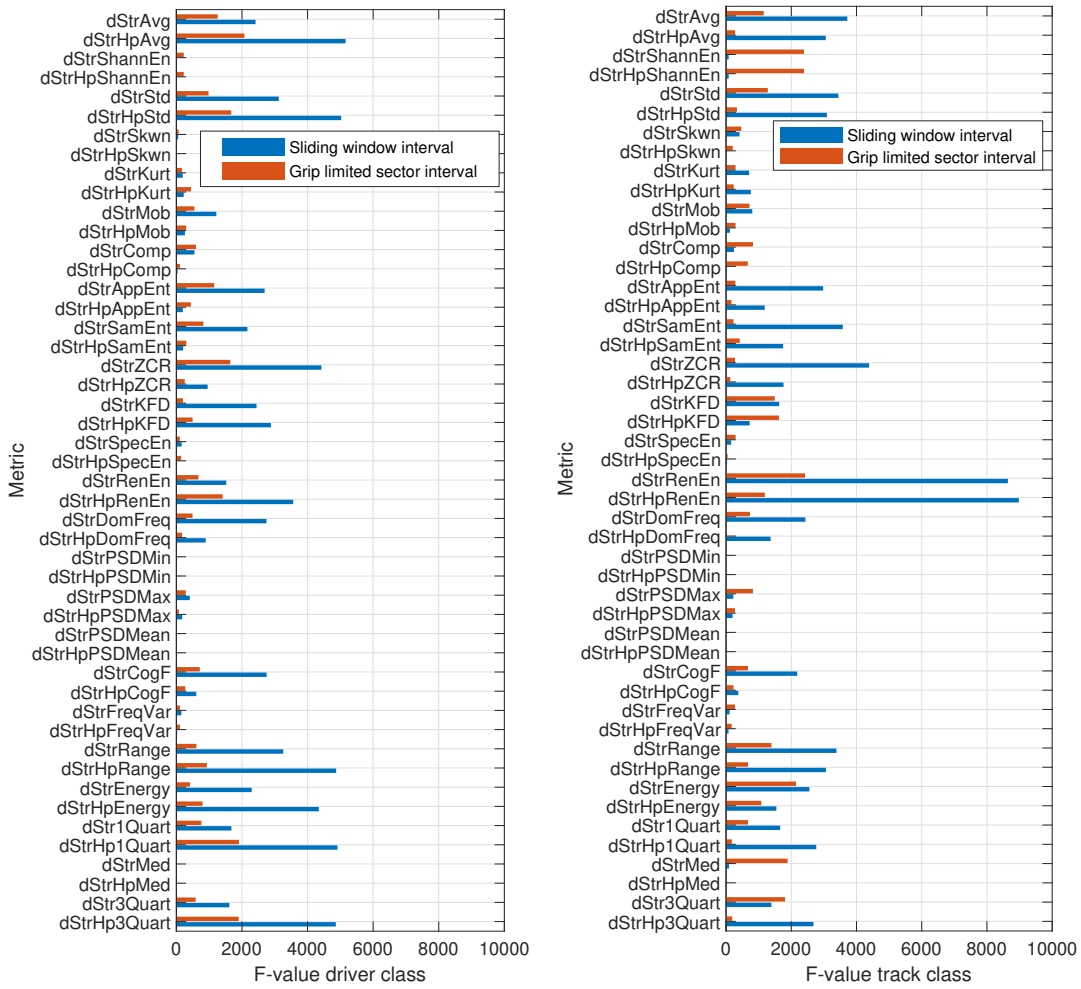
Figure 3.50 illustrates the driver and track class F-values of the statistical metrics based on the steering wheel angle calculated in sliding window intervals and in grip limited sector intervals. For most of the metrics the sliding window intervals show higher F-values at the driver and especially at the track class. A similar, but less pronounced behaviour can be observed for the statistical metrics based on the steering wheel angle rate in figure 3.51.



(a) Sliding window calculation method

(b) Grip limited sector method

Figure 3.50.: F-value driver and track class comparison of sliding window and grip limited sector feature extraction methods for steering wheel angle statistic metrics



(a) Sliding window calculation method

(b) Grip limited sector method

Figure 3.51.: F-value driver and track class comparison of sliding window and grip limited sector feature extraction methods for steering wheel angle rate statistic metrics

3.6.1. Discussion

For the statistical based metrics and the metrics initially defined on a full lap, the corresponding original feature extraction method results in higher F-values for both, the driver and especially the track class. To describe driving styles independently from the race tracks, the high variances in the metrics among the race tracks are not preferable, hence, the feature extraction interval of grip limited sectors seems to be the most suitable one for the scope of this work. Additionally, as racecar drivers are well experienced and perform on a high level, the definition of driver metrics based on the performance limit of the driver-vehicle-environment system should result in more driver specific information.

4. Conclusion and Outlook

The goal in motorsports is to maximize the average vehicle speed during competitions which is determined by the dynamic interactions of the system driver-vehicle-environment. Even if the environment is changing continuously, it can be assumed constant among the competitors. Thereupon, maximizing the performance is focussed on the manipulation of the driver-vehicle interactions. The aim of this work was to develop driver metrics to objectively describe driving styles of professional race car drivers independently of the race tracks for two racing series. The development of the objective metrics is based on a novel method introduced by Wörle et al. [WGE18]. As driving behaviours are usually strongly depending on the interactions of the driver-vehicle-environment dynamical systems, a transferability analysis of the metrics among different racing series with the background of different vehicle classes was a further point of interest.

Driver metrics were calculated in automatically detected sectors. To extract the most driving style specific information, the sectors are defined in grip limited corners with applied braking and further divided into characteristic throttle, brake and steering points. Two additional metric extraction intervals were analyzed and rejected because of a higher track dependency of the metrics.

The driver metrics were defined as driver input gradients, distances and duration times, as well as measures depending on the trajectory curvature. Further metric definitions were introduced, relying on statistical quantities for the steering wheel angle and for the steering wheel angle rate. These definitions were additionally applied for the highpass filtered signals, to remove the race track dependend curvature. Racing series specific driver metrics were calculated as shift metrics for racing series A and as energy management metrics for racing series B.

These metrics gave a lot of information on the driver input behaviour and the spatial trajectory chosen by the drivers. Nevertheless, no statements about the direct performance relevance could be made, therefore a grip optimization method was elaborated to calculate the theoretical maximum acceleration which the vehicle is capable of. Due to the choice of the optimizations initial point relying on DIL simulator data of the drivers, driver specific statements on the vehicle exploitation could be made. The biggest deviations between the optimized acceleration and the acceleration achieved by the drivers occur during braking phases which indicates the difficulty of exploiting the vehicle performance in braking maneuvers. Metrics were then defined using the grip optimization measures, but proved not to be suitable for characterizing driving styles. However, the performance relevance of the grip optimization metrics could be used to observe the correlations to other metrics and therefore gain information on the nontrivial performance

relevance of metrics which were more suitable to describe specific driving behaviours.

For validation purposes of the driver metrics regarding their ability to describe driving styles, a decision tree classification algorithm was applied to classify drivers using driver metrics. The relative prediction accuracies showed partially high variations for different drivers. The drivers with the highest prediction accuracies had distinctive patterns in the probability distributions for some metrics, proving the assumption that driving styles differ throughout different drivers. Brake pedal metrics represented the group with the highest importances for the classification algorithm.

For the combined usage of DIL simulator and real car data, the transferability among the data sources had to be analyzed. Partially, high deviations for metrics of the same driver but different data sources could be observed. Thereupon, it is suggested to only analyze the results for each data source independently.

Additionally, a transferability analysis of the driver metrics among the two racing series, showed differences in the results of a driver for two different vehicle classes. A normalization with an offset for each metric based on the median of the metric values did not lead to an acceptable outcome. However, for some driver metrics specific normalization methods proved to yield racing series independend values. Furthermore, some normalized metrics even improved the classifier accuracy on the single racing series dataset significantly.

For future works, a more in depth analysis on the individual normalization of specific driver metrics could lead to improved driving style representation within the racing series, as well as to a better transferability performance among different racing series. To use a bigger amount of data and to be independend of real car deployments, the correlation between DIL simulator and real car data has to be investigated further. At the moment the differences in the driver metrics compared to the real car are not negligible, hence, a reasonable combined usage of the data is not possible. The vehicle model of the grip optimization method can be improved as well, to sort out the deviations to the vehicle model of the simulator. Additionally, the method could be used for other applications, besides the definition and observation of driver metrics. As the goal of describing driver behaviours by objective criteria lies in maximizing the performance of the driver-vehicle interactions, the next step is the investigation of how the driver metrics can be used to influence the setup of a vehicle. Subsequently, the vehicle properties can be adjusted to meet the driver specific needs. The methods of this work could then lead to a faster identification of the performance relevant metrics and to improved driver metrics of different racing series.

List of Figures

List of Tables

Bibliography

[Arc94] Thomas J. Archdeacon. *Correlation and regression analysis: A historian's guide*. Univ. of Wisconsin Press, Madison Wis. u.a., 1. [print.] edition, 1994.

[ASEN19] Sadegh Arefnezhad, Sajjad Samiee, Arno Eichberger, and Ali Nahvi. Driver drowsiness detection based on steering wheel data applying adaptive neuro-fuzzy feature selection. *Sensors (Basel, Switzerland)*, 19(4), 2019.

[BBB16] Francesco Biral, Enrico Bertolazzi, and Paolo Bosetti. Notes on numerical methods for solving optimal control problems. *IEEJ Journal of Industry Applications*, 5:154–166, 2016.

[Bre96] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[Bre01] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[BT95] Paul T. Boggs and Jon W. Tolle. Sequential quadratic programming. *Acta Numerica*, 4:1–51, 1995.

[BTTP17] Lejo Buning, Abhishek Tomar, Frans Tillema, and Joop Pauwelussen. Adaptive support system for driver's changing role. 2017.

[CD03] Subhabrata Chakraborti and Jean Dickinson Gibbons. *Nonparametric Statistical Inference*. Taylor & Francis, Oxford, England, ed. 4 edition, 2003.

[DLG18] Nicola Dal Bianco, Roberto Lot, and Marco Gadola. Minimum time optimal control simulation of a gp2 race car. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 232(9):1180–1195, 2018.

[Don82] Edmund Donges. Aspekte der aktiven Sicherheit bei der Führung von Personenkraftwagen. *Automobil Industrie 27*, 27(2):183–190, 1982.

[Fia66] E. Fiala. Lenken von Kraftfahrzeugen als kybernetische Aufgabe. *Automobiltechnische Zeitschrift*, 68:156–162, 1966.

[FY10] Fabian Friedrichs and Bin Yang. Drowsiness monitoring by steering and lane data based features under real driving conditions. *European Signal Processing Conference*, 2010.

[GWBV02] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1):389–422, 2002.

[HHRH11] Daniel Haupt, Petr Honzík, Peter Raso, and Ondrej Hyncica. *Steering wheel motion analysis for detection of the driver's drowsiness.* 2011.

[HK11] Jan Hauke and Tomasz Kossowski. Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, 30(2):87–93, 2011.

[KHG17] John C. Kegelman, Lene K. Harbott, and J. Christian Gerdes. Insights into vehicle trajectories at the handling limits: analysing open data from race car drivers. *Vehicle System Dynamics*, 55(2):191–207, 2017.

[KJKS11] F. Kehrle, Frasch J.V, Christian Kirches, and Sebastian Sager. *Optimal control of Formula 1 race cars in a VDrift based virtual environment.* 2011.

[KL19] Michael Kofler and Cornelia Lex. Entwicklung und Anwendung einer quasi-stationären Rundenzeitsimulation. 2019.

[Kot13] S. B. Kotsiantis. Decision trees: a recent overview. *Artificial Intelligence Review*, 39(4):261–283, 2013.

[LWSG13] Gilles Louppe, Louis Wehenkel, Antonio Sutera, and Pierre Geurts. Understanding variable importances in forests of randomized trees. *Advances in Neural Information Processing Systems*, 26, 2013.

[MAT18] MATLAB. *version 9.5.0.944444(R2018b).* The MathWorks Inc, Natick, Massachusetts, 2018.

[MM95] William F. Milliken and Douglas L. Milliken. *Race Car Vehicle Dynamics.* Society of Automotive Engineers, Inc., Warrendale, 5 edition, 1995.

[Moi08] G. G. Moisen. Classification and regression trees. In *In: Jørgensen, Sven Erik; Fath, Brian D. (Editor-in-Chief). Encyclopedia of Ecology, volume 1. Oxford, U.K.: Elsevier. p. 582-588*, pages 582–588. 2008.

[MTL78] Robert McGill, John W. Tukey, and Wayne A. Larsen. Variations of box plots. *The American Statistician*, 32(1):12, 1978.

[Nil96] Nils J. Nilsson. Introduction to machine learning: An early draft of a proposed textbook. pages 175-188. http://robotics.stanford.edu/people/nilsson/mlbook.ht ml, 1996.

[PB92] Hans B. Pacejka and Egbert Bakker. The magic formula tyre model. *Vehicle System Dynamics*, 21(sup001):1–18, 1992.

[PB12] Hans B. Pacejka and Igo Besselink. *Tire and vehicle dynamics.* Butterworth-Heinemann, Oxford UK, 3 edition, 2012.

[R C19] R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna, Austria, 2019.

[Ras83] Jens Rasmussen. Skills, rules, and knowledge; signals, signs, and symbols,

and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3):257–266, 1983.

[RSt18] RStudio Team. *RStudio: Integrated Development Environment for R*. Boston, MA, 2018.

[SAK+14] Sajjad Samiee, Shahram Azadi, Reza Kazemi, Ali Nahvi, and Arno Eichberger. Data fusion to develop a driver drowsiness detection system with robustness to signal loss. *Sensors (Basel, Switzerland)*, 14(9):17832–17847, 2014.

[SAS19] Pinky Sodhi, Naman Awasthi, and Vishal Sharma. Introduction to machine learning and its basic application in python. *SSRN Electronic Journal*, 2019.

[Seg14] Jorge Segers. *Analysis Techniques for Racecar Data Acquisition, Second Edition*. SAE International, Warrendale, 2nd ed. edition, 2014.

[vSWG+19] Julian von Schleinitz, Lukas Wörle, Michael Graf, Andreas Schröder, and Wolfgang Trutschnig. Analysis of race car drivers' pedal interactions by means of supervised learning. Accepted for publication in 2019.

[WGE18] Lukas Wörle, Michael Graf, and Arno Eichberger. Objective metrics for control inputs of racecar drivers. Accepted for publication in 2018.

[WR17] Matthew J. Weinstein and Anil V. Rao. Algorithm 984: Adigator, a toolbox for the algorithmic differentiation of mathematical functions in matlab using source transformation via operator overloading. *ACM Trans. Math. Softw.*, 44(2):21:1–21:25, 2017.

[WZ17] Marvin N. Wright and Andreas Ziegler. ranger : A fast implementation of random forests for high dimensional data in c++ and r. *Journal of Statistical Software*, 77(1), 2017.

A. Appendix

Table A.1.: Dropped metrics according to too high correlations

Category	Metric	Category	Metric
Grip limited sector metrics	rd2StrRmse	Full lap metrics	pBrkMax
Statistical metrics	aStrAvg		tBrkOn
	aStrShannEn	Shiftmetrics	rdShft5Abs
	aStrStd		rdShft4Abs
	aStrAppEnt		rdShft3Abs
	aStrKFD		rdShft2Abs
	aStrRenEn		rdShft1Abs
	aStrRange		rdShft5Rel
	aStrEnergy		rdShft4Rel
	aStrHpShannEn		rdShft3Rel
	aStrHpMob		rdShft2Rel
	aStrHpAppEnt		rdShft1Rel
	aStrHpKFD		tdShft5Abs
	aStrHpRenEn		tdShft4Abs
	aStrHpRange		tdShft3Abs
	aStrHpEnergy		tdShft2Abs
	aStrHp3Quart		tdShft5Rel
	dStrAvg		tdShft4Rel
	dStrShannEn		tdShft3Rel
	dStrStd		tdShft2Rel
	dStrAppEnt	Grip optimizer metrics	rGOEntryAbs
	dStrSpecEn		rGOMidAbs
	dStrRenEn		rGOExitAbs
	dStrCogF		
	dStrRange		
	dStrEnergy		
	dStrHpAvg		
	dStrHpShannEn		
	dStrHpStd		
	dStrHpAppEnt		
	dStrHpKFD		
	dStrHpRenEn		
	dStrHpFreqVar		
	dStrHpRange		
	dStrHpEnergy		
	dStrHp3Quart		

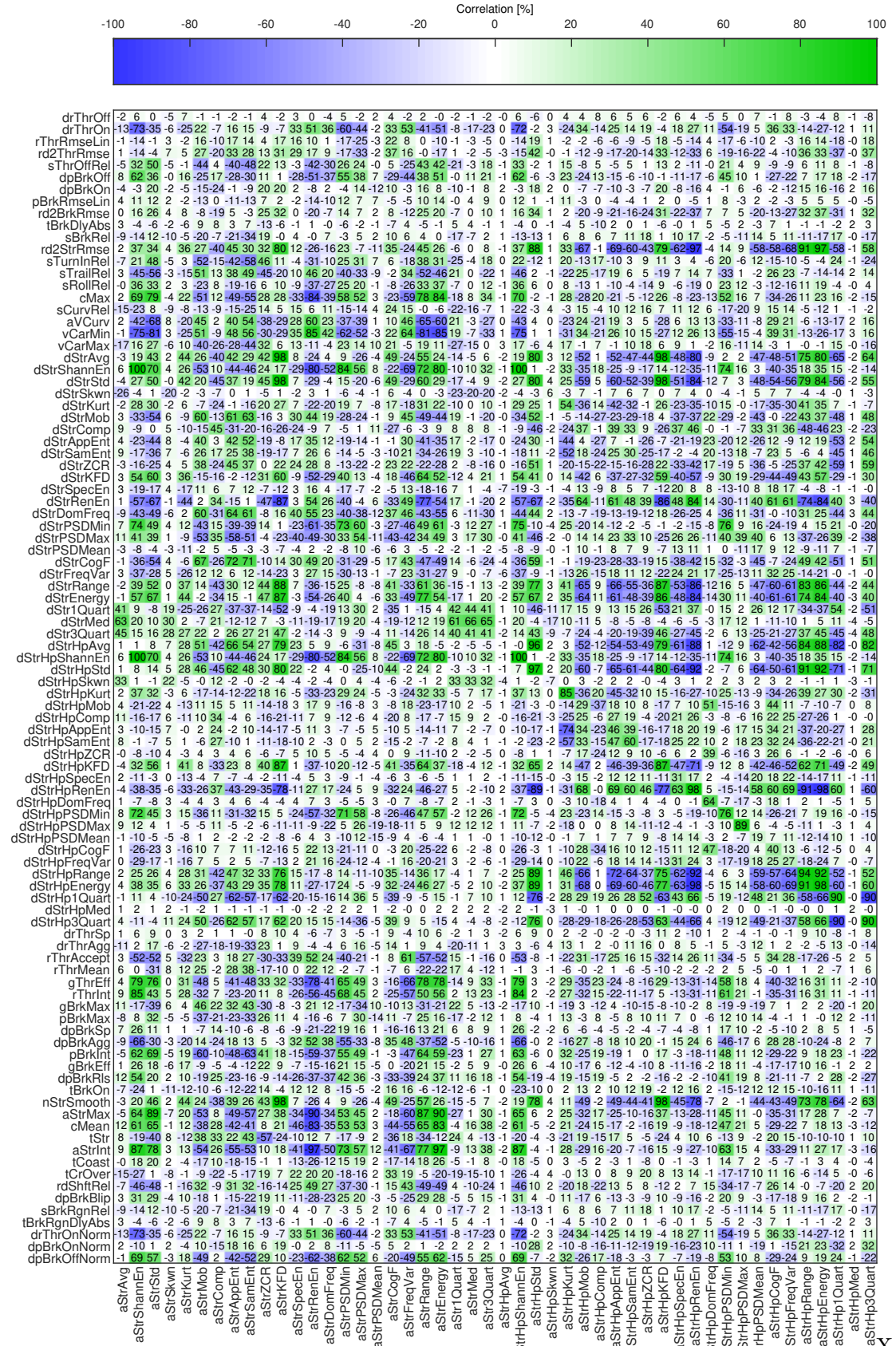


Figure A.1.: Correlation matrix of the statistical metrics based on the steering wheel angle for racing series A

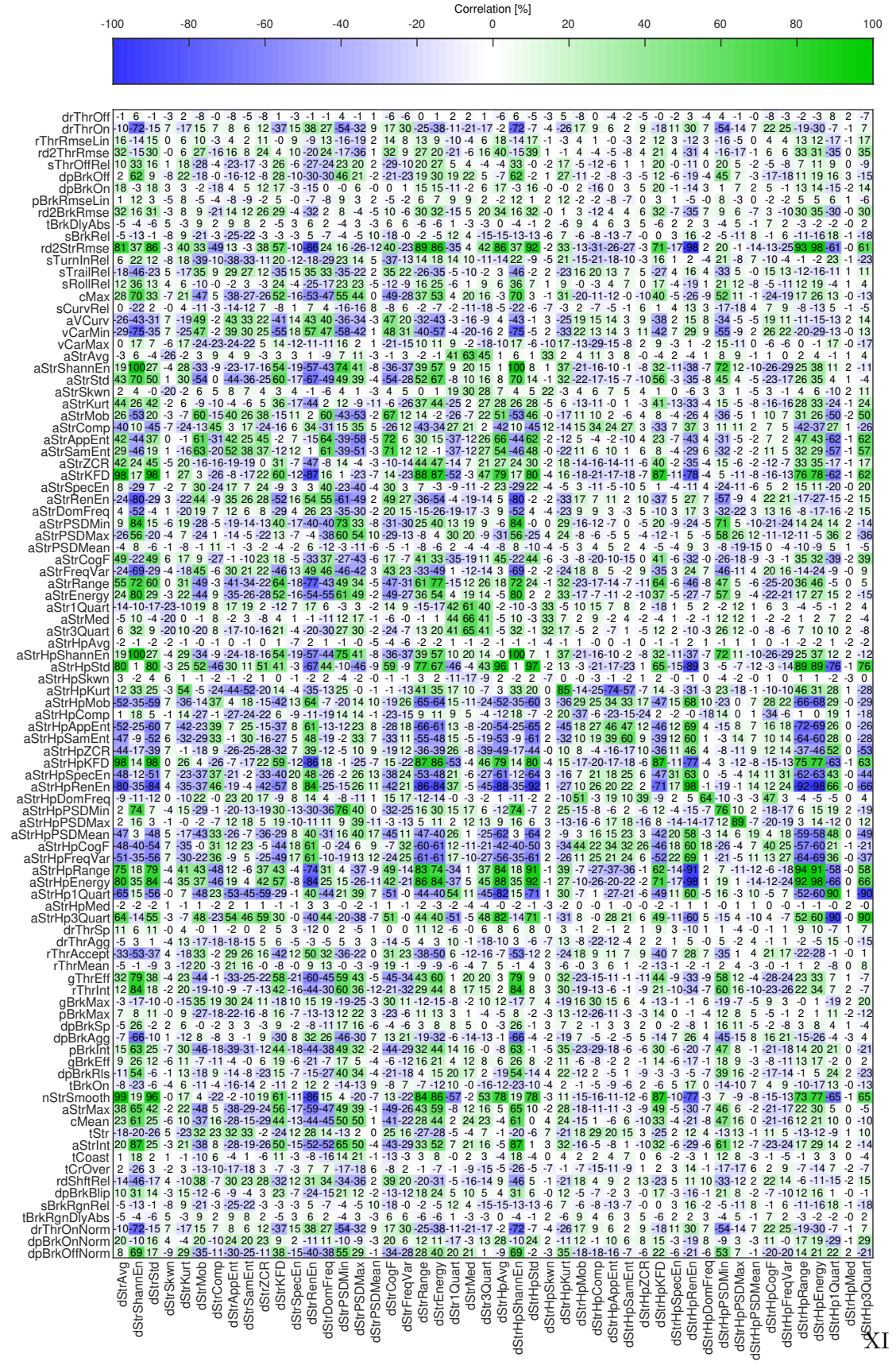


Figure A.2.: Correlation matrix of the statistical metrics based on the steering wheel angle rate for racing series A

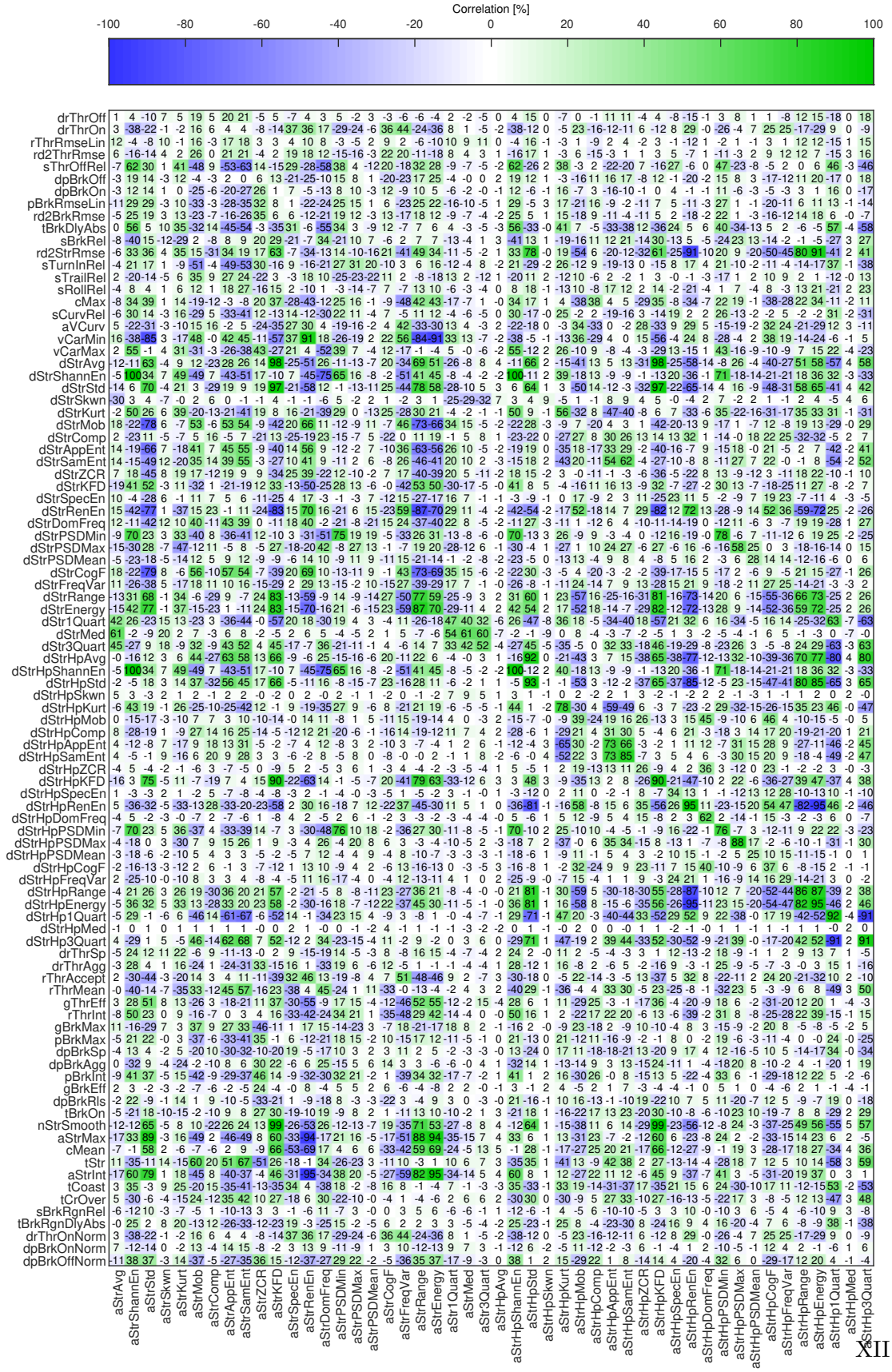


Figure A.3.: Correlation matrix of the statistical metrics based on the steering wheel angle for racing series B

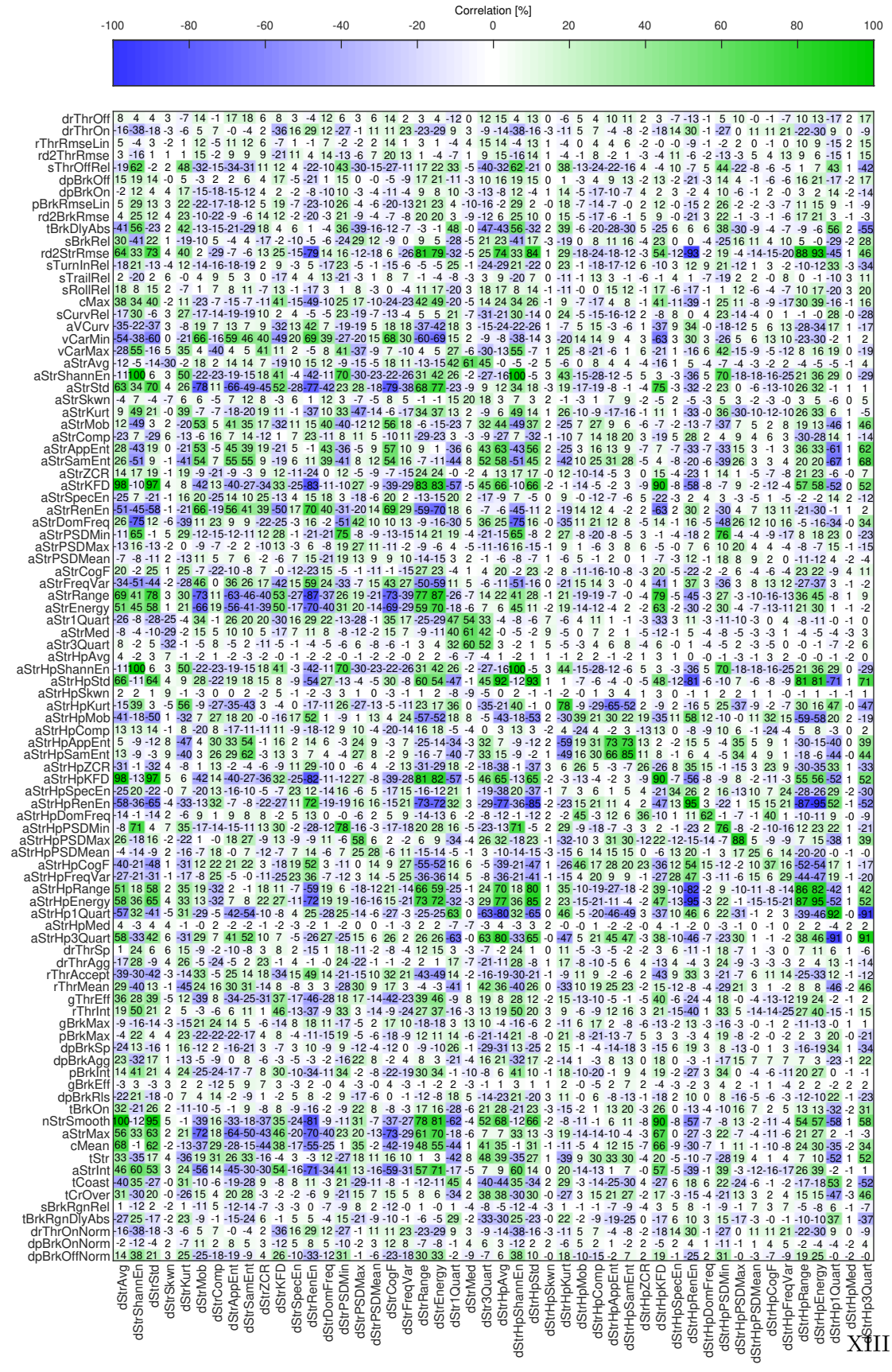


Figure A.4.: Correlation matrix of the statistical metrics based on the steering wheel angle rate for racing series B

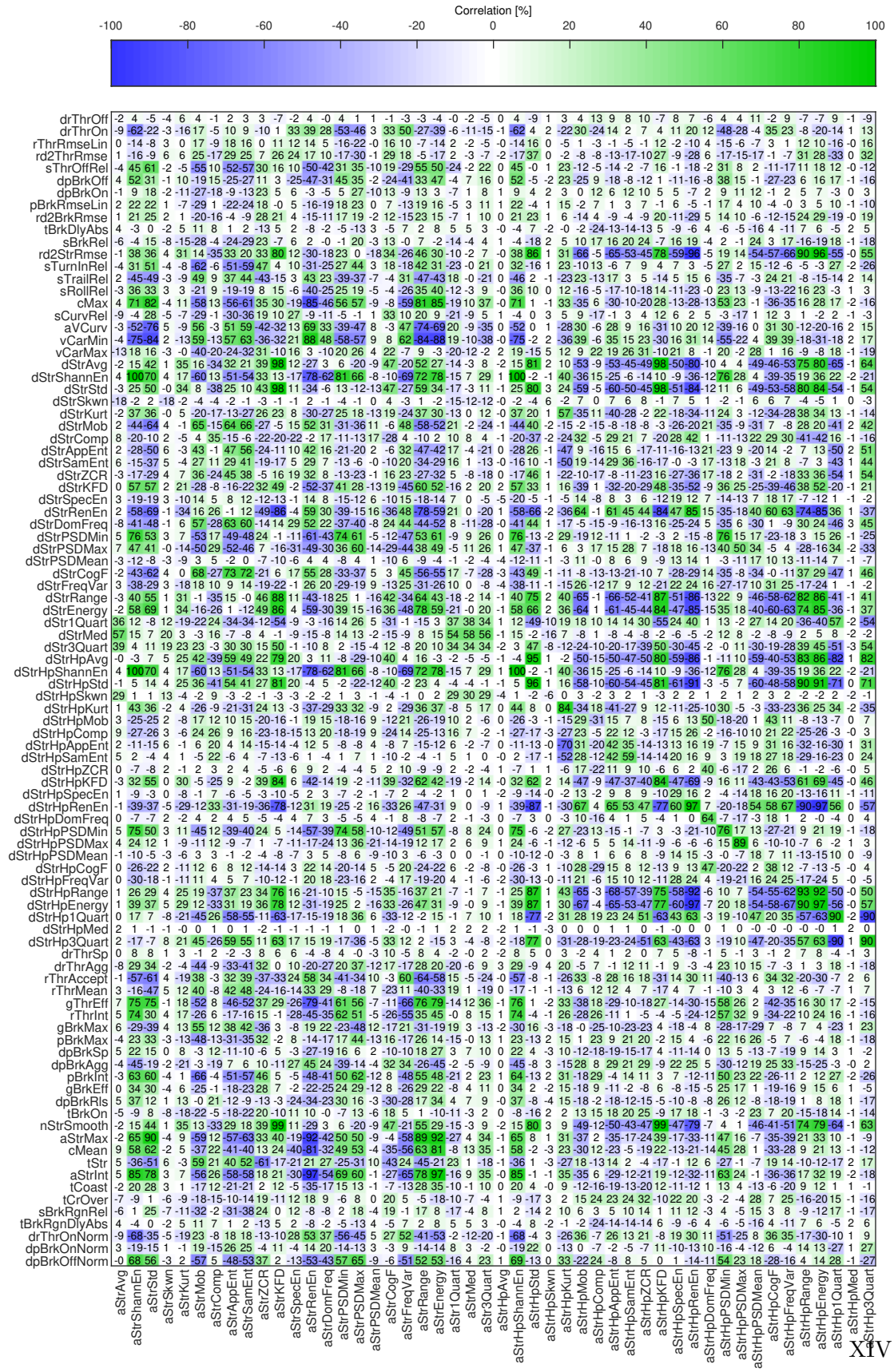


Figure A.5.: Correlation matrix of the statistical metrics based on the steering wheel angle for both racing series

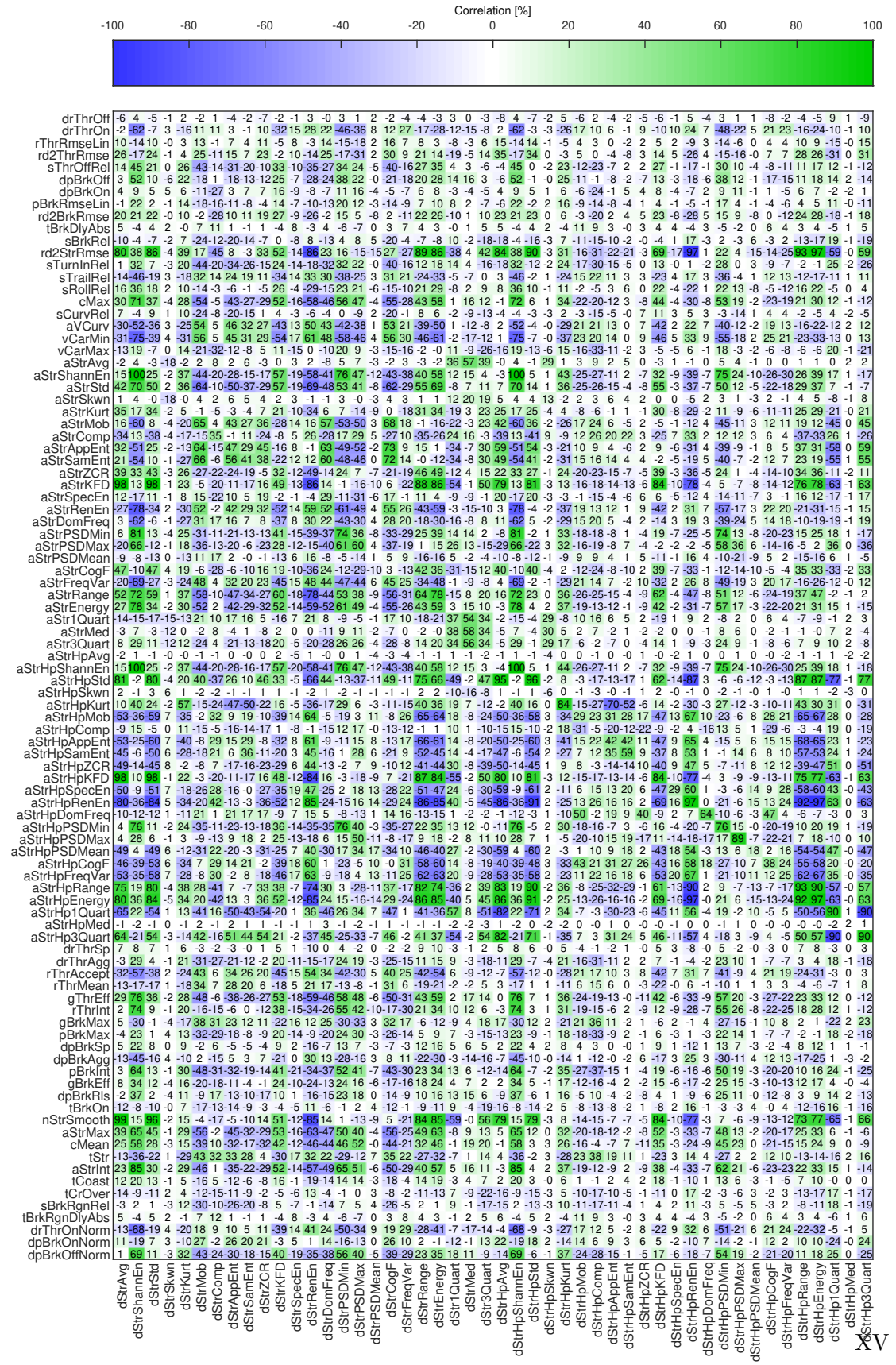


Figure A.6.: Correlation matrix of the statistical metrics based on the steering wheel angle rate for both racing series

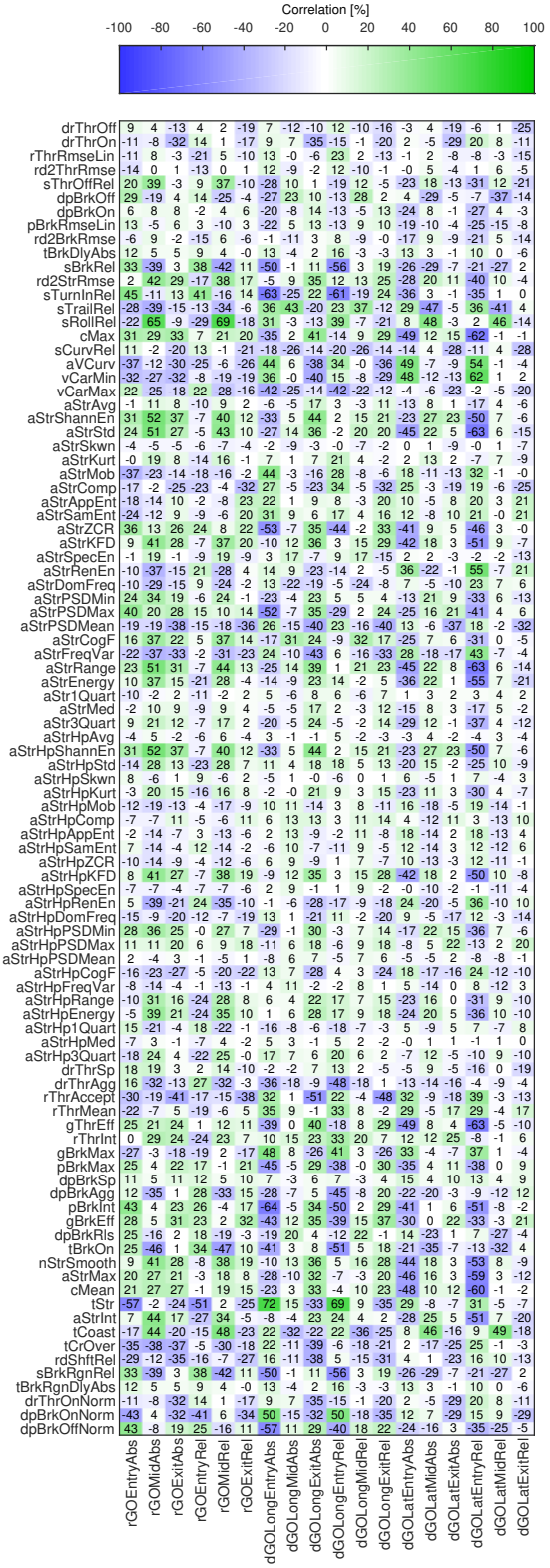


Figure A.7.: Correlation matrix of the statistical metrics based on the steering wheel angle and the grip optimization metrics for DIL simulator data of racing series A recorded during one season

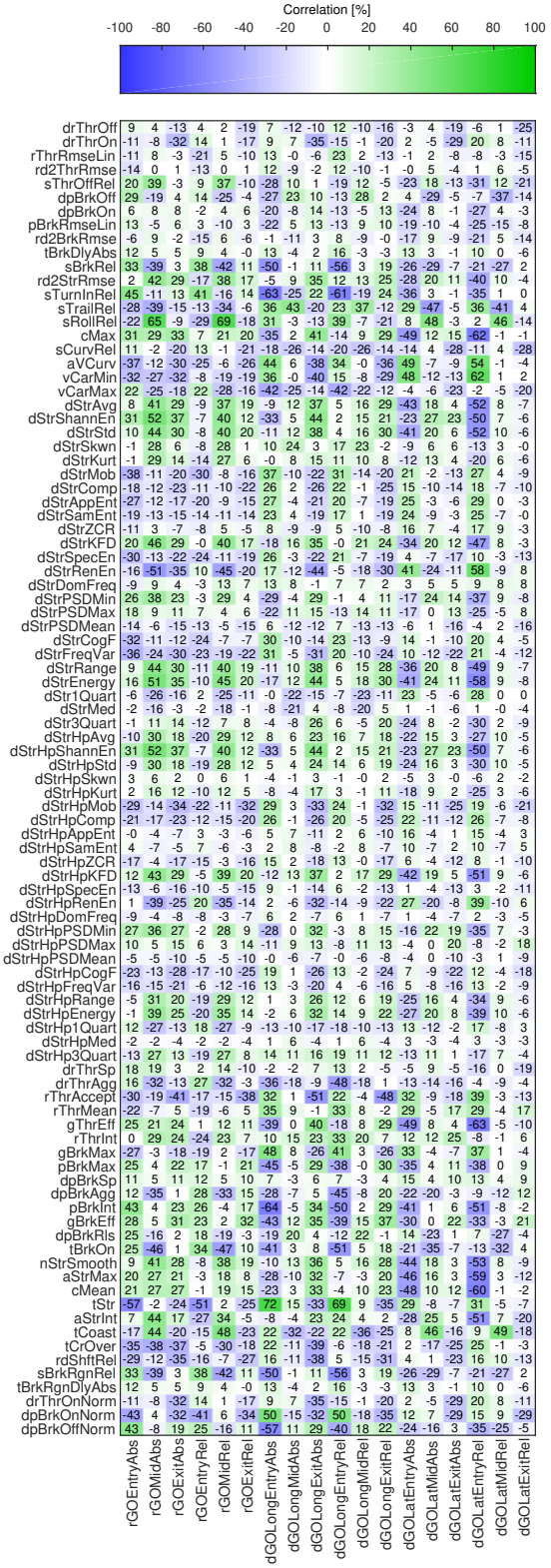


Figure A.8.: Correlation matrix of the statistical metrics based on the steering wheel angle rate and the grip optimization metrics for DIL simulator data of racing series A recorded during one season