POLITECNICO DI TORINO

Master's Degree in Automotive Engineering

Master's Degree Thesis

Probabilistic Simulator for Race Event Forecasting and Based on Telemetry and Historical Data

Politecnico di Torino

Tutors

prof. Massimiliana Carello

ing. Elia Grano

Candidate

Andrea Lorenzo Scardino

Academic Year 2024-2025

Abstract

In this thesis, a telemetry-informed race simulator for Formula One is developed to investigate the impact of strategic decisions and performance variability across full race distances. Lap time evolution is modelled as the sum of deterministic and stochastic components, calibrated using historical data and telemetry-derived parameters for fuel consumption, tire wear, driver lap time consistency and pit stop losses. Monte Carlo methods are employed to simulate multiple race iterations, reflecting the randomness inherent in motorsport events such as FCY phase, mechanical failures and overtaking maneuvers. This probabilistic modelling enables the study of not just expected results but also the confidence intervals and failure risks associated with different strategies. Through this approach, the simulator provides a framework for evaluating the robustness of race strategies and the performance envelope of each driver under varying conditions. The system lays the groundwork for data-driven predictive tools capable of supporting engineering and strategic decision-making in high variability racing environments.

Acknowledgments

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Table of contents

List of figures

List of tables

List of acronyms

F1	Formula One
FIA	Fédération Internationale de l'Automobile
VSC	Virtual Safety Car
SC	Safety Car
FCY	Full Course Yellow
FP	Free Practice
Q	Qualifying
DNF	Did Not Finish
KPH	Kilometres Per Hour
KDE	Kernel Density Estimates
PDF	Probability Density Function
SCG	SC Ghost
DRS	Drag Reduction System
MCS	Monte Carlo Simulation

Introduction

Over the past two decades, Formula 1 has undergone a substantial evolution, characterized by an increasing dependence on data acquisition, simulation and telemetry, largely due to a response to the sport's tightening regulatory framework. In an effort to reduce costs and promote fairness, the Fédération Internationale de l'Automobile (FIA) has progressively curtailed on-track development opportunities. A pivotal moment came in 2009, when in-season testing was banned entirely, forcing teams to shift much of their car development and performance validation into the virtual testing platforms. In 2018 teams were restricted to just three power units per season, despite an expanding calendar of over 20 races, placing even greater emphasis on predictive modelling and reliability analysis. The 2022 season marked a further turning point, since a substantial technical overhaul reintroduced ground effect aerodynamics, modified wheel dimensions and limited aerodynamic upgrades during the year [1]. These changes, designed to enhance competition and reduce development costs, have further elevated the role of simulation tools.

These regulatory changes have accelerated the shift towards data-driven development and simulation-based strategy planning. Modern Formula 1 teams now rely heavily on telemetry systems and computational modelling, both in the design phase and during race operations. Cars are equipped with hundreds of sensors that provide live data on vehicle dynamics, tire and brake behavior, power unit usage and energy recovery. This telemetry allows engineers to monitor performance in real time, optimize vehicle setup and respond dynamically to evolving race conditions. Beyond real-time analysis, data plays a predictive role: modelling tire degradation, estimating race pace evolution or forecasting the impact of Virtual Safety Car (VSC) and Safety Car (SC) deployments has become essential to success [2].

Within this data-centric framework, two major categories of simulators are employed: Lap Time Simulators (LTS) and Race Simulators. Lap Time Simulators replicate an individual lap under ideal conditions and are based on physics-based steady-state or quasi steady-state models. These are useful for evaluating setup changes and baseline vehicle performance, but do not incorporate traffic, degradation or multi-driver dynamics [3]. Race Simulators, by contrast, aim to replicate an entire Grand Prix through lap-wise simulations for each driver. These simulators account for tire wear, fuel mass reduction, pit stop losses, grid start randomness, overtaking interactions and the influence of Full Course Yellow (FCY) phases, such as VSC or SC. By employing Monte Carlo methods, they simulate many randomized race iterations, thereby allowing engineers to assess the robustness of a given strategy against a wide range of plausible scenarios [4, 5]. Several important research efforts have laid the foundation for modern race simulation. From early discrete-event models of pit stops and failures to more recent probabilistic simulations including SC dynamics and overtaking maneuvers, the field has progressively integrated both realism and stochastic modeling. These contributions will be analyzed in detail.

Building upon these developments, this thesis aims to contribute a further step towards comprehensive race simulation and to develop a modular and realistic Race Simulator tailored to Formula 1 applications. The simulator integrates telemetry-derived parameters and stochastic events driven by accordingly modelled logic. Each simulation is run on a per-lap, per-driver basis for a selected Grand Prix. The model incorporates random components where relevant, using statistical distributions fitted from real telemetry and timing data. The implementation is developed in MATLAB environment, while data processing and collection are performed using the FastF1 Python library.

The simulator is validated by applying it to real races from the 2023 season, while gathered data refers to previous seasons. Results, such as total race times, positional evolution, overtaking events and pit stop timings are compared against official race outcomes. Through this, the simulator's predictive accuracy, robustness and sensitivity to strategic parameters are evaluated. The final aim is to produce a

race modelling environment that is both realistic and flexible, capable of evaluating the effectiveness and repeatability of different strategic decisions under uncertainty.

1 Formula One Regulatory Framework

Formula One is the pinnacle of single-seater motorsport, founded in 1950 with the first official World Championship race held at Silverstone. Since then, the sport has undergone continuous evolution, marked by major regulatory and technical shifts. The most recent overhaul came in 2022, reintroducing ground effect and implementing cost caps and development restrictions to close the performance gap between teams. Today, Formula One is a global championship contested across more than 20 Grand Prix annually, featuring 10 teams and 20 drivers competing for both drivers' and constructors' titles [6].

A standard Formula One race weekend spans three consecutive days (Friday, Saturday and Sunday) and it is divided into several key sessions:

- Friday: Two free practice sessions (FP1 and FP2), each 60 minutes long, are held. These are primarily used for setup testing, data gathering and tire evaluation,
- Saturday: A final practice session (FP3) precedes a three-phase qualifying session (Q1, Q2 and Q3), which determines the starting grid based on lap times. The top 10 drivers in Q3 compete for pole position,
- Sunday: The Grand Prix itself is run over a distance of approximately 305 kilometers or two hours, whichever comes first.

In recent years, the format has been supplemented by Sprint weekends (which will not be treated in this thesis). On these occasions, Sprint qualifying is held on Friday ahead of a single FP session, followed by a short Sprint race on Saturday. Then the main qualifying session, which sets the grid for the main race on Sunday, is held ahead of the Sprint race [6].

Fuel in Formula One is subjected to stringent regulatory controls. Cars are powered by high-octane, unleaded petrol that must comply with a specific chemical composition defined by the FIA. Since 2022, the fuel must include at least 10% of ethanol to support environmental targets. Refueling during races has been banned since 2010, which means each car must start the race with its full fuel load. The maximum allowable fuel per race is 110 kg. Teams must also provide a 1 kg sample after the race for compliance checks, making fuel efficiency modelling a critical component for race preparation and simulation [7].

Pirelli is the exclusive tire supplier for Formula One. For each Grand Prix, three dry-weather compounds are selected from a broader range (nomenclature ranges from C0 to C5, from the hardest to the softest) and are designed as SOFT, MEDIUM and HARD for that weekend.

Each driver is allocated [1]:

- 13 sets of slick tires,
- 4 sets of intermediate tires:
    - Used on a wet track with no standing water as well as drying surface,
    - This tire can evacuate 30 liters of water per second per tire at 300 kph [8],
- 3 sets of full wet tires:
    - Used for heavy rain conditions,
    - This tire can evacuate 85 liters of water per second per tire at 300 kph [8].

Tire rules are governed by both sporting and strategic imperatives. In a dry race, drivers are required to use at least two different dry compounds, mandating at least one pit stop. During qualifying, specific compounds may be restricted to certain sessions and one set of soft tires is often reserved for use in Q3. It is not mandatory anymore to start the race with the same tire that was used during one of

the qualifying sessions. Tire degradation and compound performance differentials are among the most influential factors in determining race pace and strategy.

From the start of qualifying until the race, cars are placed under parc fermé conditions, which limit changes to the vehicle's setup. These rules are designed to ensure that cars compete in the same specification as they qualified. Teams are only permitted to make minor adjustments, such as front wing angle tweaks or brake component replacements. Key setup parameters (suspension geometry, ride height and gear ratios), must remain fixed. This regulation increases the importance of setup decisions made during the final practice session (FP3), as those configurations will carry over to both qualifying and race [9].

The outcome of an F1 race is influenced by a complex interplay of deterministic planning and stochastic elements. These include:

- Tire degradation and compound management,
- SC and VSC deployments,
- Overtaking maneuvers,
- Weather variations,
- Fuel and energy management.

Given the limited ability to adapt during a race and the cost of sub-optimal decisions, predictive modelling has become a cornerstone of competitive performance. Monte Carlo simulations, game theory [10] and machine learning can be used by the teams to forecast race evolution and evaluate strategic robustness. Race Simulator models are built upon telemetry, historical performance data and scenario analysis to determine the optimal pit window, overtaking opportunities and SC reactions.

2 State of the Art

In modern F1, the ability to simulate race weekend dynamics with accuracy is a fundamental tool for both technical and strategic success. Simulation environments allow teams to explore the behavior of the car under different configurations, test setup changes without physical limitations and evaluate alternative race strategies in advance. Moreover, predictive simulations enable the anticipation of unpredictable events, such as tire degradation, pit stop timing under FCY condition or evolving weather, supporting robust, data-driven decision-making.

The literature generally distinguishes between two principal classes of simulation tools: Lap Time Simulator and Race Simulator. These approaches are conceptually distinct but complementary in application.

Figure 2.1 clearly highlights this conceptual separation between the two simulation types, summarizing their modelling focus, scale and intended use case.

Lap time simulation	**Race simulation**
▪ Microscopic view on one lap neglecting long-term effects ▪ Physically motivated models (e.g. two-track car model, engine model) ▪ One car simulated for one lap ▪ Goal: Calculation of the exact lap time with the current configuration, e.g. to evaluate the setup	▪ Macroscopic view on the whole race including long-term effects ▪ Empirical models (e.g. tire degradation model, overtaking model) ▪ All participating cars simulated for the whole race with interactions ▪ Goal: Calculation of the final race times, e.g. to evaluate the race strategy

Figure 2.1: Comparison between lap time simulation and race simulation from Heilmeier's work [3]

Lap time simulations focus on the microscopic modelling of a single car over a single lap. These models are typically based on physically motivated formulations (e.g. multi-body dynamics, engine maps) and are used to compute the best possible lap time given a specific vehicle configuration. They neglect long-term effects and interactions with other vehicles. However, as Timings [11] notes, traditional lap time simulations assume a perfect driver and neglect real-world disturbances such as sensorimotor noise and cognitive limitations. As a result, vehicle behaviors predicted under these assumptions may be theoretically optimal but practically unmanageable. To address this, Timings introduces robustness and drivability metrics aimed at evaluating vehicle performance within a more realistic operating envelope that accounts for driver variability and external perturbations.

Race simulations take a macroscopic view, modelling the full race distance for all participating cars and incorporating long-term dynamics, such as tire degradation, fuel usage, pit stop strategies and probabilistic race events. These models usually heavily rely on empirical relation to remain computationally efficient, enabling scenario testing across thousands of conditions [3]. Vehicle dynamics is generally not taken into consideration, however they capture strategic interactions and variability essential to race planning.

Despite their different orientations, these simulators are often used in tandem. Lap time models typically generate baseline performance parameters that are fed directly into race simulation environments. Together, they create a hierarchical simulation framework that allows teams to analyze vehicle performance and strategic decision-making in a coherent and integrated manner.

The following sections will explore briefly both simulators, analyzing the main characteristics of each one of them.

2.1 Lap Time Simulator (LTS)

Lap Time Simulators (LTS) are essential tools in motorsport for estimating the minimum lap time achievable by a vehicle under a specific setup. These simulators allow engineers to predict performance outcomes and guide setup decisions long before the car hits the track. Their utility spans both the design phase, where various configurations can be tested virtually, and the development phase, where correlation with on-track telemetry can be assessed [12].

Based on Reinero's findings [12], LTS models are typically classified by their level of complexity into three main categories:

- Steady-state simulations,
- Quasi steady-state simulations,
- Transient simulations.

2.1.1 Steady-state Simulations

Steady-state LTSs are the most basic form, in which the track is divided into straights and corners and each section is analyzed assuming constant acceleration and isolated behavior (e.g. no coupling of longitudinal and lateral forces). Corners are modelled as arcs and the maximum allowable cornering speed, as shown in Equation 2.1 [12], is derived from the local radius of curvature and the vehicle's lateral acceleration limit.

$$V(i) = \sqrt{a_{y,max} \cdot R(i)}$$

Equation 2.1

This method is computationally efficient but oversimplified and thus not recommended for high-performance applications where aerodynamic and grip changes are critical.

2.1.2 Quasi Steady-state Simulations

Quasi steady-state simulations represent a refinement over simple steady-state models by introducing additional realism while avoiding the computational cost of full transient vehicle dynamics. As described by Reinero [12], this modelling technique aims to balance simplicity and fidelity, enabling a more accurate reconstruction of lap time evolution while remaining suitable for implementation in computationally efficient lap time simulators.

In this approach, the simulation is distance-based and operates along a predefined racing trajectory. The track is discretized into fine segments, typically ranging from 0.5 to 5 meters in length, each associated with a local curvature radius. This segmentation allows for segment-specific velocity calculations and enables the use of GGV diagrams, which model the feasible combinations of longitudinal and lateral accelerations. An example of GGV diagram for a race car can be seen in Figure 2.2.

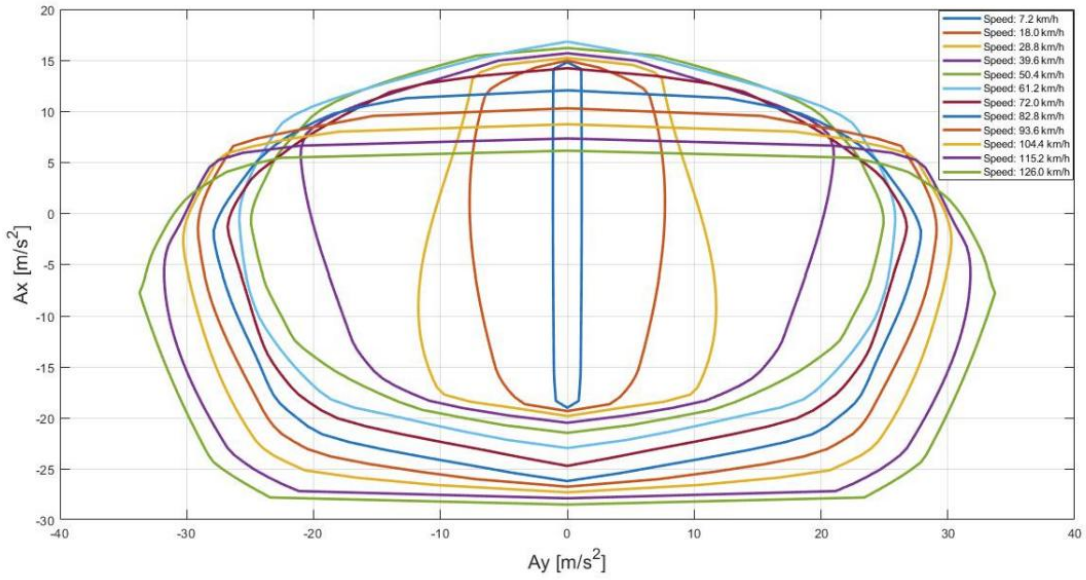

Figure 2.2: Example of a racecar GGV [12]

The procedure begins with the identification of corner apexes, defined as the local minima in the curvature profile. At these apexes, the maximum cornering speed is computed based on the lateral limits defined by the GGV diagram. Subsequently, an acceleration profile, computed by propagating forward from each apex, is constructed using the increasing radius along the corner exit and the available longitudinal acceleration to regain speed. Finally, a braking profile, computed in reverse from the next apex, is constructed by modelling the deceleration phase during corner entry under the same combined acceleration envelope.

These three profiles (cornering, acceleration and braking) are merged by selecting the minimum feasible velocity at each track segment, as shown in Figure 2.3. This ensures the vehicle remains within its physical limits throughout the lap.

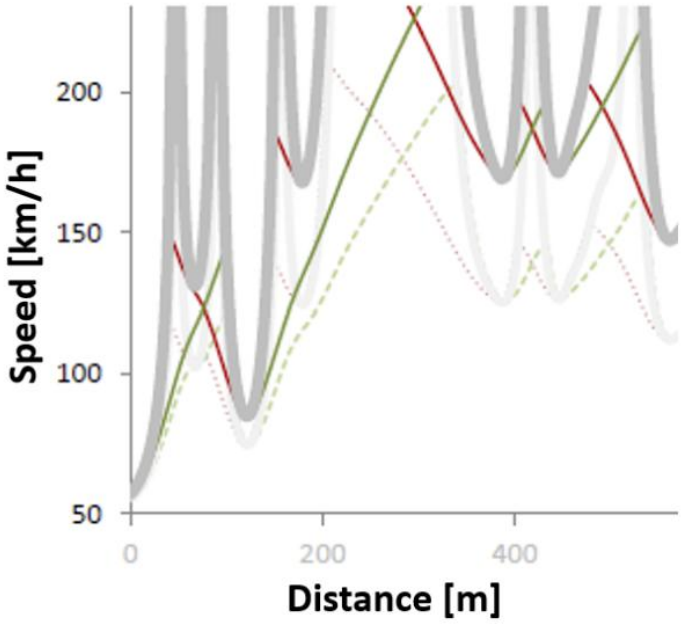

Figure 2.3: Intersection of the three speed profiles [12]

The result is a full-lap simulated velocity profile, illustrated in Figure 2.4.

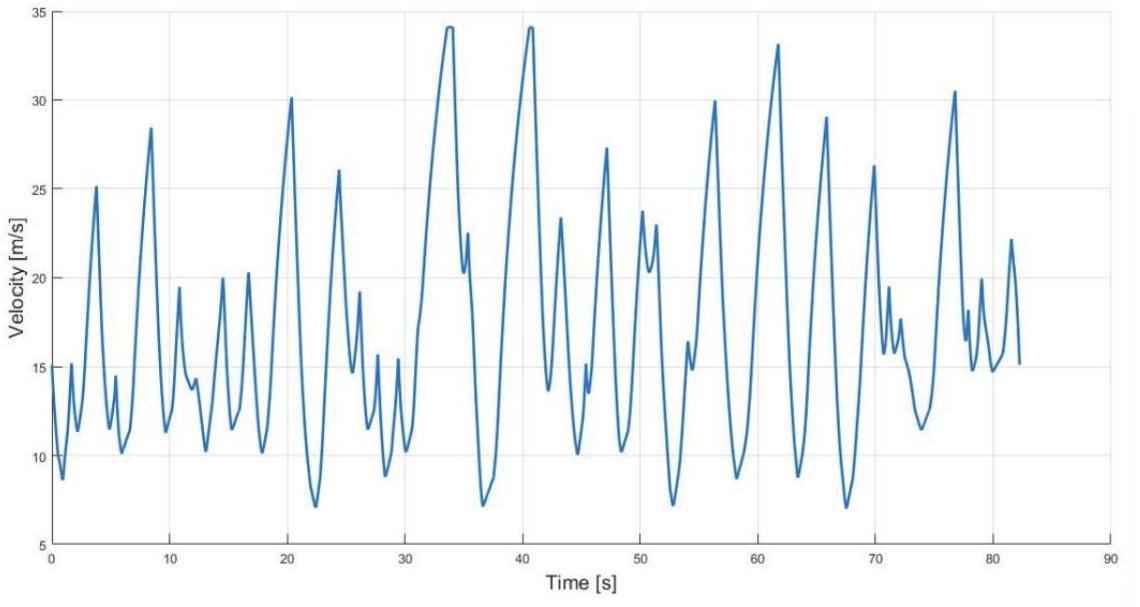

Figure 2.4: Simulated velocity profile [12]

As Reinero [12] notes, the level of vehicle modelling can vary depending on the intended application. A point-mass model is typically sufficient for use cases such as target setting or fuel and energy consumption estimation, while more advanced models, such as bicycle or four-wheel vehicle model, allow for simulation of load transfer dynamics and more accurate representation of tire behavior.

Overall, quasi steady-state simulation achieves a valuable compromise: it improves correlation with real-world data without incurring the computational burden of transient modelling. Reinero [12] highlights this as a key advantage for motorsport contexts, where flexibility, speed and model transparency are prioritized over black-box commercial solutions.

2.1.3 Transient Simulations

Transient simulations provide the highest level of fidelity by fully resolving the time-varying dynamics of the vehicle. Unlike steady-state methods, they do not assume instantaneous equilibrium. Instead, they solve the vehicle's motion through differential equations that account for unbalanced forces and moments. Full six-degree-of-freedom models, including pitch, roll and yaw, are often implemented, along with suspension, tire and aerodynamic subsystems modelled as functions of dynamic state. Driver behavior is typically incorporated using control loops for steering, throttle and braking inputs, aimed at following an optimal trajectory.

While more complex, transient models enable the integration of control strategies, component feedback and driver in-the-loop testing. However, their computational cost and data requirements limit their use to specific applications where such precision is justified.

2.2 Race Simulator

Race simulators are designed to reproduce the entire race duration by capturing long-term effects, such as tire degradation, pit stops, fuel mass loss and interaction between drivers. Unlike LTS, which operates on a single-lap, physics-heavy basis, race simulators adopt a macroscopic, empirical approach, enabling the evaluation of strategy choices across multiple laps and participants [3].

2.2.1 Brief Simulator Description

As Figure 2.5 can highlight, race strategy aspects can be divided into three categories:

- Pit stops: This category includes all decisions related to the timing and number of pit stops, as well as the tire compound choice. These factors directly influence race time due to time loss during pit entry, service and rejoining the track,
- Driving strategy: It refers to how aggressively or conservatively a driver is instructed to manage the car during the race. This affects both energy consumption and tire wear. An aggressive strategy may yield faster lap times but accelerate degradation, while a conservative approach preserves resources for later race phases or strategic flexibility,
- Response to race events: This includes the team's and driver's ability to react to unpredictable events, such as overtakes and FCY phases, where the latter can compress the field and alter the strategic landscape, opening or closing optimal pit windows and changing the value of track position.

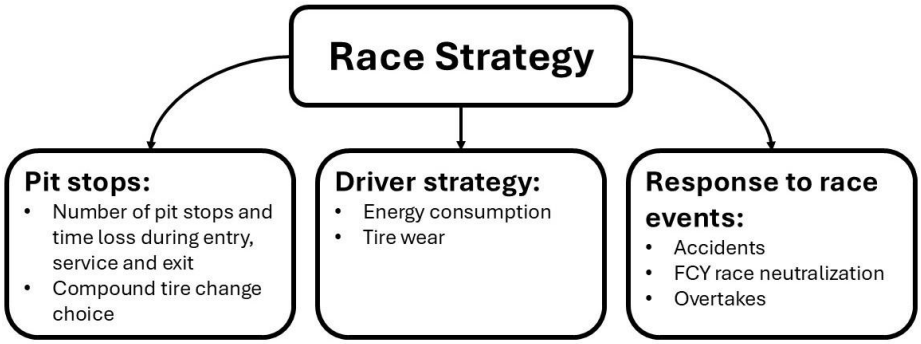

Figure 2.5: Aspects of race strategy [3]

The race is discretized lap-by-lap, allowing efficient simulation while maintaining sufficient detail for strategic assessment. Each lap includes some core phases, as depicted in Figure 2.6:

- Occurrence of probabilistic effects,
- Computation of drivers' lap times using aggregated models,
- Position update following race dynamics, including FCY, pit stops and overtakes.

The following main inputs are fed into the race simulator:

- Starting compound choice,
- Predicted pit stop strategy.

The main output is race duration, but also other useful information can be obtained, such as position, gap from driver in front and strategy modification due to FCY phases.

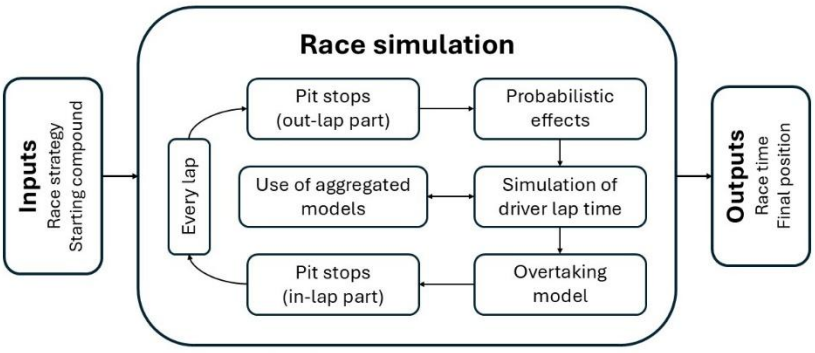

Figure 2.6: Race simulation workflow [3]

The use of simplified empirical models enables fast computation and easy integration of stochastic effects. Overtaking, pit strategy and driver-specific behavior can be layered through modular components, as demonstrated in recent works such those of Heilmeier [4], Bekker [13] and Sulsters [5].

2.2.2 Overview of Existing Race Simulation Approaches

Race simulators found in literature differ significantly in how they model race dynamics, especially under uncertainty. While most share a common lap-wise simulation structure and use empirical models for efficiency, their treatment of stochastic and event-driven effects is not uniform.

As evaluated by Heilmeier [4] and summarized in Table 2.1, six key race phenomena are often modelled with varying levels of completeness: starting performance, lap time variability, pit stop duration variability, accidents and failures, damaged car dynamics and FCY phases.

Among the reviewed simulators:

- Bekker [13] introduces a foundational time-based simulator incorporating basic modelling of starting variability, pit stop duration and failures, but does not address other race dynamics such as yellow flags and vehicle damage,
- Phillips [14] extends Bekker's model to a Formula One context, achieving excellent modelling of lap time and pit stop variability with partial implementation of failures and yellow flags,
- Salminem [15] offers a balanced approach, modelling lap time variability in detail, partially modelling accidents and failures, damaged cars and yellow flag phases, but incorrectly simulating pit stops,
- Sulsters [5] provides a strong coverage of lap time variability and overtaking models, but yellow flag phases and pit stop duration are only partially addressed.

Heilmeier proposes a simulation framework that integrates and improves all these aspects, forming a more complete modelling environment. By combining empirical sub-models with Monte Carlo methods, his implementation handles:

- Stochastic variability in driver performance and lap evolution,
- Pit stop timing and duration uncertainty,
- Failure rates and accidents scenarios,

- Dynamic responses to SC and VSC phases

This level of integration allows for a more robust assessment of strategy outcomes across a broad set of race scenarios, representing a step forward in simulation fidelity compared to earlier models.

Table 2.1: Overview of race phenomena modelling in literature [3]

Modelled Effect	Bekker [13]	Phillips [14]	Salminem [15]	Sulsters [5]
Starting performance	◔	◑	◑	◔
Variability of lap time	○	●	●	●
Variability of pit stop duration	◔	●	◔	○
Accidents and failures	◔	◑	◑	◔
Damaged car	○	○	◑	○
Full Course Yellow phases	○	◑	◑	◔

3 Methodology

This thesis project presents the development of an updated and modular race simulation model tailored to the current context of F1 racing. The simulator is specifically designed to reflect the characteristics of the 2023 season and it aims to provide a more accurate and flexible framework for evaluating race strategies under contemporary conditions.

The motivation for this update lies in the substantial regulatory changes introduced in 2022, which have significantly impacted vehicle behavior and strategic planning. The return of ground-effect aerodynamics, the switch to 18-inch wheels and the implementation of cost caps and development restrictions have redefined how teams approach performance optimization. These changes have altered how races unfold and are managed strategically, especially in terms of driver performance variability, pit stop timing and race-event handling.

To support this objective, the first step in this thesis involves a comprehensive data acquisition phase. Data is extracted using the Python-based FastF library, which provides access to historical telemetry, timing data, pit stop logs, tire usage and session metadata. In addition to retrieving raw lap times, telemetry is used to derive compound-specific stints and pit lane travel durations. These data are pre-processed and structured to serve as direct inputs to the simulator.

Once the data is organized, a modular race simulator is constructed using MATLAB. The simulation operates on a lap-by-lap basis, with each driver's race time accumulated incrementally. The simulator integrates several empirical sub-models that influence lap time evolution: tire degradation, fuel mass reduction, pit stop time loss and driver specific variability. All these components interact to produce a realistic approximation of race progression and timing. Strategic inputs such as pit stop laps and tire compound selection are provided manually and each lap update dynamically adjusts the race state based on evolving conditions.

In order to reflect the inherent variability of real-world races, probabilistic effects are incorporated into the simulation. These include random pit delays, mechanical failures, inconsistent starting performance and the possibility of FCY intervention. Each of these components is modelled using empirically grounded probability distributions derived from historical event frequency or variance in telemetry data. This probabilistic layer enhances the model's realism and its ability to simulate a wide range of plausible race outcomes.

The developed simulator is first validated through a single-event simulation, selecting arbitrary the Grand Prix to simulate. The simulation output includes the total race durations, final classifications and a trace of in-race events such as overtakes, pit stops and FCY phases.

Finally, to evaluate the robustness of the model, Monte Carlo method is applied to the simulator, to repeat it over a large number of iterations. The resulting distribution of race outcomes is then analyzed to assess the simulator's stability and predictive power. These results are compared with real race data from the 2023 Formula One season to determine how closely the simulation output replicates the actual race dynamics and classifications.

This methodology provides a foundation for developing a race strategy optimization simulator that is not only aligned with modern Formula One regulations and data but also robust in other possible motorsport scenarios. The subsequent chapters will detail the data handling procedures, modelling assumptions and validation outcomes that support the use of this simulation tool for performance and strategy evaluation.

4 Data Acquisition

The development of a race simulation model relies on a robust and coherent dataset that accurately captures both the structural attributes of race events and the performance behavior of drivers and teams across a championship season. Since one of the objectives of this thesis is to construct a simulator reflective of the modern Formula One environment, the 2023 season has been selected as the reference framework. This choice ensures alignment with the current technical regulations, team configurations and tire allocation schemes.

To support this endeavor, extensive data acquisition and preprocessing are required. A wide array of variables must be considered, ranging from circuit information, lap time distributions, tire usage and driver classification results. These parameters form the foundation upon which key simulation models, such as lap time calculation, tire degradation estimation and stochastic event handling are built.

The primary tool used for gathering time-series, session and telemetry data is the Python-based FastF1 library [16], which provide access to publicly available Formula One data though an interface with the official F1 timing service. Additional performance metrics, such as team and driver retirement rates, have been integrated using external sources including StatsF1.com and selected academic literature [4, 5, 14, 17].

This chapter outlines the structure of the datasets collected, the rationale for their inclusion and how each dataset contributes to the modelling and calibration of the simulation work.

For each Grand Prix in the 2023 season, circuit data were gathered on round number inside the season, circuit name, circuit classification (whether it is street or permanent), number of race laps and lap distance in kilometers.

An estimate of pit lane trave time was also computed for each venue. This was done by analyzing telemetry data from race sessions to identify sections where the car was under pitlane speed limit conditions (depending on the circuit, it can be either 80 kph or 60 kph). By calculating the distance covered while the speed limiter was active and dividing it by the constant pit lane speed, an approximation of pit lane trave time was obtained. A reference to the content inside the telemetry file obtained using the FastF1 library can be shown in Table 4.1.

Table 4.1: Example content inside the data file relative to circuit-specific telemetry laps

Session Time	Time	Speed [kph]	Distance [m]
0 days 01:23:54.316	0 days 00:01:41.295	302	3384.901

In addition to physical track characteristics, qualitative weather conditions during both qualifying and race sessions were recorded (categorized as dry or wet). This parameter relies his importance on the successive data utilization of the simulator for the training set, since one of the limitation of this kind of simulator is to be suitable for dry weather conditions only, struggling to model wet weather conditions due to high variability in track surface grip level and quantity of standing water.

Tire allocation was also tracked for each race using the C-nomenclature (C0, C1, C2, C3, C4, C5) provided by the official tire supplier, Pirelli. The compounds used at each event were mapped under three standard categories, SOFT, MEDIUM, HARD. A reference of the content inside the circuit data file obtained using the FastF1 library is shown in Table 4.2.

Table 4.2: Example content inside the data file relative to circuit information

Round	Circuit Name	Type	Laps	Lap Distance [km]	Pit Lane Travel [s]
01	Bahrain Gran Prix	Permanent	57	5.412	20.057

Qualifying Weather	Race Weather	SOFT	MEDIUM	HARD
Dry	Dry	C3	C2	C1

To establish baseline driver performance, the fastest non-deleted lap time for each driver was extracted from the qualifying session of each Grand Prix. For every entry, the dataset includes the round number, driver name, lap time, circuit name, qualifying weather and race weather conditions. These lap times are later used to derive performance differentials between drivers and to support the initialization of base lap times within the simulator. A reference to the content inside the driver-specific qualifying data file obtained using the FastF1 library can be shown in Table 4.3.

Table 4.3: Example content inside the data file relative to driver-specific qualifying information

Round	Driver	Lap Time	Circuit Name	Qualifying Weather	Race Weather
01	ALB	0 days 00:01:31.461	Bahrain Grand Prix	Dry	Dry

Historical driver performance was analyzed over the 2022 and 2023 seasons to identify patterns of race completion and retirement. For each driver and each event, the starting grid position, final classification status (finished or not), round number, circuit, year and race weather were recorder. This dataset enables the modelling of driver-specific retirement probabilities, including accident-driven DNFs, as part of the simulator's probabilistic layer. A reference to the content inside the driver-specific results data file obtained using the FastF1 library can be shown in Table 4.4.

Table 4.4: Example content inside the data file relative to driver-specific race results information

Round	Driver	Starting Position	Finished	Circuit Name	Year	Race Weather
01	ALB	14	Yes	Bahrain Grand Prix	2022	Dry

Race time data were collected for all drivers across the 2023 season, recorded by lap-by-lap timings. Two different collections were performed, one including all laps of the race (considering also eventual SC and VSC laps and in-laps and out-laps relative to pit stops) and the other one excluding them by creating a filtered dataset following the 107% rule [18]. According to this regulation, only laps withing the 107% of the session's best lap time are retained for analysis, unless explicitly excluded by race control. This filtering allows for a cleaner estimation of actual race pace by reducing the influence of abnormal laps. A reference of the content inside the circuit-specific and driver-specific clean laps data file obtained using the FastF1 library can be shown in Table 4.5.

Table 4.5: Example content inside the data file relative to circuit-specific and driver-specific clean laps information

Driver	Lap Time	Lap Number
ALB	0 days 00:01:40.430	2

Tire usage data were compiled for each driver across the 2023 calendar, separated by compound type. For each tire compound (SOFT, MEDIUM, HARD) the collected data include driver name, lap time, stint number, compound name and the lap number relative to when the compound was first mounted. These data were subsequently reclassified inside MATLAB environment to align with the standardized C-nomenclature, based on compound allocations defined per circuit. It is important to note that the initial wear state of the tires at the time of fitting is not known and, due to the absence of such information, all stints are assumed to begin with new tires. While this simplification may introduce minor inaccuracies, it is a necessary assumption given current data limitations. A reference of the content inside the compound-specific, circuit-specific and driver-specific stint lap time data file obtained using the FastF1 library can be shown in Table 4.6.

Table 4.6: Example content inside the data file relative to compound-specific, circuit-specific and driver specific stint lap times information

Driver	Lap Time	Stint	Compound	Lap Number
ALB	0 days 00:01:37.503	3	HARD	28

Finally, supplementary failure and retirement data are retrieved. Two categories are used:

- Driver-specific accident data from the StatsF1.com website, including the number of Grand Prix starts, retirements and relative retirement percentage,
- Team-level failure data, used to model mechanical breakdowns and reliability, which are actually derived from statistics referenced in Heilmeier's work [4].

With the data infrastructure now established, the next step involves the construction of the simulator itself. The following chapter will present the simulation architecture, explain how the collected data are integrated into the modelling logic and describe the key components that govern driver behavior, race timing and strategic dynamics within the simulation environment.

5 Simulator Structure and Modelling

After gathering the necessary data and information to prepare the simulator building, the actual structure of the simulator itself can be modelled. This thesis presents the development of a race simulator tailored to predict driver-specific outcomes under well-defined GP conditions and, although the simulator is developed for application within Formula One, the modelling principles adopted here are broadly extensible to any closed-circuit, multi-participant motorsport discipline. What has to be accounted for is the correct setup data to be retrieved in order to correctly perform the simulation and model the needed parameters.

The goal of the simulator is to forecast race results prior to the event by integrating strategic, empirical and probabilistic modelling layers. The outputs of each simulation include:

- Final classification and position changes across the race,
- Total race time for each driver,
- Number of successful overtaking maneuvers,
- Strategy executed by each driver in terms of pit stops and tire usage.

To achieve this, the simulator models each driver's race performance as the cumulative sum of lap-by-lap results, integrating multiple influences that contribute to lap time variability. These influences include physical parameters (such as fuel mass or tire age), stochastic elements (such as DNFs and FCY phases) and strategic inputs (such as pit stop timings). The simulator is structured to follow a lap-based discretization, which simplifies modelling while retaining sufficient resolution for strategic analysis.

The used approach follows the architecture proposed by Bekker and Lotz [13], who modelled Formula One races as discrete-event systems by summing timed events across race segments. However, while their model used pre-2010 regulations, the simulator here developed builds upon their structure using updated parameters and race logic consistent with the 2023 season. Similarly, the framework incorporates methodological insights from Heilmeier's Monte Carlo-based design, which emphasizes robustness over determinism by including key probabilistic effects such as lap time variability, DNFs and FCY conditions [4]. Sulsters' implementation also informs this work by highlighting practical challenges in empirical modelling, such as starting grid mixing, overtaking and pit stop simulation.

The simulator operates on several assumptions [5]:

- Each driver begins with a predefined strategy regarding tire compounds and pit stop laps,
- Cars do not experience the slipstream effect, as it is not possible to model this phenomenon as input parameter of this simulation model due to the lack of available data,
- Fuel consumption of each car remains constant for the duration of the race,
- Driver performance is highly affected by the qualifying time,
- Simulation model can only acquire data from races in dry weather conditions and it does not change for the entire duration of the race,
- Each driver can overtake only one car per lap, except for the first lap (affected by mixture of cars at the end of the first lap due to the start of the race) and for the pit stop laps (in which drivers can lose and gain different positions due to the long wait at the pits),

The structure of the lap time calculation follows the decomposition proposed by Bekker [13] and adopted in subsequent studies, such as Sulsters' one [5]. Each lap time is expressed as the sum of several contributing terms:

$$t_{lap}(l) = t_{base} + t_{fuel}(l) + t_{tire}(a_{tire}, c_{tire}) + t_{car} + t_{driver} + t_{grid}(l, p_g) + t_{pit,in/out}(l)$$

Equation 5.1

Where:

- t_{base} is the base lap time, representing the minimum achievable time for the driver under ideal conditions: fresh tires, low fuel and no external disturbances. This value is driver-specific and circuit-specific and is typically extracted using qualifying data,
- $t_{fuel}(l)$ captures the effect of fuel mass reduction. As fuel burns off during the race, the car becomes lighter, typically resulting in a reducing lap time loss over the course of the race duration. This term is modelled as a monotonic function of lap number l, decreasing in magnitude as the race progresses,
- $t_{tire}(a_{tire}, c_{tire})$ accounts for the performance degradation due to tire wear. It depends on the compound age a_{tire} and type c_{tire}, reflecting its stiffness and degradation profile. The function to obtain it is calibrated using empirical stint data from race telemetry,
- t_{car} and t_{driver} reflect the vehicle-specific performance deltas and the driver's intrinsic variability and consistency, accounting for their tendency to oscillate around the expected pace. Drivers with higher variability or lower skill may accumulate lap time losses even under optimal conditions. This component may be modelled deterministically or stochastically, based on observed lap time distributions,
- $t_{grid}(l, p_g)$ describes the impact of starting position. During the first lap, the simulator accounts for the distance between the driver's grid position and the start/finish line through a deterministic offset. It also considers the phenomenon of start phase mixing, which captures the random gain or loss of positions caused by interactions among cars at the race start. This effect is represented probabilistically, using statistical distributions derived for historical data on race starts [5],
- $t_{pit,in/out}(l)$ represents time losses associated with in-lap and out-lap phases during pit stops. It includes the time required to enter the pit lane, perform the pit stop (including potential random delay due to pit crew efficiency) and rejoin the track. These time deltas are applied only to laps affected by pit activity and change based on the pit lane length.

In addition to these structured parameters, several probabilistic race dynamics are incorporated to improve realism:

- FCY events (SC or VSC) are simulated based on probability distributions on occurrence and duration of each phase and lead to a temporary and global increase in driver-specific lap time computation,
- Accidents and mechanical failures are modelled using historical DNF rates at both driver and team levels. When triggered, they cause the driver's simulation to terminate at the respective retirement lap,
- Overtaking logic is handled outside the lap time equation but affects driver positions between laps. Circuit specific thresholds and probabilities are used to evaluate whether an overtake attempt is made and whether it succeeds.

Equation 5.1 defines the core of the race simulator. Its modularity allows each term to be independently calibrated and updated, ensuring the flexibility to adapt to different datasets, seasons or categories. The next sections will detail each of these modelling components.

5.1 Base Lap Time Model

The base lap time serves as the core reference for modelling race lap times during the race event. It represents the theoretical lap time a driver could achieve under ideal race conditions: clean air, no degradation, minimal fuel load and no interactions or disturbances. This value provides the foundation of all other time-related modifications, such as tire degradation, pit stop penalties or fuel burn benefits.

As proposed by Bekker [13] and further formalized by Heilmeier [3], the base lap time is commonly derived from the qualifying session, where conditions are most controlled and performance is maximized. During qualifying, drivers run on low fuel, fresh tires and are fully committed to extracting peak performance, making qualifying laps a strong proxy for a car's intrinsic speed potential. However, race laps are inherently slower due to factors such as fuel management, engine preservation, race traffic and strategic constraints. Therefore, a correction term must be applied to bridge the gap between qualifying and race conditions.

In this thesis, the formulation presented in Equation 5.2 follows the original approach introduced by Heilmeier [3], which explicitly model the time gap between race pace and qualifying pace:

$$t_{base} = t_{qualifying} + t_{gap,racepace}$$

Equation 5.2

Where:

- $t_{qualifying}$ is the fastest recorded qualifying lap for a given driver and circuit during the 2023 season.
- $t_{gap,racepace}$ is a deterministic correction term that accounts for the average time loss from qualifying conditions to actual performance.

To quantify this correction term, a statistical analysis was performed comparing each driver's qualifying lap with their fastest lap during the final phase of the race, when the car carries its lowest fuel load, resulting in reduced weight and therefore faster lap times. In this phase, drivers who have established a comfortable gap to the car behind may elect to make an additional pit stop to fit a fresh set of softer compound tires. This strategy aims to maximize grip and exploit the full performance potential of the tires over a single flying lap. The motivation for this tactic is the additional championship point awarded for achieving the fastest lap during the race. Consequently, teams and drivers often weigh the strategic advantage of securing this point against the minimal time loss from an extra pit stop, particularly if their track position is secure. Such late-race attempts at the fastest lap combine the advantages of lower fuel mass and optimal tire performance to achieve the quickest possible time.

The statistical analysis approach avoids the variability of mid-race conditions, such as traffic or race management strategies and provides a cleaner estimation of race pace potential. The lap time gaps (defined as the difference between the fastest race lap times and the qualifying lap regarding the same driver) were collected across a wide range of drivers and races, then aggregated into a distribution for analysis.

Figure 5.1 presents the resulting distribution of time gaps between qualifying laps and fastest race laps. The histogram displays the normalized frequency of these gaps across all sampled events. A smooth red line denotes the fitted probability density curve, while the vertical blue dashed line highlights the mean value, which in this case is approximately equal to $\overline{t_{gap}} = 3.413\ s$. This indicates that, on average, the fastest lap a driver achieves during the race is roughly 3.4 seconds slower than their qualifying best. This delta primarily captures differences in car mode and engine settings

between qualifying and race sessions, such as reduced engine mapping, conservative driving programs and less aggressive deployment of energy recovery systems, implemented to preserve reliability over race distance.

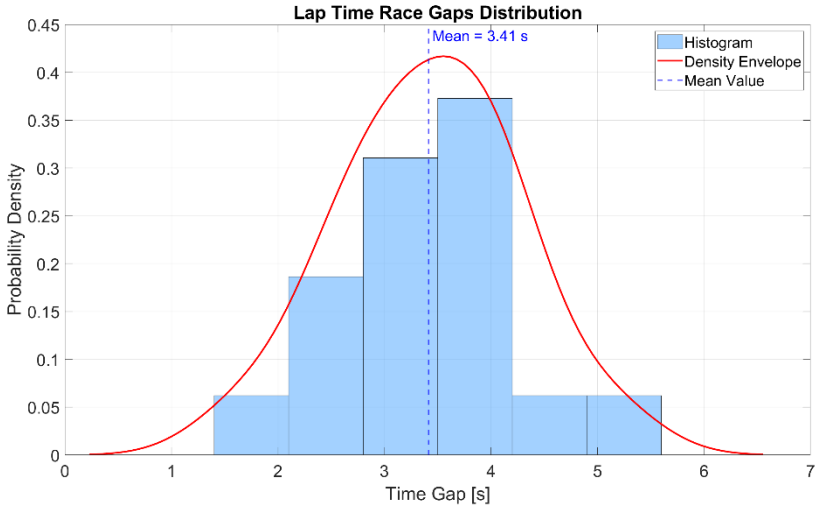

Figure 5.1: Distribution of time gaps between qualifying and fastest race laps. The mean value (dashed blue line) is used as a deterministic race pace correction

The choice to use the mean value from this distribution is based on its ability to represent the expected value across varying circuits and driver styles. While using the median or a percentile could limit the influence of outliers, the shape of the distribution in Figure 5.1 is approximately symmetric and centered, justifying the use of the arithmetic mean as a reliable and interpretable statistic for this purpose.

Thus, for each driver and circuit combination, Equation 5.3 allows to express the final base lap time, computed as:

$$t_{base} = t_{qualifying} + \bar{t}_{gap} \approx t_{qualifying} + 3.413\,s$$

Equation 5.3

This formulation ensures a consistent and realistic foundation for modelling race performance while preserving driver-specific and circuit-specific variation.

5.2 Fuel Consumption Model

An essential contributor to lap time variation over the course of a race is the evolving fuel mass carried by the car. At the start of a Grand Prix, cars are filled with enough fuel to complete the entire race distance, often close to the regulatory limit of 110 kg [19]. As the race progresses, this mass is gradually reduced, resulting in a measurable performance gain due to lower inertia and reduced tire load. To account for this effect, a dedicated fuel model is implemented in the simulator, which modifies the lap time of each drier based on the amount of fuel consumed at any given lap.

This modelling approach follows the structure proposed by Heilmeier [3], who defines the lap time benefit from fuel consumption as a linear function of the remaining fuel mass. The formulation considers the fuel consumed per lap, the total fuel mass at the start of the race and the sensitivity of lap time to vehicle mass. Equation 5.4 [3] expresses the fuel lap time model contribution as:

$$t_{fuel}(l) = \left(m_{fuel,tot} - m_{fuel,consumed}(l)\right) \cdot s_{lap,mass}$$

Equation 5.4

Where:

- $t_{fuel}(l)$ is the lap time effect from fuel consumption at lap l,
- $m_{fuel,tot}$ is the total fuel mass at the race start (assumed to be 110 kg),
- $m_{fuel,consumed}(l) = B_{fuel,perLap} \cdot l$ is the cumulative fuel consumed up to lap l (the fuel consumption rate term $B_{fuel,perLap}$ is defined in Equation 5.5 [3]),
- $s_{lap,mass}$ is the mass sensitivity coefficient, which quantifies how much lap time decreases per kilogram of fuel mass lost.

This formulation is computationally efficient and physically interpretable. It captures the monotonic trend of performance improvement over the race and can be generalized to include refueling logic in other racing formats. In Formula One, where refueling is banned, this results in a strictly decreasing lap time contribution across the race duration.

For implementation purposes, a constant fuel consumption rate is assumed for each race, computed as:

$$B_{fuel,perLap} = \frac{m_{fuel,tot}}{N_{laps}}$$

Equation 5.5

Where N_{laps} is the number of scheduled race laps for a given circuit. This assumption, while idealized, provides a reasonable approximation due to the limited lap-to-lap variation in fuel burn in dry and uninterrupted conditions, although, as it will be better explained in the following chapters, this value will be adjusted according to the eventuality of a FCY phase.

The mass sensitivity coefficient $s_{lap,mass}$ is drawn from literature estimates and engineering judgment [20]. In the current implementation, as McLaren's paper states [20], it is set to 0.03 s/kg, meaning that each kilogram of fuel lost results in a 0.03 second lap time improvement. This value may be adjusted for future calibration using telemetry or regression analysis.

While Sulsters [5] proposes a data-driven regression approach that relates lap time directly to the remaining fuel percentage, the current version of the simulator opts for a physically grounded model that assumes known fuel usage and a predefined sensitivity coefficient. Sulsters' method requires a large volume of clean race lap data and per-driver regression fitting, which, though precise, is

computationally intensive and introduces additional uncertainty in estimating parameters like race start fuel mass or lap-specific disturbances.

Figure 5.2 illustrates a sample output of the fuel model, showing the lap-by-lap fuel time contribution for a full-length race (i.e. 57 laps). The time loss decreases linearly as the fuel mass decreases, resulting in a difference of approximately 3 seconds from the first to the last lap purely due to fuel effect.

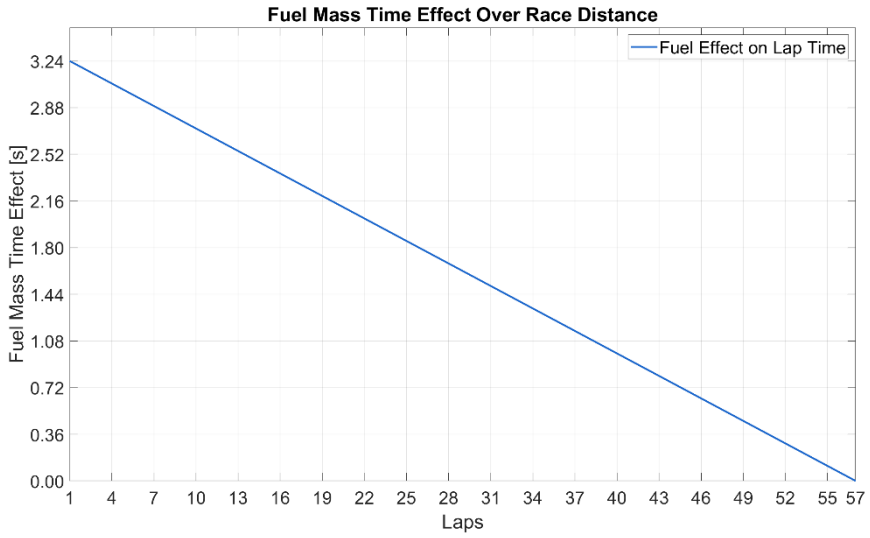

Figure 5.2: Fuel mass time effect over the race distance

In the simulator structure, this fuel effect is added to the base lap time as described by Equation 5.1 and accumulated lap-by-lap for each driver. Secondary effects such as fuel-induced changes to tire degradation or cornering dynamics are neglected in the current implementation for simplicity and because their impact is relatively minor compared to the primary linear mass effect.

5.3 Tire Degradation Model

Tire degradation is a crucial element in modelling race performance, as it reflects the gradual reduction in grip and overall performance that tires undergo during a stint. As tires wear, they provide less traction and stability, leading to slower lap times and diminished handling capabilities for the car. As a tire's age increases, its ability to maintain optimal contact with the asphalt deteriorates, resulting in incremental lap time losses. Capturing this phenomenon accurately is essential to simulate race dynamics, particularly in the context of pit stop planning, compound selection and stint length optimization.

Multiple modelling approaches exist in literature to approximate this degradation trend, differing in both mathematical formulation and empirical assumptions. A widely adopted strategy is to define tire degradation as a function of the fitted compound type and the number of laps completed on that compound, referred as tire age.

Heilmeier [3] proposes two primary formulations:

- Logarithmic model, expressed in Equation 5.6 [3], which better represents the early non-linear performance drop of softer compounds, preferred for high degradation scenarios,
- Linear model, expressed in Equation 5.7 [3], reserved for contexts with sparse data or low degradation sensitivity.

These models are defined as follows:

$$t_{tire,log}(a_{tire}, c_{tire}) = log\big(a_{tire} \cdot k_{1,log}(c_{tire}) + 1\big) \cdot k_{2,log}(c_{tire}) + k_3(c_{tire})$$

Equation 5.6

$$t_{tire,lin}(a_{tire}, c_{tire}) = a_{tire} \cdot k_{2,lin}(c_{tire}) + k_3(c_{tire})$$

Equation 5.7

Where:

- c_{tire} represents the fitted compound,
- a_{tire} represents the tire age, expressed as laps,
- k_1, k_2, k_3 are compound-dependent coefficients determined empirically.

F1Metrics [14] analysis similarly models degradation with a quadratic function, assuming a degradation rate that accelerates over time as the tires wear out. A key highlight from this study is the concept of the compound crossover point, highlighted in Figure 5.3. This point marks the stage in a stint when a heavily worn softer tire can become slower than a newer, harder compound. Recognizing this moment is crucial for race strategy, as it enables a driver who switches early to fresh harder tires to gain an advantage. By pitting sooner, this driver may achieve faster lap times and potentially overtake a competitor who remains on the deteriorating softer tires, a strategy known as undercut.

It should be noted that the labeling used in Figure 5.3 reflects the older conventional terminology from a period when regulations permitted only two tire compounds to be used during the race [21]. In this context, the "option" tire refers to the softer compound, while the "prime" tire indicates the harder compound. Consequently, the figure illustrates the comparison limited to two consecutive tire types, highlighting how performance evolves between a softer and a harder compound over the course of a stint.

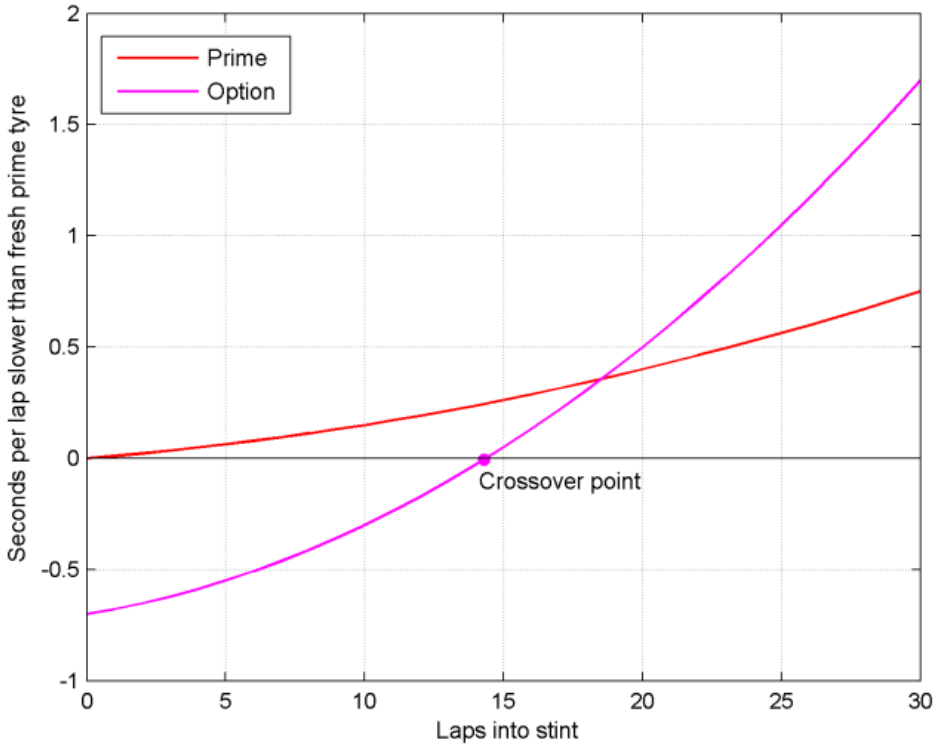

Figure 5.3: Crossover point: used softer (option) tires start to be slower than fresh harder tires (prime) [14]

The model adopted in this thesis builds upon both the formulations of F1Metrics [14] and Sulsters [5], where tire degradation is captured through a quadratic non-negative regression model. This approach enables a flexible representation of lap time increases due to tire wear, with the only model constraint being that the second-order coefficient remains non-negative, ensuring that the time loss profile is either linear or increases with tire age, reflecting realistic tire performance trends over a stint.

The resulting degradation function expressed in Equation 5.8 [5] is defined in as:

$$t_{tire,quad}(a_{tire}, c_{tire}) = a_{tire}^2 \cdot k_{1,quad} + a_{tire} \cdot k_{2,quad}(c_{tire}) + k_3(c_{tire})$$

Equation 5.8

A notable distinction of this work lies in the treatment of tire compound classification. Unlike prior studies that refer to compounds using generic nomenclature of SOFT, MEDIUM and HARD, this thesis adopts the official C-Pirelli nomenclature (from C0 to C5). These labels represent the actual compound spectrum provided by Pirelli over the Formula One season, with C0 being the hardest and C5 the softest. For each Grand Prix, three of these six available dry compounds are selected and designated as SOFT, MEDIUM and HARD for public communication and broadcasting purposes. By referencing the C-compound directly in the degradation model (i.e. using c_{tire} {C0, C1, C2, C3, C4, C5}), the simulation retains fidelity to the true physical properties of the tires, rather than relying on the race-specific labelling conventions.

In addition, this formulation allows the model to adapt to various degradation patterns based on the observed data for each compound and circuit. Rather than estimating coefficients on a driver-specific basis, an approach used in Heilmeier [3] and Sulsters [5] studies, this thesis generalizes the degradation model at the circuit-compound level. This decision is motivated by the intent to isolate the structural impact of the circuit and compound while neutralizing driver-related variability. As a result, all drivers at a given event are modelled with the same degradation profile, ensuring consistency and comparability across simulation runs.

5.3.1 Implementation Strategy

To estimate tire-related performance losses during a race stint, a systematic methodology has been implemented in MATLAB. This procedure transforms raw lap data into corrected, circuit-specific degradation profiles through a series of structured preprocessing and fitting steps. The objective is to generate compound-specific degradation parameters for each circuit, applicable across all drivers, thereby decoupling tire performance modelling from driver variability.

The procedure begins by standardizing the lap time format. Raw data, initially stored as string-formatted durations, is converted into numeric values expressed in seconds. This conversion is executed individually for each driver and for every compound used during the race. Such normalization of data allows for consistent and reliable processing in all subsequent computational steps.

To isolate tire degradation from other overlapping race effects, the influence of decreasing fuel mass is explicitly removed. This is achieved by subtracting, from each lap time, the precomputed time loss associated with fuel consumption, based on the previously modelled fuel mass time effect. The outcome of this subtraction is a corrected lap time that reflects only tire-related performance loss while excluding the decreasing time loss derived by fuel burn. This ensures a clean baseline for studying degradation dynamics.

Corrected lap times are then grouped into valid stints. A valid stint is defined as a sequence of at least six consecutive laps that share the same tire compound and belong to the same race file (i.e. uninterrupted by pit stops). Each stint is characterized by the progression of lap times, lap numbers and a unique stint identifier. These data are aggregated in a hierarchical structure sorted by circuit and compound, enabling later statistical analysis and visualization.

Figure 5.4 and Figure 5.5 show an example of this data structure. For a set of selected drivers, every recorded stint at the Italian Grand Prix is plotted, visually highlighting the degradation behavior for each compound and confirming the validity of stint segmentation.

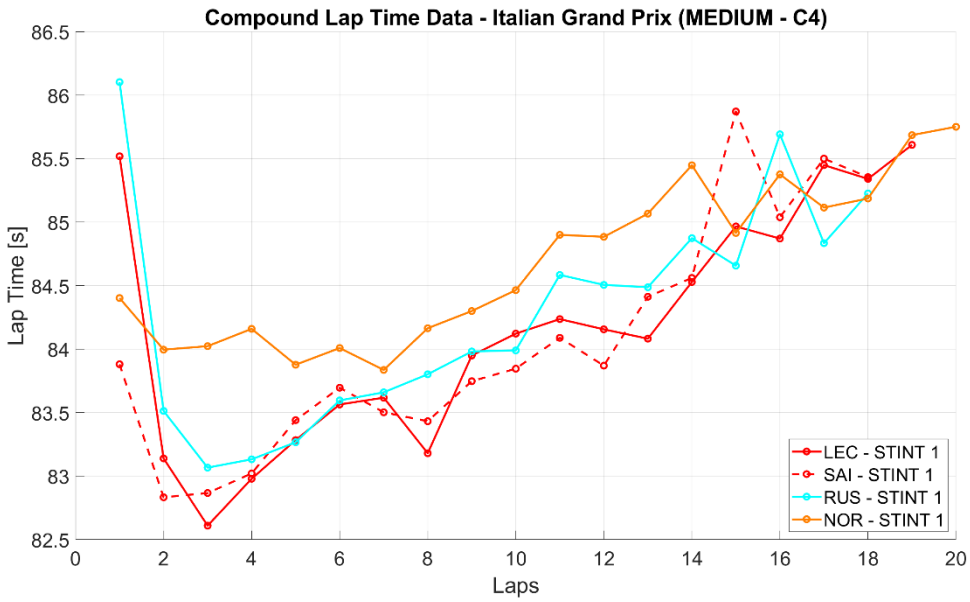

Figure 5.4: MEDIUM (C4) compound lap time data for the 2023 Italian Grand Prix

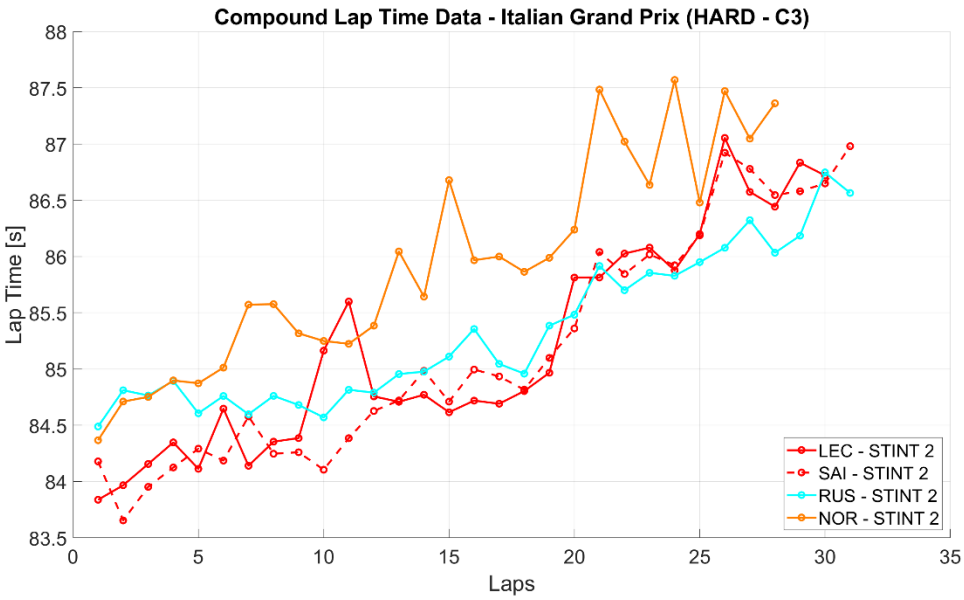

Figure 5.5: HARD (C3) compound lap time data for the 2023 Italian Grand Prix

For each recorded stint, a non-negative quadratic fitting model of the form expressed in Equation 5.8 is applied. The fitting is constrained to ensure that the second-order term is always non-negative, avoiding the modelling of unrealistic concave behaviors. Each stint is fit independently and the resulting quadratic term $k_{1,quad}$ and the linear term $k_{2,quad}$ are stored alongside stint length metadata. This approach allows to capture either linear or upward-bending trends, depending on the compound and circuit characteristics, providing a flexible and physically consistent model of tire performance.

Figure 5.6 and Figure 5.7 visually compare the original and fitted lap times for each stint across selected drivers. This comparison confirms that the chosen fitting model captures the underlying degradation pattern while adhering to the physical constraint of non-negative quadratic term. Furthermore, it is particularly evident in Figure 5.6 that the first lap time is significantly higher than subsequent laps. This phenomenon reflects several contributing factors unique to the opening lap, such as the initial acceleration phase from a standstill position in the drivers' respective grid slot, heightened risk of incidents due to tightly packed cars and limited space and increased overtaking activity as drivers attempt to gain positions while running in close proximity and in turbulent air. These factors collectively result in the first lap standing apart in its duration, emphasizing why it is often treated separately in both analysis and simulation modelling.

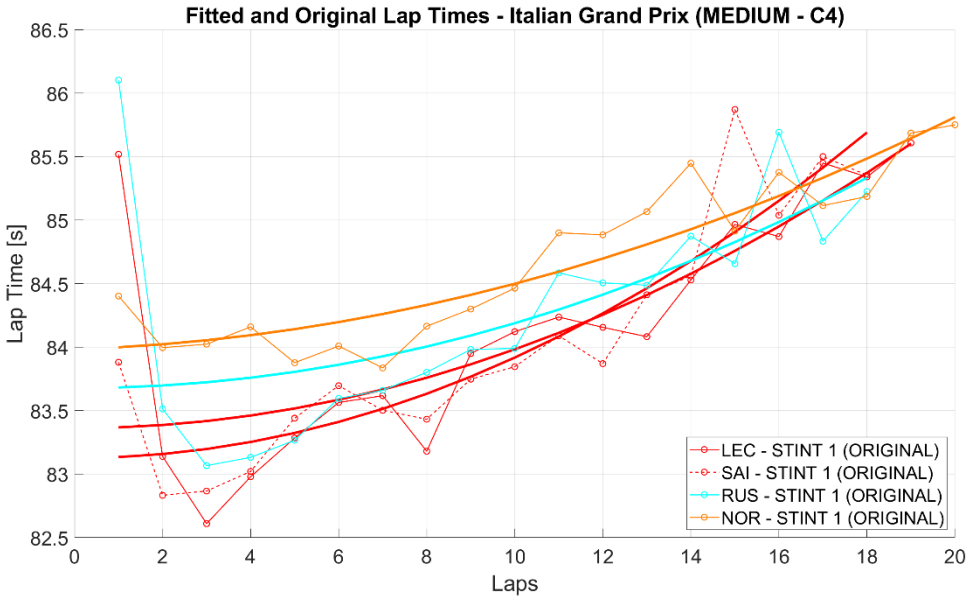

Figure 5.6: Non-negative quadratic fitting operation compared with original lap times regarding MEDIUM (C4) compound for the 2023 Italian Grand Prix

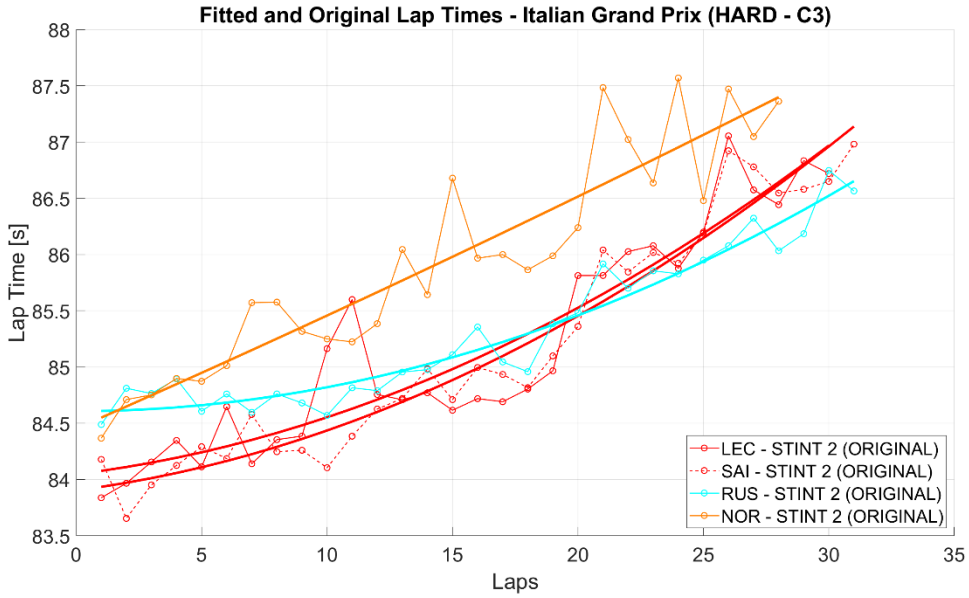

Figure 5.7: Non-negative quadratic fitting operation compared with original lap times regarding HARD (C4) compound for the 2023 Italian Grand Prix

After individual stint fitting, the most representative stint for each driver and compound at a specific circuit is identified. This selection is based on two prioritization criteria:

- The magnitude of the quadratic coefficient, indicative of stronger degradation patterns,
- The length of the stint, for which longer stints provide more robust fits.

Once these criteria are satisfied for all drivers, the corresponding quadratic and linear coefficients are averaged across drivers to yield the final degradation parameters for that circuit-compound pair. This ensures that tire degradation behavior is circuit-specific rather than driver-specific, improving generalizability and simulation robustness. Table 10.1 reports each quadratic and linear parameter for each circuit of the 2023 season and each used compound.

In the final stage of the modelling procedure, a visualization step is introduced to assess the distribution of degradation parameters across different drivers for a selected circuit. This process aims to verify the variability and consistency of the non-negative quadratic fitting outputs and to facilitate validation of the average parameter values derived in the previous step. Both quadratic and linear parameters are visualized using normalized histograms overlaid with kernel density estimates (KDEs), one plot per parameter type:

- Quadratic Parameter Distribution: Reflects the intensity of non-linear degradation across stints. A higher value indicates a more pronounced acceleration in lap time loss over the stint duration,
- Linear Parameter Distribution: Captures the constant rate of degradation independent of non-linear effects.

For each compound, the mean of the parameter is also displayed, offering a reference point for interpreting the KDE envelope and histogram spread. These visualizations serve both as diagnostic tools and as validation for the selection of average values to be used in the race simulation model. Figure 5.8 illustrates the distribution of quadratic and linear degradation parameters across drivers for the HARD (C3) and MEDIUM (C4) compounds at the Italian Grand Prix. The left subplot shows that the C4 compound exhibits a noticeably higher mean quadratic parameter than C3, suggesting a more pronounced late-stint performance drop due to non-linear tire wear. Conversely, the right subplot reveals that the C3 compound has a higher linear degradation parameter, implying a more uniform and steady degradation pattern throughout the stint. These observations support the use of compound-specific degradation models and justify the use of the average values as representative parameters in the simulator.

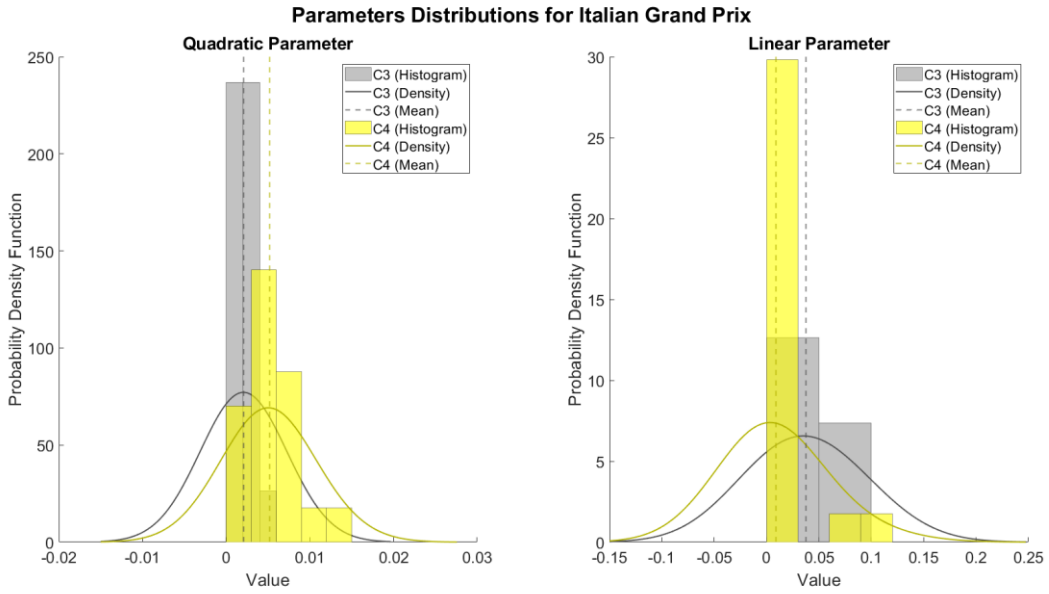

Figure 5.8: Quadratic and linear parameters distributions regarding MEDIUM (C4) and HARD (C3) compounds for the 2023 Italian Grand Prix

5.4 Lap Time Variability Model

As part of the broader effort to construct a comprehensive and data-driven race simulation framework, this section introduces the modelling of lap time variability. While previous chapters have addressed deterministic contributors such as fuel mass reduction (Chapter 5.2) and tire degradation (Chapter 5.3), these alone cannot fully account for the irregularities observed in real-world lap time data. In reality, even under stable conditions free from pit stops, traffic or SC interventions, drivers seldom replicate the exact same lap time twice. These subtle fluctuations stem from a combination of factors, including steering precision, throttle modulation, local grip variability, external disturbances like wind and psychological pressure.

Capturing such random but bounded effects is crucial to simulate racing behavior with authenticity. Ignoring them would lead to overly deterministic and unrealistic race outputs, where lap times evolve too predictably. Converse, incorporating a well-calibrated measure of lap time variability allows the simulator to better represent overtaking chances, pit stop strategy windows and intra-driver performance stability.

The method adopted in this thesis builds directly upon the framework proposed by Heilmeier [4] in his study of probabilistic race simulations using Monte Carlo methods. There, lap time variability is extracted subtracting a fitted trend line from real lap time sequences, isolating the random component from structured race dynamics. This idea is portrayed in a similar way in the statistical modeling proposed by Sulsters [5], who embeds a normally distributed residual term within his full lap time formulation, accounting for the same unpredictability after removing systematic effects like tire wear and fuel consumption.

In this thesis, the core of the method involves segmenting race data into clean, uninterrupted lap sequences and fitting a quadratic polynomial to each. This polynomial approximates the expected la time trend within a series of consecutive laps using the lap number as the independent variable. The fitting curve is expressed in Equation 5.9 [4] as:

$$t_{lap,poly}(p) = k_2 \cdot p^2 + k_1 \cdot p + k_0$$

Equation 5.9

Where:

- p is the continuous sequence of laps recorded without any interruption or anomalies, respecting the 107% rule [18],
- $t_{lap,poly}(p)$ represents the quadratic fitted lap time and the expected lap time evolution,
- k_2, k_1, k_0 are the coefficients derived from the polynomial fitting.

Each real time $t_{lap}(p)$ subjected to the fitting procedure is then compared to the corresponding fitted value and the deviation is calculated as described in Equation 5.10 [4]:

$$t_{lap,dev}(p) = t_{lap}(p) - t_{lap,poly}(p)$$

Equation 5.10

The resulting deviation $t_{lap,dev}(p)$ quantifies the random, non-repeatable portion of the lap time. By design, it excludes effects which were already addressed in previous chapters through deterministic models and captures only the noise component intrinsic to race performance.

5.4.1 Implementation Strategy

The implementation of this model begins by selecting only clean laps recorded under dry-weather conditions. These laps are extracted using the FastF1 telemetry toolkit, previously introduced in Chapter 4, and filtered to comply with the 107% rule [18]. Applying the rule, laps influenced by traffic, pit stops or neutralization periods such as yellow flags or SC phases are excluded to isolate pure race pace.

Once the data is curated, the lap sequences are segmented based on lap number continuity. Each race stint is scanned for consecutively numbered laps and whenever a discontinuity is detected the current segment is closed and a new one is initiated. This segmentation ensures that the fitting process is applied only to uninterrupted race segments where a continuous performance trend can be reasonably assumed.

For each valid segment, a second-order polynomial, described by Equation 5.9, is fitted to the lap time data, modelling the typical pace evolution over the stint. Subtracting the fitted curve from the actual lap times, as described by Equation 5.10, yields lap-specific deviations, which represent the stochastic component of driver performance.

This process is applied across every dry circuit of the 2023 Formula 1 season and for every driver in the dataset. Once the deviations are calculated, their average values are computed to derive a single driver-specific variability index. This index quantifies the typical amplitude of a driver's lap time oscillations around the expected trend: small values suggest consistent performance, while large values indicate more pronounced variability. A complete summary of these driver-specific mean variability values is provided in Table 10.2.

To operate this variability within the race simulation, a final step involves the construction of a normally distributed random component for each driver. During the simulation, this is implemented by drawing random samples from a zero-mean Gaussian distribution [5] whose standard deviation matches the driver-specific variability index. Mathematically, this is expressed in Equation 5.11 [5] as:

$$\varepsilon_{lap} \sim \mathcal{N}\left(0, \sigma^2_{driver}\right)$$

Equation 5.11

Where:

- σ_{driver} is the computed variability for that driver,
- ε_{lap} is the random term to be added to each driver-specific lap time.

This generated sequence provides the stochastic noise to be incorporated into each lap of the simulation, allowing the model to reflect real-world unpredictability. This ensures that the same driver and strategy might yield slightly different outcomes across multiple simulation runs, an essential feature when adopting a Monte Carlo approach to race prediction and strategy evaluation.

To illustrate this methodology, Figure 5.9 presents a representative case for driver Carlos Sainz at the 2023 Italian Grand Prix. The upper subplot shows the consecutive raw lap times fitted with a second-order polynomial, the lower subplot displays the calculated deviations between the actual and the fitted lap times. These deviations are symmetrically distributed around zero, with no systematic drift, indicating a balanced fluctuation pattern, consistent with the assumption of normally distributed noise.

It is important to clarify a potential source of confusion that may arise when comparing the fitted lap time trends from the variability analysis with the deterministic model of tire degradation and fuel consumption presented earlier in this thesis. The tire degradation model is constructed to impose a non-negative penalty on lap times, reflecting the physical principle that tire wear reduces grip and

progressively slows the car. Thus, it produces a monotonically increasing or, in the limit, constant lap time penalty across a stint. However, in the polynomial fits derived for lap time variability (as illustrated in Figure 5.9), it is sometimes observed that lap times initially decrease before rising again later in the stint. This initial decrease is not contradictory to the degradation model because it can be explained by the simultaneous effect of fuel mass reduction and maximum initial exploitation of the new fitted tire. As race progresses, the car becomes lighter, leading to faster lap times which can temporarily outweigh the negative impact of tire wear. Therefore, the fitted polynomial captures the combined effect of these overlapping phenomena.

Nevertheless, the purpose of the variability fitting in this thesis is not to replace or contradict the deterministic components already modelled for fuel and tire effect. Instead, it serves solely to isolate the stochastics, lap-to-lap fluctuations around the expected pace once the broader trends have been accounted for. While the deterministic models explain systematic effects like fuel mass decrease and tire degradation, the polynomial fit in the variability model provides a local approximation of the net trend in each stint, accommodating any transient balance between fuel savings and tire penalties. The residuals extracted from this fit represent purely random deviations, which are crucial for simulating realistic racing scenarios. Thus, introducing a polynomial fitting step in the variability modelling does not imply the creation of a new deterministic model, but rather a methodological tool for quantifying and isolating stochastic behavior.

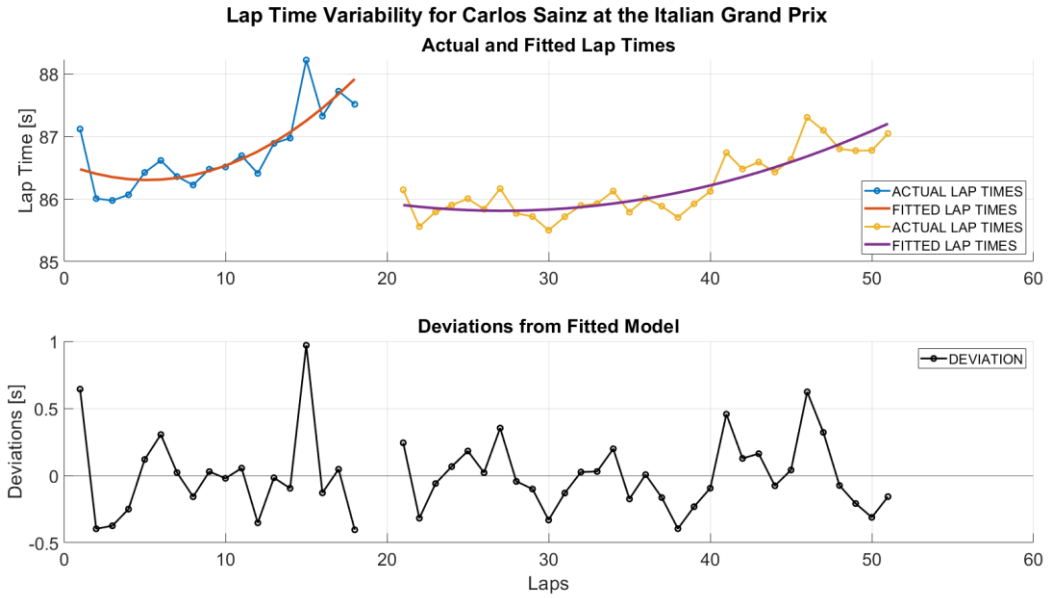

Figure 5.9: Polynomial fit and lap time deviations for Carlos Sainz at the Italian Grand Prix

5.5 Starting Performance Model

In a complete and realistic race simulation framework, the treatment of the race start plays a critical role in determining the evolution of the early race order. While previous modules of the simulator focus on fuel, tire degradation and lap time variability, starting performance represents a sudden and discrete effect that occurs exclusively at the onset of the race, influencing the outcome of the first lap alone. However, from the driver's perspective, the outcome of the start is crucial for the development of the most suitable race strategy.

The model adopted in this thesis follows the deterministic-probabilistic dual framework proposed by Heilmeier [4], which provides a physically grounded and statistically nuanced method for modelling the start phase. This model is preferred over discrete or empirical sampling approaches, like the one proposed by Sulsters [5], as it ensures interpretability and continuity across the grid, avoiding the sparsity and volatility associated with raw positional gain distributions.

The deterministic portion of the model, called "average starter" by Heilmeier [4], captures the baseline time a driver needs to travel from their grid slot to the start/finish line, assuming uniform conditions and average driver skill. This is modelled using Equation 5.12 [4], a square root formulation derived from constant-acceleration kinematics:

$$t_{start,deterministic} = \sqrt{\frac{2(p_g - p_s) \cdot s_{grid}}{a_{avg}}} + t_r$$

Equation 5.12

Where:

- p_g is the driver's starting grid position,
- p_s is the fractional offset of the pole position from the start line (typically set to 0.8 [4]),
- s_{grid} is the standard spacing between grid slots (set to 8 m),
- a_{avg} is the average longitudinal acceleration during launch, estimated by least-square fitting to historical Formula One start data (set to 11.2 m/s^2 [4]),
- t_r is a fixed human reaction time (assumed to be 0.2 s [4]).

This deterministic time, visually represented in Figure 5.10, effectively quantifies the physical component of launch performance, assuming each car launches cleanly from its starting position with uniform grip and optimal execution. It thus forms the baseline over which driver-specific deviations can be layered.

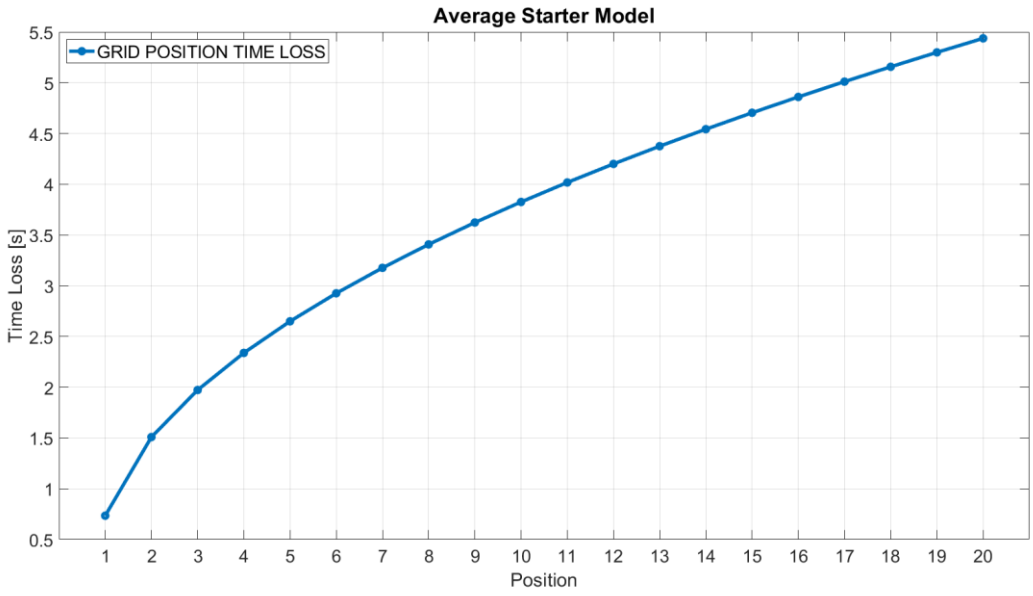

Figure 5.10: Average starter deterministic model taking into account ideal starting grid slot-specific launch conditions [4]

Real-world observations, however, show that even under similar conditions, some drivers consistently outperform or underperform the deterministic baseline. To reflect this, a stochastic performance modifier is introduced, as shown in Equation 5.13 [4]:

$$t_{start,performance} \sim \mathcal{N}\left(\mu_{start,performance}, \sigma^2_{start,performance}\right)$$

Equation 5.13

Each driver is assigned a unique normal distribution centered at a mean value $\mu_{start,performance}$ with a standard deviation $\sigma_{start,performance}$, both empirically derived from video analysis and telemetry evaluations conducted by Heilmeier [4]. These parameters capture how each driver deviates from the "average starter" profile and account for skill, reaction consistency and car-clutch interaction quality at launch. Driver-specific distribution values are reported in Table 10.2, as they are taken from Heilmeier's work and adjusted to the 2023 drivers.

The overall time cost from the race start until crossing the start/finish line is therefore modelled as the sum of the deterministic and probabilistic components, highlighted in Equation 5.14 [4]:

$$t_{start} = t_{start,deterministic} + t_{start,performance}$$

Equation 5.14

The total start time is only applied once, on the first lap of the race, and directly influences the cumulative race time from which all subsequent lap positions are derived. Drivers with consistently negative values of $t_{start,performance}$ are more likely to gain positions off the line, while those with positive deviations will tend to fall back.

Compared to position-based models such as the smoothed empirical distribution proposed by Sulsters [5], which draw from discrete histograms of positional gain/loss at the start and finally creates a discrete shift in position, this time-based approach offers several key advantages:

- It integrates seamlessly into a race-time simulation framework without requiring direct manipulation of positions,
- It allows fine-tuned probabilistic variation within a continuous domain, avoiding the rigidity of categorical jumps,

- It maintains physical realism by anchoring every performance deviation to a measurable outcome: time to cross the start/finish line,
- It provides a fairer representation for drivers who frequently start from pole position [4], as these drivers have limited opportunity to gain positions at the start. Unlike categorical models, which can introduce bias by underestimating variability for front-runners, the adopted time-based modelling approach captures subtle differences in start performance even for those starting at the front of the grid.

Ultimately, the incorporation of this start model enhances the simulator's fidelity during the critical early moment of the race, ensuring that launch dynamics are neither oversimplified nor excessively randomized.

5.6 Pit Stop Performance Model

In the context of a comprehensive race simulation framework, accurately modelling pit stop performance is essential for capturing the strategic and stochastic elements that influence race outcomes. Pit stops introduce discrete time losses that can significantly affect a driver's position and overall race time. This section delineates the components of pit stop time and presents a modelling approach that integrates both deterministic and probabilistic elements, drawing upon established methodologies in the literature.

5.6.1 Components of Pit Stop Time

As it can be stated by referencing Heilmeier's work [4], the total time loss associated with a pit stop can be decomposed into two primary components, noted in Equation 5.15:

$$t_{pit} = t_{inlap} + t_{outlap}$$

Equation 5.15

Where:

- t_{pit} is the time loss associated with a pit stop,
- t_{inlap} is the time lost during the entry into the pit lane, which qualitatively captures the additional time required for the driver to maneuver towards the pit entry line, decelerate from racing speed and activate the pit limiter system (it is typically modelled as a small time loss between 1.5 s – 2 s),
- t_{outlap} is the time lost starting from the moment in which the pit limiter is engaged and encompasses several sub-components, highlighted in Equation 5.16 [4]:

$$t_{outlap} = t_{pitdrive} + t_{crew} + t_{penalty}$$

Equation 5.16

Where:

  - $t_{pitdrive}$ is the time required to traverse the pit lane at the speed limit imposed by circuit-specific race regulations,
  - t_{crew} is the duration of the stationary period during which the pit crew performs operations on the car, such as tire changes and wing adjustments,
  - $t_{penalty}$ is the additional time incurred due to penalties, such as stop-and-go or time penalties imposed by race stewards.

The $t_{pitdrive}$ value has been determined by analyzing driver telemetry during pit stop sequences, as it was described in Chapter 4. Specifically, the entry and exit points of the pit limiter were identified by detecting when the car's speed first dropped (and later rose) around the prescribed pit lane limit, accounting for a tolerance margin of ±2%. From these time stamps, the duration of pit lane traversal was calculated. The resulting $t_{pitdrive}$ values, alongside the corresponding pit lane speed limits for each circuit, are listed in Chapter 6, Table 10.5.

The $t_{penalty}$ component is event-driven and incorporated based on race incidents and steward decisions. For the sake of simplicity and given the complexity of accurately modelling such event-driven occurrences, this factor is excluded from all current simulation scenarios.

The t_{crew} component exhibits variability influence by team efficiency and operational precision necessitating a probabilistic modelling approach.

5.6.2 Pit Crew Efficiency Modelling

The variability in pit crew performance, encapsulated in the t_{crew} component, is modelled using the Fisk distribution also referred to as the log-logistic distribution due to its demonstrated effectiveness in representing the empirical distribution of pit stop durations observed in Formula One. Unlike symmetric distributions such as the Gaussian, which fail to capture the distinct asymmetry of pit stop times, the Fisk distribution offers a flexible, right-skewed shape that mirrors real-world behavior, where the majority of pit stops are tightly grouped around a minimal mode, but occasional operational inefficiencies result in a long tail of delayed stops. This modelling choice is directly inspired by the methodology proposed by Philips [14] and Heilmeier [4] who, after evaluating pit stop data from multiple Formula One seasons, noted that top-performing teams exhibited pit stop durations consistently close to the race minimum, while others showed more dispersed outcomes with a higher incidence of longer stops. In Heilmeier's work [4], the Fisk distribution was identified as the best-fitting statistical model to account for this asymmetric dispersion and its parameters were fitted individually for each team using race data from 2014-2019 seasons, adjusted to be representative of the 2023 team list.

The Fisk distribution, often referred to as the log-logistic distribution, is a continuous probability distribution suitable for modelling non-negative, right skewed data. It is particularly effective when values cluster near a central tendency but exhibit a heavy right tail, a common characteristic in real-world phenomena such as income distribution [22] and, in this context, Formula One pit stops duration.

Given three positive parameters (shape, scale and location), the Probability Density Function (PDF) of the Fisk distribution is given by Equation 5.17 [22]:

$$Fisk(x; shape, loc, scale) = \frac{\frac{shape}{scale} \cdot \left(\frac{x - loc}{s}\right)^{scale-1}}{\left[1 + \left(\frac{x - loc}{scale}\right)^{shape}\right]^2}, \qquad x > loc$$

Equation 5.17

Where:

- x is the Fisk value used for generating random samples (defined in detail by Equation 5.18),
- $shape$ value controls the tail behavior, in particular a lower shape value increases the probability of very large values (long tail),
- $scale$ value stretches or compresses the distribution horizontally,
- loc value, which is null in this thesis work to simplify the modelling, shifts the distribution along the x-axis.

The Fisk value, x, is defined in Equation 5.18 [22] as:

$$x = loc + scale \cdot \left(\frac{u}{1 - u}\right)^{(1/shape)}, \qquad u \sim \mathcal{U}(0,1)$$

Equation 5.18

The resulting random variable x exploits the heavy right tail typical of pit stop distributions, where most values concentrate around a centrale mode, but rare, longer delays due to inconveniences during the pit stop remain probabilistically possible.

To ensure consistency with historical performance, the generated random values from the Fisk distribution are aligned with the team's observed average pit stop time $\bar{t}_{crew}$ [23]. This is achieved by shifting the distribution accordingly: the average time is added directly to the sampled variable, producing a distribution centered around each team's typical execution level. Equation 5.19 shows the adjusted pit stop crew efficiency time loss re-centered on the team-specific target mean:

$$t_{crew} = \bar{t}_{crew} + x$$

Equation 5.19

Moreover, since no pit stop can realistically occur below a physical feasible threshold, which has been established at 1.80 s [23] during the 2023 Formula One season, values are generated to remain strictly above this minimum. This approach ensures statistical representativeness while maintaining physical plausibility and aligns with simulation methodologies described by Sulsters [5] and Heilmeier [4]. Detailed information on each team's fitted parameters and their statistical justification can be found in Chapter 10, Table 10.6, which adapts the data found by Heilmeier [4] to the 2023 Formula One grid.

To visualize the impact of pit crew efficiency on pit stop durations, Figure 5.11 considers the following comparative analysis:

- Top team (Ferrari), whose distribution of t_{crew} is sharply peaked, indicating high consistency with minimal variability. Most pit stops are completed near the mode, with few instances of extended durations,
- Midfield team (Haas), whose distribution is flatter and more spread out, reflecting greater inconsistency. This results in increased probability of both quicker and significantly longer pit stops than the mode.

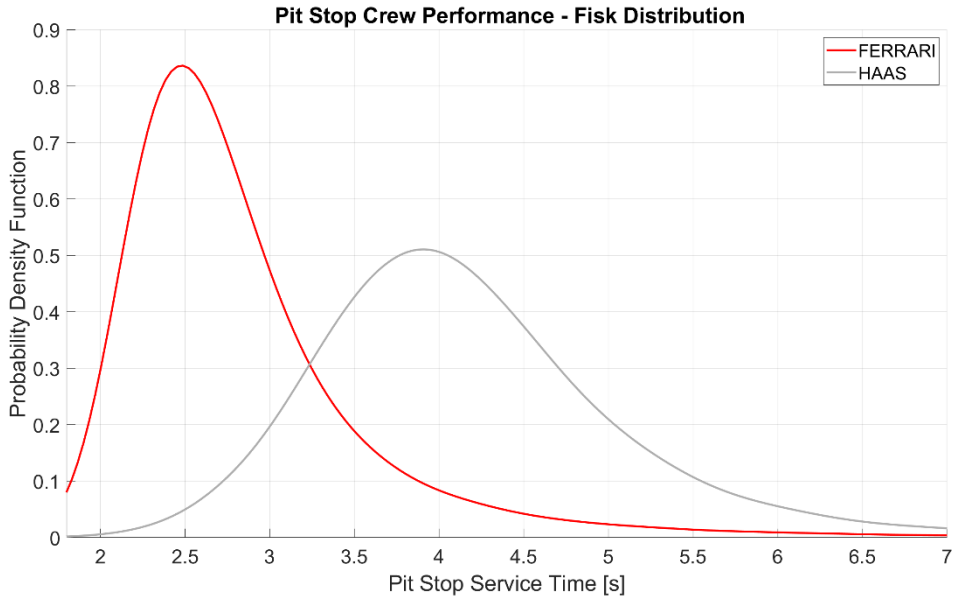

Figure 5.11: Comparative analysis of the pit crew efficiency on pit stop durations

This comparison underscores the importance of pit crew performance in race strategy and outcomes. Efficient and consistent pit stops can contribute to maintaining or improving track position, while variability can introduce risks that may compromise race performance.

5.7 Did Not Finish (DNF) Probability Model

In the stochastic modelling of Formula One races, one of the most impactful sources of unpredictability is the occurrence of race-ending events, commonly referred to as Did Not Finishes (DNFs). DNFs arise due to either accidents or mechanical failures and their influence is twofold: they directly remove competitors from the classification and they indirectly shape race dynamics by triggering FCY phases such as SCs and VSCs. This chapter presents the modelling approach adopted to simulate such stochastic events, combining statistical inference with domain-specific assumptions. This marks the first truly probabilistic event in the simulator pipeline, laying the foundation for dynamic and reactive race simulations.

The methodology for computing DNF probabilities follows the Bayesian framework introduced in Sulsters' thesis [5] and in Heilmeier's work [4]. This method enables the integration of prior knowledge with observed racing data, addressing the challenge of limited observations for certain drivers while ensuring a consistent probabilistic foundation.

Race retirements are not rare outliers: analysis of race data collected from StatsF1.com over the 2019-2023 seasons reveals that in dry conditions, on average, between two and four drivers per Grand Prix fail to reach the checkered flag [17]. This corresponds to approximately 10% to 20% of the starting grid per race. These events significantly distort predicted finishing positions, invalidate strategy assumptions and frequently alter the final classification by enabling opportunistic gains for trailing drivers. As noted by Catapult Sports [2], predictive analytics in Formula One must incorporate reliability and failure modelling to anticipate race-altering scenarios. Therefore, excluding DNFs from simulation would compromise the realism and predictive value of the model.

5.7.1 Accidents and Failures Modelling

To accurately simulate DNFs in Formula One races. Retirements are divided into two main categories:

- Accidents, which are primarily driver-dependent and often occur due to collisions or misjudged maneuvers,
- Mechanical failures, which are largely team-dependent, stemming from technical malfunctions, engine reliability issues or system failures.

This categorization enables a hybrid estimation strategy.

Accident probability $P_{accident}$ is estimated through Bayesian inference applied to individual race histories from 2019 to 2023. This method addresses the challenge of limited observations per driver by incorporating prior knowledge and ensuring stable probability estimates. The dataset manually curated from StatsF1 [17] contains statistics including the number of started races and DNFs recorded for each driver, enabling the posterior estimation of DNF likelihoods through Beta-distributed priors.

Mechanical failure probability $P_{failure}$ is derived from historical team-specific failure frequencies documented in Heilmeier's study covering the seasons from 2015 to 2019 [4]. Since team reliability evolves over time, these rates are re-mapped to reflect the most recent team composition in the 2023 grid. This ensures that failure probabilities remain aligned with contemporary technical performance.

A key refinement in this model is the necessary adjustment applied to accident probabilities during the opening lap. Empirical data and academic studies [4, 5] indicate a heightened risk of incidents at the start of the race due to dense driver formations and aggressive position changes. To reflect this, an enhanced probability multiplier is applied, as shown in Equation 5.20 [4, 5]:

$$P_{accident,lap1} = 10 \cdot P_{accident}$$

Equation 5.20

This adjustment captures the increased likelihood of first lap accidents.

5.7.2 Bayesian Inference Framework for DNF Estimation

Retirements are modelled by estimating the DNF probability for each driver based on the finishing status recorded in historical race data. A naive method would assign each driver a DNF probability equal to the fraction of races they failed to finish. However, such estimates are unreliable for small sample sizes and result in unrealistic zero probabilities for drivers with no retirements [5]. To overcome this, a Bayesian inference approach is employed, as detailed in Sulsters [5] and Heilmeier [4].

Bayesian inference is a statistical method that updates the probability of a hypothesis as more evidence or information becomes available. It combines prior belief about a parameter with new data to yield a posterior belief, thereby systematically refining uncertainty estimates. This approach is particularly useful in scenarios with limited or noisy data, where it enables the incorporation of reasonable prior assumptions. It typically involves conjugate priors to simplify computations, such as the Beta distribution when dealing with Bernoulli or Binomial processes [24].

As is portrayed in Sulsters' thesis [5], each race start is modelled as an independent Bernoulli trial with a binary outcome: finish or DNF. Let y_j represent the number of non-finishes for driver j out of n_j races. The likelihood model follows a binomial distribution, reported in Equation 5.21 [5]:

$$y_j \sim Bin(n_j, \theta_j)$$

Equation 5.21

Where θ_j denotes the undelying DNF probability of the driver.

Since the parameter must lie in the interval [0, 1] and it is believable that it is more likely that a driver finishes than retires, it is possible to adopt a right-skewed Beta distribution as a prior, shown in Equation 5.22 [5]:

$$\theta_j \sim Beta(\alpha, \beta)$$

Equation 5.22

This distribution is flexible and defined on [0, 1] and serves as a conjugate prior to the Bernoulli/Binomial likelihood, as Equation 5.23 [5] is showing. The conjugacy allows the posterior distribution to also be Beta-distributed:

$$\theta_j \leftrightarrow y_j \sim Beta(\alpha + y_j, \beta + n_j - y_j)$$

Equation 5.23

This posterior distribution reflects both the prior beliefs and the empirical data, leading to more stable and realistic estimates. Even drivers with no DNFs obtain non-zero DNF probabilities, reflecting the inherent uncertainty.

The estimated shape parameters $\hat{\alpha}, \hat{\beta}$ are derived in Equation 5.24 from the empirical mean and variance of the DNF rates across the population using the method of moments [5]:

$$\hat{\alpha} = \left(\frac{1-\hat{\mu}}{\hat{\sigma}^2} - \frac{1}{\hat{\mu}}\right)\hat{\mu}^2, \;\; \hat{\beta} = \hat{\alpha}\left(\frac{1}{\hat{\mu}} - 1\right)$$

Equation 5.24

Where:

- $\hat{\mu}$ is the average DNF rate,
- $\hat{\sigma}^2$ is its variance across all drivers.

The Bayesian estimation approach ensures a principled, data-informed probabilistic robust treatment of driver-specific retirement probabilities.

As previously mentioned, to compute accident probability $P_{accident}$, a Beta distribution is fitted using empirical retirement rates from 2019 to 2023. This process leverages prior knowledge and observed race data, preventing unreliable zero-probability estimates for drivers without recorded DNFs. Table 10.3 in Chapter 10 lists the posterior Beta distribution parameters, α and β, alongside the retirement percentage upon which they are computed. Retirement percentage is used instead of the number of retirements to normalize the retirements across the disputed races by every driver, since not every driver has competed along all the seasons between 2019 and 2023. This tabulated representation highlights the interplay between sample size and inferred accident probability.

In parallel, driver-specific probabilities are further detailed in Chapter 10, Table 10.4, where $P_{accident}$, $P_{accident,lap1}$ and $P_{failure}$ are explicitly listed. This structured breakdown enables targeted assessments of both driver-induced and mechanical related DNFs.

To enhance visual intuition, comparative graphical representations in Figure 5.12 illustrate the PDF envelopes of selected driver distributions. These distributions contrast a driver with low retirement percentage, characterized by a sharp, right-skewed Beta distribution (higher value of β, leading to a sharp probability peak near 0), and a driver with a high retirement percentage, whose posterior distribution exhibits a broader, left-skewed profile reflecting greater uncertainty (higher value of α). The envelope diagrams highlight how the Bayesian approach refines individual probability estimates while maintaining realistic bounds.

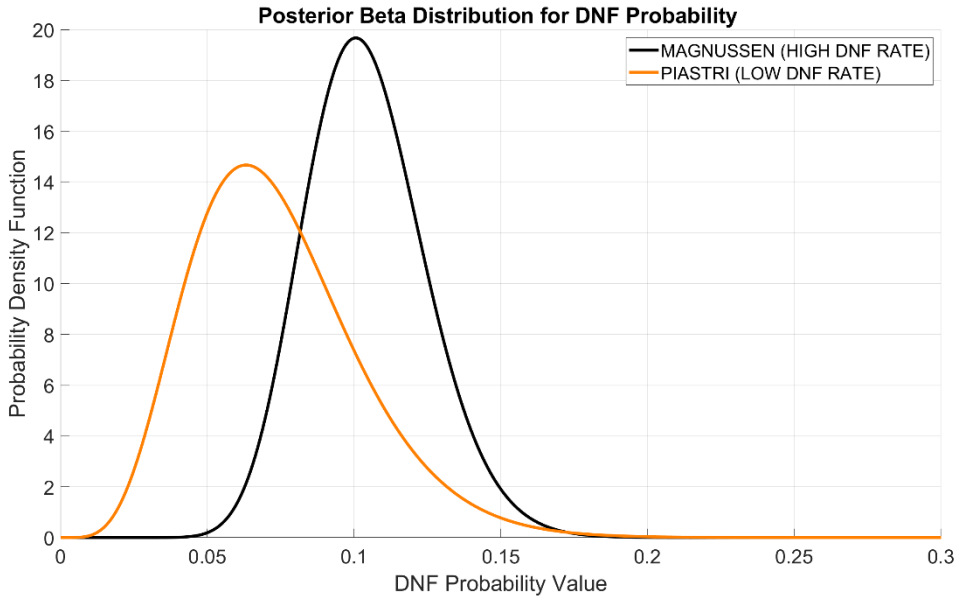

Figure 5.12: Posterior Beta distributions of two drivers with respectively low retirement percentage (Piastri) and high retirement percentage (Magnussen)

5.8 Full Course Yellow (FCY) Phases Model

One of the most critical elements in accurately reproducing the unpredictable nature of a Formula One race is the integration of FCY phases. These interruptions, which manifest as either SC or VSC deployments, are vital safety procedures triggered by on-track incidents or hazardous conditions. While introduced for safety reasons, they have considerable strategic consequences. Their effects permeate not only lap times and time gaps between competitors but also alter pit stop timing and overall race flow.

The present chapter is structured to first introduce and explain the nature, classification and race influence of FCY phases in a comprehensive manner. Subsequently, the implementation strategy used in the simulator is described in full detail, including the methods used to define deployment frequency, timing and lap-time impact.

5.8.1 Understanding FCY Phases

In Formula One, FCY phases are officially declared when conditions on the track require the temporary neutralization of racing speeds across the circuit. These conditions typically arise from accidents, debris or vehicle breakdowns that pose risks to drivers or marshals. The deployment of FCY phases ensures that safety procedures can be conducted without interference from competitive on-track activity. Unlike local yellow flags, which only influence specific sectors of the track, FCY phases apply uniformly to the entire circuit.

There are two main variants of FCY phases, each with their own characteristics and effects:

- VSC is a virtual speed control system. When deployed, it requires all drivers to reduce speed and comply with a minimum lap time constraint, set to around 140% of the lap recorded under green flag conditions on the basis of Heilmeier's work [4]. This regulation ensures that all cars slow down uniformly. Crucially, under VSC, overtaking is prohibited and the time intervals between cars are mostly preserved.
- SC introduces a physical car onto the circuit. This car, dispatched from the pit lane, positions itself in front of the race leader and sets a significantly reduced pace. Drivers form a queue behind it and must refrain from overtaking. As the field bunches up, any pre-existing time gaps between drivers effectively vanishes. SC lap times are generally set around 160 % of the unaffected lap time on the basis of Heilmeier's work [4].

Moreover, both FCY types share a common strategic implication: they reduce the relative time lost when performing a pit stop. This is because the in-lap and out-lap, which are typically slower than normal race pace, become proportionally less costly when the entire field is circulating at reduced speed. AS a result, FCY phases present prime opportunities for teams to perform pit stops with minimal penalty.

Figure 5.13 illustrates the evolution of lap times during the 2018 French Grand Prix, showing SC and VSC periods and their respective slow-sown levels relative to the green flag condition lap time. Horizontal markers at 140% and 160% of lap time are used to delineate the expected boundaries.

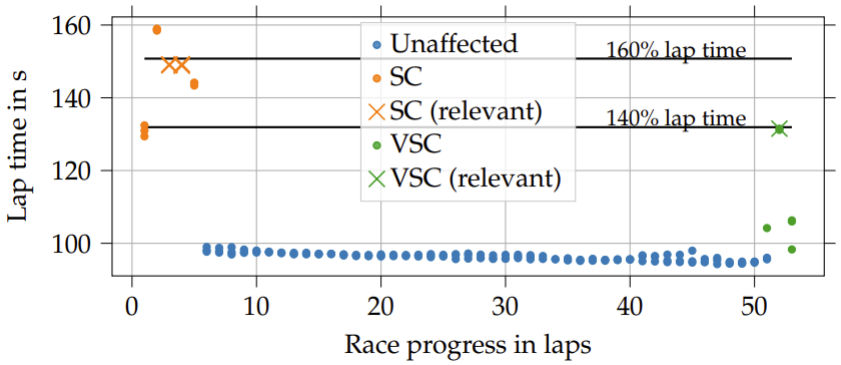

Figure 5.13: Driver lap times relative to the French Grand Prix. Crosses mark the average lap times chosen to identify the lap time increase during SC and VSC phases. Horizontal lines indicated the identified lap time increase in both SC and VSC phases with respect to the baseline lap time [4]

5.8.2 Empirical Insights and Statistical Behavior

The characterization of FCY phases in the simulator is grounded in the empirical analysis of Formula One data from the 2014 to 2019 seasons, as discussed in Heilmeier's study [4]. The implementation of these phases follows a carefully defined sequence that guarantees both statistical realism and race fidelity.

The first operation in the modelling of FCY events occurrence during a race simulation involves the determination of SC phases and their relation to on-track accidents. The number of SC phases expected in a race is sampled from empirical distributions and, for each selected phases, a specific start interval is chosen across the race distance. These intervals are refined using uniformly distributed random perturbations to simulate the fact that SC do not necessarily align with exact lap boundaries. Durations are also drawn from real-world data. Every SC event is considered to be the consequence of an accident, but for simplicity, only one driver per SC event retires, selected according to the individual crash probabilities reported in Table 10.4.

The second step concerns the occurrence of mechanical failures. These are randomly assigned to drivers based on the individual failure probabilities reported in Table 10.4. Notably not every failure result in a VSC as damaged cars may be able to get back to the pits or to stop at a safe location. A conditional rule reported in Equation 5.27 is applied to determine whether a failure should trigger a VSC event or not.

The third key aspect is converting the timing of FCY events from race progress, which is lap-based, to absolute race time. This ensures that all drivers experience the phase at the exact same race moment, regardless of their position or whether they are lapped. Without this conversion, inconsistencies would emerge, such as lapped drivers being affected a full lap earlier than race leaders. Aligning FCY activations to race time guarantees synchronized behavior across the grid.

To preserve realism, the ending of an FCY phase should not overlap with the beginning of another FCY phase during the race. To ensure realism, FCY phases are spaced to avoid overlap by imposing a temporal separation condition, expressed in Equation 5.25 and Equation 5.26 [4]:

$$r_{fcy,start,new} \leq r_{fcy,end,existing} + r_{fcy.distance}$$

Equation 5.25

$$r_{fcy,start,existing} - r_{fcy,distance} \leq r_{fcy,end,new}$$

Equation 5.26

Where:

- $r_{fcy,start,new}$ is the start of the new FCY phase,
- $r_{fcy,end,existing}$ is the end of the existing FCY phase,
- $r_{fcy,distance}$ is the minimum required gap between FCY phases (set to one lap in the simulation),
- $r_{fcy,start,existing}$ is the start of the existing FCY phase,
- $r_{fcy,end,new}$ is the end of the new FCY phase.

The two inequalities ensure that a new FCY begins only after the previous one has ended, maintaining a minimum required gap of at least one lap between consecutive FCY events and that it concludes without overlapping any earlier FCY start. This avoids any unrealistic stacking of FCY events [4].

5.8.3 FCY Simulation Implementation

To translate the statistical behavior and functional impact of FCY phases into the simulation environment, a structured and pre-defined approach is required. Based on the framework introduced earlier, the simulator reproduces SC and VSC events by determining their number, timing and characteristics before the race begins. This ensures that their influence on race dynamics is applied consistently across all drivers.

5.8.3.1 SC Phase Generation

The probability distribution of SC deployments across races shows that approximately 45% of races experience no SC, around 41% have one SC and the residual percentage includes two or more events. This frequency $P_{sc,quantity}$ is graphically represented in the bar chart of Figure 5.14 and summarized in Table 5.1 [4].

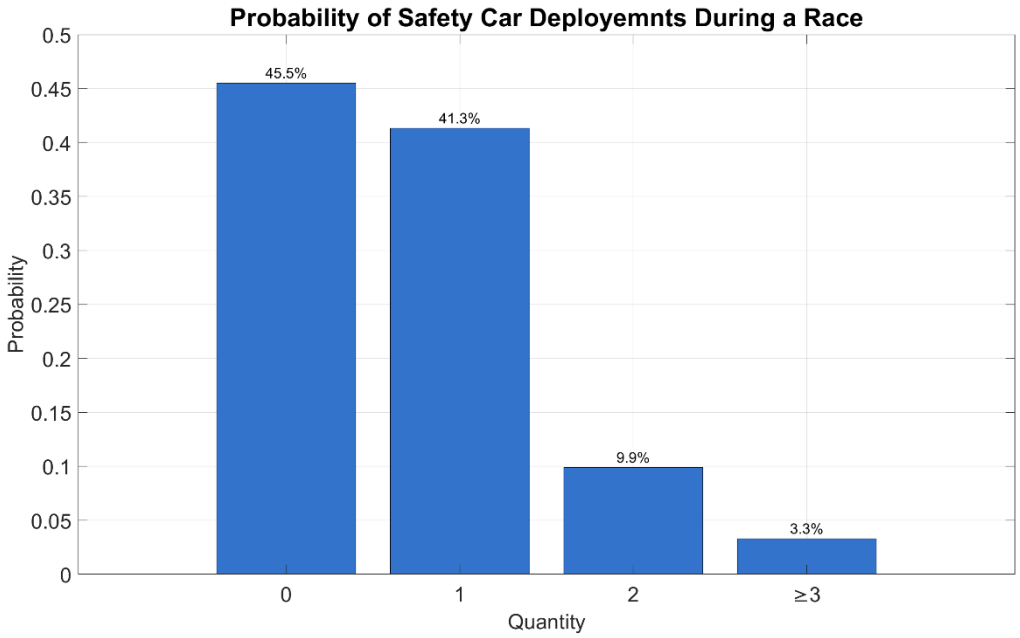

Figure 5.14: Number of SC deployments [4]

Table 5.1: Probabilities relative to the historical quantities of SC deployment during a single race [4]

Probability	**0 SC**	**1 SC**	**2 SC**	**≥ 3 SC**
$P_{SC,\,quantity}$	0.455	0.413	0.099	0.033

SC phases occur more frequently at the beginning of the race, especially in the first lap where over 36% of SCs are triggered, mainly due to large unpredictability of position changes following the race start. The race is divided into six intervals based on the progress (in particular first lap, ≤20%, ≤40%, ≤60%, ≤80%, ≤100%) and each bin has a specific deployment probability $P_{sc,start}$, as modelled by Heilmeier in Figure 5.15 and summarized in Table 5.2 [4].

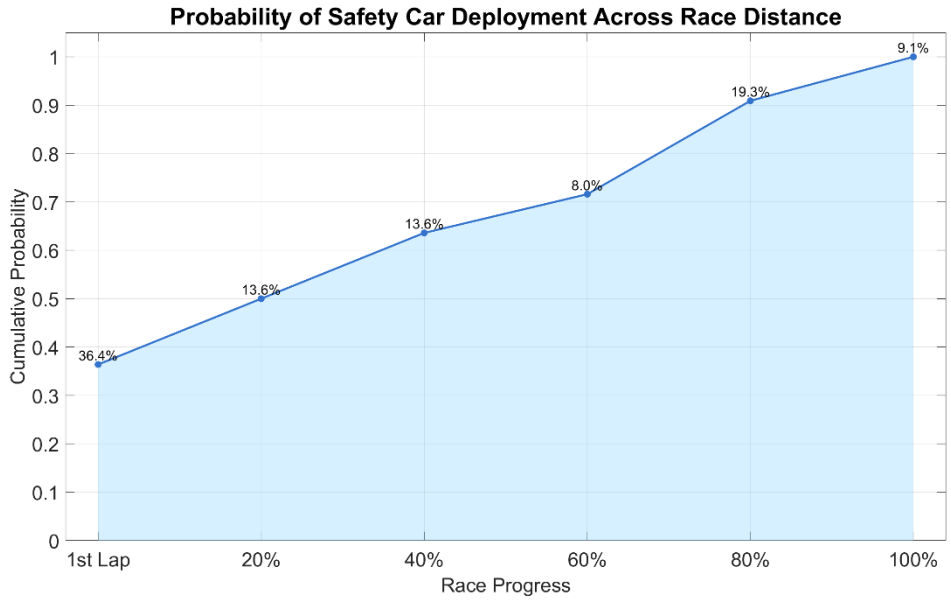

Figure 5.15: Cumulative probability distribution of the start of SC phases during the course of a race [4]

Table 5.2: Probabilities relative to the start of the SC during one of the six race distance groups [4]

Probability	1st Lap	2nd Lap to 20%	20% to 40%	40% to 60%	60% to 80%	80% to 100%
$P_{SC,\,start}$	0.364	0.136	0.136	0.08	0.193	0.091

From Heilmeier's study [4], Table 5.3 summarizes the empirical duration probabilities of SC phases $P_{sc,duration}$. It can span from 2 to 8 laps, with 3-lap and 4-lap durations being the most common. This variability reflects real-world dynamics where track clearance and incident severity vary greatly. The start of the SC phase is further modified by introducing a uniform distribution $\mathcal{U}(0,1)$ to take into account the fact that the SC phase could start at a random point within a lap.

Table 5.3: Probabilities relative to the lap-based duration of the SC [4]

Probability	2 Lap	3 Laps	4 Laps	5 Laps	6 Laps	7 Laps	8 Laps
$P_{SC,\,duration}$	0.182	0.250	0.227	0.193	0.057	0.068	0.023

Every SC phase is assumed to be caused by an accident. Although real races may involve multiple cars, the simulator simplifies this by retiring only one driver per SC event [4]. This driver is selected based on their individual accident probability $P_{accident}$, derived through Bayesian inference model introduced in Chapter 5.7 and explicitly listed in Table 10.4.

5.8.3.2 VSC Phase Generation

VSC phases are generated independently and are associated with mechanical failures rather than accidents. Each driver is assigned a probability of experiencing a failure, $P_{failure}$, based on historical trends or team-specific reliability metrics, listed in Table 10.4. When a failure occurs, the simulator decides, based on a predefined probability threshold reported in Equation 5.27, whether it escalates into a VSC phase.

$$P(VSC|failure) = \frac{n_{VSC}}{n_{failures}} = 0.227$$

Equation 5.27

Where:

- n_{vsc} is the number of VSC phases,
- $n_{failure}$ is the number of failures.

This formulation is based on data recorded by Heilmeier [4] from the 2015 to the 2019 seasons, since the VSC was first introduced in 2015 as part of the regulation after the fatal incident involving the Marussia Racing Team driver Jules Bianchi during the 2014 Japanese Grand Prix.

Based on Equation 5.27, most failures result in no race neutralization at all, reflecting real-world situations where drivers are able to retire in a safe location. If the failure is deemed disruptive enough, a VSC phase is triggered.

Once triggered, the VSC is assigned a start point randomly across the remaining race time and it is constrained to not overlap with existing FCY events. The duration of a VSC is generally shorter than that of SC phases and it is drawn from a narrower empirical range, reported in Table 5.4, reflecting the typically faster recovery time of minor stoppages. To reflect the fact that a SC phase can begin at any moment within a lap, the model uses a uniform random distribution between 0 and 1, $\mathcal{U}(0,1)$, to randomly place the start of the SC period inside the lap. For VSC phases, this same approach applies both to the start and the end of the VSC period, allowing these events to occur at any point within a lap rather than strictly at lap boundaries.

Table 5.4: Probabilities relative to the lap-based duration of the VSC [4]

Probability	**1 Lap**	**2 Laps**	**3 Laps**	**4 Laps**
$P_{VSC,\,duration}$	0.479	0.396	0.021	0.104

5.8.3.3 Conversion to Race Time

A critical step in the FCY implementation is the conversion of all event triggers from race progress (lap number) to absolute race time. This ensures that each driver, regardless of their current position or lapped status, experiences the effect of the SC or VSC at the exact same moment in the simulation timeline.

As explained by Heilmeier [4], if this step were omitted and triggers remained tied to lap-based progress, it would result in inconsistencies: lapped drivers would encounter the FCY earlier than race leaders, which contradicts actual Formula One behavior. For instance, a SC deployed at the moment the leader crosses the line to start lap 30 should simultaneously impact a lapped driver still on lap 29.

To prevent such discrepancies, the simulator uses cumulative race time (in seconds) as the universal reference. Before the main simulation is run, a pre-simulation with a single driver is executed to estimate when each stage of race progress occurs in real time. This enables the translation of lap-based FCY triggers into consistent race-time markers. Although minor deviations may exist between the pre-simulation and actual race [4], these shifts affect all drivers equally, ensuring that every FCY phase remains synchronized across the field.

All the drivers selected to be involved in accidents and failures are simply not taken into consideration during the simulation of the laps following their retirement.

5.8.4 VSC Modelling and Implementation

The VSC provide an alternative method to neutralize the race with respect to the physical SC by reducing speeds across the entire track without requiring cars to bunch up. It allows marshals to safely intervene during minor incidents while minimally disrupting the flow of the race. This section details the modelling rationale and the computational strategy used in the simulator, drawing directly from Heilmeier's methodology [4].

According to Heilmeier, the VSC is modelled by artificially increasing the lap time of any driver affected by the phase. When a VSC phase is active, each affected driver's lap time is artificially extended. The base lap time during such conditions is set to 140% of the base lap time, as mentioned in Equation 5.28, simulated the reduced pace mandated by race control:

$$t_{base,VSC} = 1.4 \cdot t_{base}$$

Equation 5.28

In addition to slower lap times, the VSC phase significantly lowers mechanical strain. The simulator captures this by adjusting fuel consumption and tire degradation. Fuel usage is reduced to 50% of the usual per-lap rate, reflecting the reduced throttle input and engine load, as mentioned in Equation 5.29:

$$B_{fuel,VSC} = 0.5 \cdot B_{fuel,normal}$$

Equation 5.29

Adjusting the fuel mass consumed during VSC conditions as:

$$m_{fuel,consumed,VSC}(l) = B_{fuel,VSC} \cdot l = 0.5 \cdot m_{fuel,consumed,normal}(l)$$

Equation 5.30

This represents the lower engine load associated with SC conditions. The remaining 50% of the expected fuel is not discarded, rather it is virtually stored as a fuel saving buffer. After the VSC phase ends, this stored fuel is automatically redistributed across the remaining race distance. At each subsequent lap, the simulator recalculates the per-lap fuel consumption as:

$$B_{fuel,adjusted} = \frac{m_{fuel,remaining}}{n_{laps,remaining}}$$

Equation 5.31

Where:

- $m_{fuel,remaining}$ is the driver's fuel level at the end of the SC phase,
- $n_{laps,remaining}$ is the number of laps left.

This

Equation 5.8 as follows in Equation 5.32:

$$t_{tire}(a_{tire}, c_{tire}) = (0.5 \cdot\ a_{tire})^2 \cdot k_{1,quad}(c_{tire}) + (0.5 \cdot a_{tire}) \cdot k_{2,quad}(c_{tire}) + k_3(c_{tire})$$

Equation 5.32

However, since VSC phases can begin or end at any moment within a lap, the simulation must account for partial-lap effects. In such cases, the lap is segmented into two fractions:

$$f_{normal} = \frac{t_{start}^{VSC} - t_{end}^{lap-1}}{t_{lap,normal}}$$

Equation 5.33

$$f_{VSC} = 1 - f_{normal}$$

Equation 5.34

Equation 5.33 and Equation 5.34 apply when a VSC phase begins during the current lap but was not active in the previous one, resulting in the lap time being defined as in Equation 5.35:

$$t_{lap} = f_{normal} \cdot t_{lap,normal} + f_{VSC} \cdot t_{lap,VSC}$$

Equation 5.35

If the VSC covers the entire lap, then $f_{normal} = 0$ and $f_{VSC} = 1$, resulting in the lap time being fully defined by $t_{lap,VSC}$.

In these expressions:

- t_{start}^{VSC} denotes the race time at which the VSC phase begins,
- t_{end}^{lap-1} refers to the race time of the driver at the end of the previous lap,
- $t_{lap,normal}$ is the driver's normal lap time under green flag conditions,
- $t_{lap,VSC}$ is the lap time computed with the VSC correction.

These parameters are then used in the simulator to dynamically adjust lap times for each driver depending on when the VSC phase begins relative to their current position on track.

This formulation allows the simulation to replicate the partial and dynamic effects of a VSC phase with high fidelity, reinforcing the stochastic realism of the race environment without compromising computational efficiency.

5.8.5 SC Modelling and Implementation

SC phase is one of the most impactful and strategically significant events in a Formula One Grand Prix. When deployed, the SC physically enters the race track to neutralize the event, requiring all drivers to slow down and form a queue. This results in the compression of gaps between competitors and the suspension of on-track battles, leading to a reset of race dynamics. In order to capture this phenomenon within a simulation environment, a model known as the SC Ghost (SCG) has been adopted, following the methodology introduced by Heilmeier [4].

Unlike a real-world SC that influences all drivers at once, the SCG model created a fictitious version of the SC for each individual driver. This is needed because the simulator processes the race in discrete laps. In a real race, a driver's location on the track at the moment the SC is deployed determines when they are affected. If a single SC model is applied to all the drivers simultaneously, those who were lapped or in different parts of the track would not experience the SC phase accurately, which is the case of a lapped driver who might end up being slowed down one full lap too early. The SCG approach, together with the conversion in race time of the SC start though a pre-simulation as mentioned in Chapter 5.8.3.3, ensure that the SC phase starts at the correct absolute moment for everyone, no matter their track position.

The SCG becomes active for a driver once their cumulative race time exceeds the predefined SC start time. From this moment, the simulation enters a two-stage process, as it is better illustrated:

- Run-up stage: The driver slows down to a lap time equivalent to FCY conditions. This is meant to simulate the reduced pace needed to approach and eventually reach the SCG, which drives at a slower pace than the drivers to bunch them up,
- Following stage: After a few laps, the driver catches up to the SCG, which drives at a lap time set to 160% of the base lap time, defined in Equation 5.36 [4]. During this stage, the driver must remain behind the SCG and cannot overtake it. Their progress becomes synchronized with the ghost, reproducing the real-world situation where drivers follow the SC in formation.

$$t_{lap,SCG} = 1.6 \cdot t_{base}$$

Equation 5.36

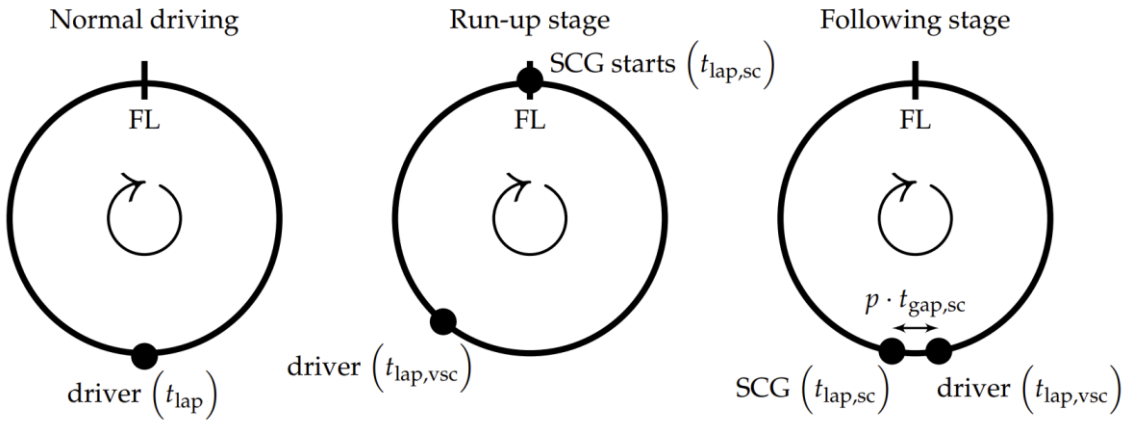

Figure 5.16: Illustration of the SCG concept, highlighting the normal, run-up and following stages [4]

During the SC phase, several important race dynamics are altered. First, normal fuel consumption is applied until the SC phase begins. From that point on, fuel usage is reduced to 25% of its nominal value [4], as it is shown in Equation 5.29:

$$B_{fuel,SC} = 0.25 \cdot B_{fuel,normal}$$

Equation 5.37

Adjusting the fuel mass consumed during SC conditions as:

$$m_{fuel,consumed,SC}(l) = B_{fuel,SC} \cdot l = 0.25 \cdot m_{fuel,consumed,normal}(l)$$

Equation 5.38

This represents the lower engine load associated with SC conditions. The remaining 75% of the expected fuel is not discarded, rather it is virtually stored as a fuel saving buffer. After the SC phase ends, this stored fuel is automatically redistributed across the remaining race distance. At each subsequent lap, the simulator recalculates the per-lap fuel consumption in the same way as previously mentioned in Equation 5.31.

This ensures a smooth consumption of the entire tank by the end of the race, keeping energy balance consistent.

A similar approach is used to model tire degradation reduction. During SC conditions, the simulator scales the rate to only 25% of a lap, according to Heilmeier's research [4]. This is achieved by increasing the lap counter used to identify the tire age inside the degradation function by 0.25 rather than 1.0 per lap, modifying $t_{tire,quad}(a_{tire}, c_{tire}) = a_{tire}^2 \cdot k_{1,quad} + a_{tire} \cdot k_{2,quad}(c_{tire}) + k_3(c_{tire})$

Equation 5.8 as it is shown in Equation 5.39:

$$t_{tire}(a_{tire}, c_{tire}) = (0.25 \cdot a_{tire})^2 \cdot k_{1,quad}(c_{tire}) + (0.25 \cdot a_{tire}) \cdot k_{2,quad}(c_{tire}) + k_3(c_{tire})$$

Equation 5.39

Because the tire model is quadratic with respect to the lap counter, this modelling approach results in a realistic mitigation of performance loss.

In scenarios where the SC phase begins within a lap (i.e. not precisely at the lap boundary), partial-lap handling is necessary. This approach, identical to what has been defined for the VSC in Chapter 5.8.4, involves computing two fractions:

- f_{normal} is the portion of the lap completed before the SC starts,

- $f_{SC} = 1 - f_{normal}$ is the portion affected by the SC phase.

The resulting mixed lap time t_{lap} is then computed by weighting the normal and SC-affected durations, as show in Chapter 5.8.4, Equation 5.35 (recalling that if the SC phase involves the entire lap, $f_{normal} = 0$ and $f_{SC} = 1$).

As the race proceeds, each driver continues catching their respective SCG. Once the driver's cumulative race time becomes less than or equal to that of the ghost, the simulator enforces a lock on their position, preventing any virtual overtaking. The driver is now considered to be in the following stage, not in the run-up stage anymore.

To reproduce the characteristic bunching effect observed during SC phases, the simulator applies a minimum temporal gap between drivers based on their on-track position. This ensures that the field compresses but does not overlap, maintaining a realistic restart formation. The time gap per driver is defined as in Equation 5.40:

$$t_{gap,i} = p_i \cdot t_{gap,SC}$$

Equation 5.40

Where:

- p_i is the driver's position in the current race order,
- $t_{gap,SC}$ is the spacing unit.

The minimum spacing unit is defined in Equation 5.41 as:

$$t_{gap,SC} = \frac{10 \cdot L_{car}}{L_{lap}} \cdot t_{lap,SCG}$$

Equation 5.41

This formulation reflects a core requirement from the FIA Formula One Sporting Regulations, which state that during a SC period, drivers must not fall more than 10 car lengths behind the car ahead while following the SC [25]. Given that a typical Formula One car length L_{car} is approximately 5.65 meters [26], this leads to a spacing of 56.5 meters. To adapt this rule to any circuit, the model divides this length by the lap length L_{lap} deriving the proportion of a full lap that this gap represents. This fraction is then scaled by the SCG lap time $t_{lap,SCG}$ (defined in Equation 5.36) to yield consistent time-based gap. This ensures that the spacing remains physically and strategically accurate across tracks of varying lengths.

At the conclusion of the SC phase, the SCG is deactivated for each driver at the end of that lap. The entire procedure captures the strategic implication of the SC, including compressed fields, neutralized gaps and reduced fuel consumption and tire wear, in a manner that is both computationally robust and more realistic.

5.8.6 Pit Stop Time Loss Adjustments

Pit stops represent a critical tactical moment in any Formula One race and their associated time loss plays a major role in shaping race outcomes. Under normal green flag conditions, the in-lap and out-lap of a pit stop contribute significantly to the total time loss. However, during FCY conditions, this

time loss is substantially reduced. The simulation therefore incorporates a refined model that dynamically adjusts the pit entry and exit penalties whenever a stop occurs under FCY conditions.

According to Heilmeier's analysis [4], cockpit camera footage from the 2018 and 2019 Formula One seasons (sourced from F1TV [14]) was used to empirically determine the reduction in pit stop time loss under neutralized conditions. The findings revealed that the in-lap and out-lap time losses under SC and VSC phases were considerably lower than during green flag laps. For instance, during the 2019 Spanish Grand Prix, the time lost entering and exiting the pit lane dropped from 19.04 seconds in normal conditions to 10.03 seconds under VSC and just 7.88 seconds under SC conditions.

In reality, these values are track-dependent and not always available, therefore the simulator generalizes these effects using practical multipliers:

$$t_{pit,in/out,VSC} = 0.5 \div 0.7 \cdot \ t_{pit,in/out}$$

Equation 5.42

$$t_{pit,in/out,SC} = 0.4 \div 0.6 \cdot t_{pit,in/out}$$

Equation 5.43

This reduction is justified by the fact that while the field is slowed on track, the pit lane speed limit remains constant. As a result, drivers pitting during a neutralization phase gain an effective advantage, losing less time relative to their competitors.

In addition to dynamically scaling pit time losses, the simulator features a proactive pit strategy adaptation mechanism. This approach goes beyond Heilmeier's original model by allowing pit stops to be anticipated if a VSC or SC period begins within a short window before the originally planned stop.

Specifically, if the difference between the lap at which a FCY phase begins and the driver's originally planned pit stop is within a range of approximately two to five laps [27], a window which is consistent with real-world strategic planning for potential SC interventions, the simulator anticipates the stop and executes it in the current lap. This adjustment reflects real-world strategic practices where teams exploit neutralized conditions to minimize time loss during pit stops. In practical racing operations, strategy teams continuously evaluate scenarios in which a SC might be deployed in the subsequent laps. They proactively determine whether a stop under these conditions would offer a competitive advantage. When the team judges that pitting under a SC would be beneficial, the driver is informed that they are in the SC window. This communication ensures that the driver can pit immediately without requiring further confirmation, while the pit crew stands by ready to perform the stop as soon as the SC is deployed. Once the stop is moved forward, the tire compound and stint number are updated accordingly. For multi-stop strategies, subsequent scheduled stops are adjusted in parallel to maintain consistent stint distribution across the race.

This approach captures the dynamic nature of Formula One race strategy, where teams often reconsider their scheduled pit stops to take advantage of the reduced pit lane time loss during FCY phases.

5.9 Overtaking Model

Among the most dynamic and impactful aspects of Formula One race, overtaking maneuvers represent a vital feature in shaping race outcomes. Unlike deterministic lap time simulations that evolve in isolation, a realistic race simulator must also consider the interactions between competitors sharing the track. For this reason, a dedicated overtaking model is introduced within the simulator to reflect both strategic and stochastic influences behind positional changes during the race.

The core purpose of including an overtaking model is to simulate how trailing drivers may attempt to pass a leading competitor, under defined conditions that regulate both the feasibility and the probability of success. These interactions are not only central to the realism of a race simulation but also serve to influence cumulative lap times and final classifications.

The overtaking logic implemented in this simulator is based on the model proposed by Sulsters [5] and integrates additional design insights taken from Heilmeier [3] and Bekker [13].

5.9.1 Main Overtaking Parameters

A successful overtaking maneuver is governed by a structured set of conditions and parameters, as explained by Sulsters' work [5]:

- Minimum time difference δ_{min} is the closest permissible distance between two cars and it ensures that no driver may be unrealistically close to each other in cumulative race time ($\delta_{min} = 0.2s$),
- Overtaking threshold $\alpha_{overtake}$ combines both the strategic complexity of a specific track and the necessary performance gap between competitors. First, it acts as a strict requirement: a driver behind must be fast enough relative to the car in front to justify an overtaking attempt. Second, it reflects how difficult overtaking is at a particular circuit. For example, narrow and twisty tracks will typically have much smaller $\alpha_{overtake}$ (its value must always be negative), making overtakes rare, while circuits with long straights will allow a more permissive value. Mathematically, an overtaking attempt can only happen if the race time gap $\delta_{trailing/leading}$ between the trailing and leading driver is more negative than the following threshold:

$$\delta_{trailing/leading} = t_{trailing} - t_{leading} < \alpha_{overtake}, \qquad \alpha_{overtake} < 0$$

Equation 5.44

$\alpha_{overtake}$ is track-dependent and it is derived using a linear interpolation based on the number of overtakes recorded at each track during the 2022 season [28]. The formulation is defined as:

$$\alpha_{overtake} = -0.6 + 0.4 \cdot \frac{N_{overtakes} - N_{min}}{N_{max} - N_{min}}$$

Equation 5.45

Where:

- $N_{overtakes}$ denotes the number of on-track overtakes observed at a given circuit,
- N_{min} and N_{max} represent the minimum and maximum number of on-track overtakes across all circuits considered,

- $C \cdot \frac{N_{overtakes} - N_{min}}{N_{max} - N_{min}}$ represents the normalization of the number of overtakes into a [0,1] range, equal to 0 when $N_{overtakes} = N_{min}$ and equal to 1 when $N_{overtakes} = N_{max}$.

The resulting value is constrained within the arbitrary range [-0.6, -0.2], with more negative values assigned to circuits where overtaking is rare (i.e. Monaco Grand Prix) and less negative values where overtaking is more frequent (i.e. Brazilian Grand Prix). All circuits are interpolated linearly between these two extremes.

This approach ensures that $\alpha_{overtake}$ becomes more permissive (closer to zero) as the overtaking tendency of a circuit increases, requiring less lap time gap between trailing and leading drivers. Conversely, circuits that historically exhibit fewer passing opportunities impose stricter conditions, requiring the trailing driver to demonstrate a more significant pace advantage to initiate an overtaking maneuver. This continuous scaling guarantees consistency between real-world overtaking patterns and simulated race dynamics.

- Overtaking probability $P_{overtake}$ determines, once an overtaking window is validated by $\alpha_{overtake}$, whether the attempt is successful. Also, the probability $P_{overtake}$ is track-dependent.

$$0 \leq P_{overtake} \leq 1$$

Equation 5.46

Similarly to $\alpha_{overtake}$, the probability of successful overtaking maneuver $P_{overtake}$ is computed using a linear scaling based on the same overtaking reference dataset. Its formulation is given by:

$$P_{overtake} = 0.2 + 0.4 \cdot \frac{N_{overtake} - N_{min}}{N_{max} - N_{min}}$$

Equation 5.47

The resulting probability is constrained within the arbitrary interval [0.2, 0.6], where higher values correspond to circuits that historically allow easier and more frequent overtakes (like Austria and Bahrain), while lower values reflect more layout-constrained venues (like Monaco and Singapore). All circuits are interpolated linearly between these two extremes.

This scaling captures the intrinsic likelihood that a valid overtaking attempt, once permitted by the $\alpha_{overtake}$ threshold, will actually succeed.

- Overtaking time penalties t_{win} and t_{lose} are assigned to respectively the overtaking and overtaken drivers due to sub-optimal racing lines performed due to the overtake. In particular, they are defined as:

$$t_{win} < t_{lose}, \qquad t_{win/lose} > 0$$

Equation 5.48

- Drag Reduction System (DRS) activation window $t_{DRS,window}$ identifies the time gap between the leading and the trailing car to allow the trailing car to activate the DRS (by FIA regulations, $t_{DRS,window} = 1\ s$). If the trailing driver is able to stay below the $t_{DRS,window}$, then he can activate the DRS to facilitate the overtaking maneuver. The DRS is a mechanical system controlled by the driver in pre-determined zones of the circuit, which allows the upper flap of the rear wing to be opened in order to reduce drag, thus increasing speed along the straights [29]. DRS effect $t_{DRS,effect}$ exploits the gain in lap time due to an activated DRS, modelled as a negative contribution. If the trailing car is not fast enough to trigger an

overtaking attempt but it is within the DRS activation window, it receives the DRS effect bonus in the following lap, in order to give a more distinct advantage to attempt the overtake.

Therefore, considering all the boundary conditions to allow the overtake being performed, the overall time loss generated by an overtaking maneuver is modelled in Equation 5.49 as the sum of the DRS effect, if present, and the imposed overtaking penalties:

$$t_{overtake} = t_{DRS,effect} + t_{overtake,win/lose}$$

Equation 5.49

The circuit-specific values of $\alpha_{overtake}$ and $P_{overtake}$, computed using the formulations detailed in Equation 5.45 and Equation 5.47 respectively, are reported in Chapter 10, Table 10.5. These values are derived directly from historical overtaking data and serve as calibrated input parameters for each race circuit modelled in the simulation.

It is important to highlight that overtaking attempts are explicitly forbidden during FCY phases. This aligns with real-world FIA regulations, where such periods are meant to neutralize the race and ensure on-track safety. When such conditions are active, the simulation freezes all positional dynamics and maintains the race order as inherited from the previous lap.

5.9.2 Simulation Flow and Execution Logic

Once the individual lap times are computed for each driver, the overtaking logic is executed following a stepwise flow that captures both structural and probabilistic racing dynamics.

The first check evaluates whether a FCY phase is active during the current lap. In such cases, overtaking is entirely disabled and the order of drivers is preserved exactly as it was at the end of the previous lap. This condition ensures compliance with race-neutralization protocols, as seen in actual FIA regulations.

If no FCY is active, the simulation proceeds to evaluate potential overtaking situations. For each pair of consecutive drivers in the race order, the model first computes the relative time gap. It then verifies whether the trailing driver meets the circuit-specific overtaking threshold, $\alpha_{overtake}$. This step ensures that the following car is not only behind in track position, but also sufficiently faster in pace to justify a pass attempt.

When the threshold condition is satisfied, the model introduces a probabilistic element by performing a random draw governed by the probability parameter $P_{overtake}$. This models the uncertainty and competitive nature of real overtaking attempts, where success is not guaranteed even with a performance advantage.

If the overtake is successful, the positions of the two drivers are swapped. In addition, overtaking time penalties are applied: the overtaking driver receives a smaller time increment t_{win}, while the overtaken driver incurs a slightly larger penalty t_{lose}. These penalties reflect the time cost of deviating from the optimal racing line during a positional battle. The updated lap times are then added to each driver's cumulative race time.

In the case of a failed overtake, the trailing driver is flagged as DRS-eligible for the next lap, assuming the current time gap lies within the DRS activation window $t_{DRS,window}$. This models the advantage a car can gain in subsequent attempts due to reduced aerodynamic assistance.

This simulation structure ensures overtaking is only processed when a driver is demonstrably faster, positioned close enough and statistically fortunate to convert the move. Each overtaking event,

whether successful or failed, is logged within a summary structure that tracks total attempts and completed maneuvers, to be used in post-race analysis.

To further clarify the interaction between overtaking conditions, Figure 5.17 [5] illustrates the logic of the overtaking model through a visual representation. In the figure, driver d_j denotes the leading care, while driver d_k represents the trailing car attempting an overtake.

The vertical arrows show the difference in cumulative race time between two drivers at the end of a given lap. For an overtaking attempt to be considered, the trailing driver must be faster than the leading one by an amount exceeding the overtaking threshold $\alpha_{overtake}$, indicated on the left of the picture. This situation is shown in the left portion of the diagram, where the time delta between the leading and trailing drivers is $\delta_{j,k} < \alpha_{overtake}$, allowing a probabilistic check for overtaking to occur. If this probabilistic check is passed, the positions are switched, as shown in the center part of Figure 5.17, where the red car is overtaken by the blue car. However, if the time gap is insufficient or the overtake attempt fails the probabilistic condition, the drivers retain their original order, as shown in the right-hand side of the picture. In both scenarios, the model ensures that a minimum time gap δ_{min} is respected to maintain physical plausibility.

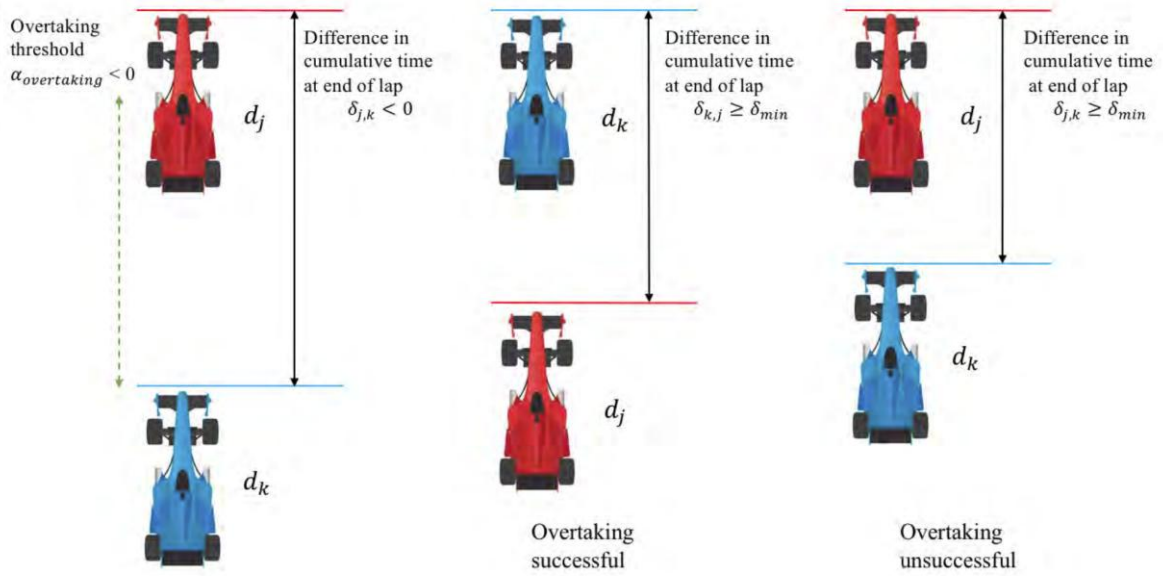

Figure 5.17: Visual representation of the overtaking model between a leading driver d_j and a trailing driver d_k [5]

6 Simulation Results

After the complete definition of all necessary parameters, the simulator is now ready to be employed in a real-case scenario. This chapter presents the results obtained from a single simulation instance. Rather than focusing on statistical ensembles or Monte Carlo simulations at this stage, performing multiple instances of the main simulation, the objective here is to provide a clear and through walkthrough of what the simulator performs, how it processes input data and what kind of information is extracted after the race is completed. The complete race simulation process is driven by a sequence of coordinated actions executed in a modular and scalable framework. This section details, in a comprehensive and discursive manner, the entire simulation pipeline: from global initialization to the lap-by-lap progression and final result computation. Each step mirrors the real-world dynamics and uncertainties of Formula One racing, relying on functions described in the previous chapters. MATLAB has been chosen as the reference environment on which the simulator will run.

6.1 Pre-Requisite Actions Before Simulation Start

The simulation begins by associating each driver with the corresponding team. This association is one of the core inputs required by the simulator, since this linkage is not only essential for preserving the structure of the actual race grid, but it also enables the simulator to correctly retrieve all team-dependent and driver-dependent parameters throughout the simulation and combine them when useful. For instance, pit stop behavior is inherently tied to team affiliation, as each team exhibits its own performance distribution during pit operations. Consequently, accurate team-driver mapping ensures the simulator assigns realistic pit stop outcomes and strategic dynamics to the right driver and team at the same time, faithfully replicating the operational differences observed across the Formula One field.

Subsequently, the circuit where the simulation will be executed is selected. This input is used to retrieve several circuit-specific parameters: lap length, total number of laps, pit lane travel time and Pirelli C-compound selection for the race weekend. These compound mappings are particularly important, as they allow the simulator to switch between the local compound nomenclature (SOFT, MEDIUM, HARD) to the C0-C5 classification issued by Pirelli for each venue. These parameters are critical to accurately reflect the unique characteristics of each Grand Prix venue and are used throughout the simulation to define the time evolution of a driver's performance. Among the most important are the tire degradation coefficients, which include both quadratic and linear terms obtained from race data analysis associated to each compound of the 2023 Formula One season, varying from circuit to circuit. In parallel, the fuel consumption model is also tailored to each circuit based on its lap length and number of laps, governing the lap time impact of decreasing car weight across the race distance. It is crucial to emphasize that the simulator is limited to dry weather conditions only and any circuits marked as wet in the database are automatically excluded from valid simulation runs. This constraint reflects a limitation of the modelling approach, which does not account for variation of the previously computed parameters under wet conditions.

In order to maintain realism and accuracy, the simulator incorporates logic to account for mid-season driver swaps, as occurred in the Formula One championship regarding drivers such as Nyck de Vries, Daniel Ricciardo and Liam Lawson who all replaced each other during some of the races of the 2023 season. The simulator ensures that only drivers who actually participated in the chosen circuit's Grand Prix are included in the simulation, by filtering out invalid pairings based on historical race data. This allows the use of circuit-specific performance data, such as qualifying lap times and tire degradation profiles, which are otherwise unavailable for non-participating drivers.

One of the most critical aspects in configuring the simulation is the initial race strategy definition for each driver. The simulator assumes that the strategy is known before the race begins and that it will not change unless specific race conditions arise (such as the deployment of a FCY phase). The strategy includes a series of structured inputs: the number of planned pit stops, the starting compound and the compound fitted during the pit stops. The compound inputs are defined by using the local circuit nomenclature SOFT, MEDIUM, HARD and automatically adjusted to C-compound nomenclature to fit the specifications for the selected circuit. In addition to the compound choices, the timing of each pit stop is also defined as a fraction of the total race distance. According to inferred strategy ranges, SOFT tires are assumed to last between 17% and 23% of the race, MEDIUM tires between 29% and 37% and HARD tires between 41% and 64%. These figures are not officially published by the FIA or Pirelli but are derived from strategic analyses of past races and race engineering practices [2].

In order to enhance realism, the simulator also includes logic to handle teammate pit stop overlaps. In Formula One, teammates share the same physical pit box during the race and double stacking (pitting both cars on the same lap) is generally avoided unless under exceptional circumstances (like FCY deployment). The simulator checks whether both drivers from the same team are scheduled to stop at the same lap. If so, the system prioritizes the driver ahead on track, allowing him to pit on the intended

lap, while delaying the second driver's stop by one lap. The strategy plan is automatically updated to reflect this change, ensuring that pit stop congestion is avoided without introducing performance penalties that would be unrealistic under normal race operations.

At the end of this initialization phase, the simulator provides a summary table for each driver, showing their assigned strategy. The table includes the driver abbreviation, the associated team, the number of planned stops and the compounds mounted at each race phase. A simplified example is shown in Table 6.1:

Table 6.1: Example of a summary table related to driver-specific expected race strategy input

Circuit	Italian Grand Prix
Driver	Charles Leclerc
Team	Ferrari
Expected Number of Pit Stops	2
Starting Compound	MEDIUM (C4)
Expected First Stop Lap	14
First Stop Compound	HARD (C3)
Expected Second Stop Lap	34
Second Stop Compound	HARD (C3)

This table serves not only as a confirmation of the input parameters but also as a reference to verify the consistency and correctness before launching the actual race simulation. Once all preconditions are satisfied, the simulation can proceed to compute lap-by-lap race performance, incorporating all physical models and stochastic elements described in the previous chapters.

6.2 Main Simulation Workflow

After the initialization of the circuit-specific parameters, driver-teams association and the driver-specific race strategy inputs, the main simulation workflow can be exploited. The successive steps following the pre-requisite inputs are:

1. FCY deployment check,
2. Pre-simulation for FCY time estimation start,
3. Driver-specific initialization,
4. Lap-by-lap simulation core loop.

In this section each step will be treated and explained separately.

6.2.1 FCY Deployment Check

Before the actual simulation begins, the system probabilistically determines whether a SC or a VSC phase will be deployed during the race. These stochastic events are modelled using probability distributions drawn from historical data [4, 5]. The SC deployment is directly linked to the occurrence of an accident and the simulator uses a predefined probability distribution to determine which driver is likely to be involved based on lap-specific accident likelihood. In contrast, VSC phases are only considered if a driver suffers a mechanical failure. However, not all failure events trigger a VSC: the simulator evaluates the failure against an independent probability threshold to decide whether it warrants a race neutralization. Once triggered, the SC or VSC start lap and duration are selected from empirical distributions and a driver is designated as the source of the incident. Overlap between SC and VSC phases is actively avoided by designated control logic.

6.2.2 Pre-Simulation for FCY Time Estimation Start

To accurately map race progress into temporal values, a pre-simulation is performed for a reference driver. This step estimates the clean lap time and the cumulative race time vector, allowing SC and VSC triggers to be converted from fractional laps to seconds, in order to allow every driver, independently from their race location and position, to be affected by the FCY phase at the same time as everyone else.

6.2.3 Driver-Specific Initialization

Each valid driver in the simulation is initialized individually. This involves:

- Assigning the race strategy and tire plan,
- Retrieving the base lap time for the selected circuit,
- Loading tire degradation coefficients relative to the selected circuit (both quadratic and linear),
- Computing launch delays from the grid based on grid slots obtained by qualifying results and individual performance randomness,
- Retrieving team-based pit stop timing parameters.

All these values are stored by the simulator in order to track each state evolution for each driver across laps.

6.2.4 Lap-by-Lap Simulation Core Loop

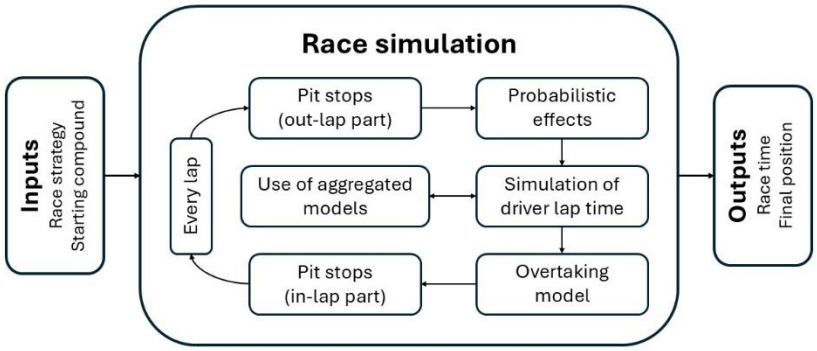

Figure 6.1: Race simulation workflow [4]

The race progresses lap by lap. At each lap, the following processes occur sequentially:

1. Driver-specific Lap Time Simulation
   Each driver completes one simulated lap through the internal routine responsible for lap time computation. This step combines circuit-specific pace, tire degradation effects, remaining fuel load, pit stop windows and random performance variability. Lap time simulation is adapted in case of first lap actions (it adds the starting performance contribution), in case of FCY phases (it slows down the pace of the remaining drivers and, eventually, adjusts the pit stop timings) and in case of pit stops (adds the in-lap and out-lap components).

2. DNF Handling
   If a driver is designated as retired due to a failure or accident, they are removed from the active race pool. Retirement time and reason are stored for post-processing purposes and FCY typology trigger respectively. Their last entry in the lap table is marked as “RETIRED”.

3. Provisional Order
   Active drivers are not yet ranked by cumulative race time, since the overtaking logic has not been used yet. The only applied modification concerns pitting drivers, since they are inserted into the order based on their expected pit loss. This ensures realistic reshuffling during pit phases and does not consider them into the overtaking logic.

4. Overtaking Logic
   The validated order is passed to the overtaking model, which applies probabilistic overtaking decisions over the non-pitting drivers. This model evaluates several factors, including the relative tire compounds between drivers, the time gap separating them and circuit-specific overtaking coefficient. A probabilistic function determines whether an overtaking attempt is initiated and, if successful, it adjusts the relative time performances accordingly. Furthermore, minimum enforced gaps are imposed in the case of failed overtakes to simulate the aerodynamic and strategic penalties associated with close racing without position change. This logic ensures that overtakes occur only through these authorized simulation steps. The overtaking logic also freezes the pecking order in case of SC and VSC phases, since overtaking is forbidden during such conditions.

5. Position Updates
   The final classification for the lap is recorder. Position changes are logged, the leader is identified and each driver's lap result is stored in a cumulative table, recording cumulative and lap-specific times, tire compound, tire age, fuel level, pit status and race position. A consistency check ensures that no driver is erroneously classified ahead of a competitor with a faster total time.

This process results in a complete race timeline, compound usage trace and performance breakdown per driver, ready for post-simulation visualization and statistical analysis. In Figure 6.1 is shown the simulation workflow regarding a single simulation, highlighting the input parameters to be fed into the simulation, the main simulation loop and the output parameters.

6.3 Visual Representation of a Single Simulation Without FCY Phases

Following the execution of a single race simulation, performed under dry conditions and without any FCY phases, a variety of tables and graphs are generated to illustrate the results in a clear and complete way. These outputs are designed to help understand how the race developed over time, what strategies were used and how each driver performed.

The first important output in analyzing the race simulation is the lap-by-lap table for each driver. While the full simulation naturally includes data for every lap completed by every driver, only a selection of representative laps is presented in the final report for clarity and readability. Table 6.2 offers a snapshot of this information and includes key race parameters such as the lap time, the cumulative race time up to that point, the driver's position at the end of the lap, the fitted tire compound, the current tire age, the remaining fuel level and flag that indicate whether a given lap is an in-lap or an out-lap.

This table plays a crucial role in helping the reader reconstruct the driver's race in detail. By following the sequence of laps, it is possible to track the progression of time and race events for a specific driver across the entire race distance. The lap number clearly indicates the current stage of race, while the individual lap time shows the duration required to complete that specific lap. Importantly, the cumulative race time aggregates all completed lap times up to that point, providing a real-time indication of how the driver's total performance is building up over the course of the event.

Several aspects stand out when analyzing the timing data. For example, the first lap is typically affected by the initial mixing at the start, where drivers accelerate from standstill and are subjected to more interactions between each other. This leads to a longer lap time compared to the following laps, as it includes the time spent launching from the grid, finding position and reaching race speed. This initial time loss is consistent with real-world race behavior and is accurately reflected in the simulation.

Further into the race, other timing variations become evident through pit stop operations. When a driver is about to enter the pit lane, he/she is no longer driving at full racing speed for part of the lap, leading to a moderate increase in lap time due to deceleration and pit entry maneuvers. On the subsequent lap, the driver exits the pit lane after the crew performs the pit stop. This lap typically exhibits a significantly longer lap time, as it includes the time lost in the pit lane, comprising pit lane travel, service duration and the return to racing speed. The in-lap and the out-lap together capture the time penalty associated with a pit stop.

Another key aspect visible in the table is the management of tire wear, as reflected by the tire age parameter. Tire age increases by one with every completed lap on the same set of tires. However, when a pit stop is performed and a new set is mounted, regardless of whether the compound is different or the same, the tire age counter resets to 1 on the out-lap. This reflects the beginning of a new stint on fresh tires. For example, considering Lecler as the reference driver in Table 6.2, lap 13 is marked as his in-lap, showing that he enters the pit lane during that lap. Then, on lap 14, his out-lap is registered, indicating the actual pit stop and the start of his next stint. From that point onward, the table shows that he continues on the HARD compound, transitioning from his earlier MEDIUM stint. The tire age restarts accordingly and his lap times begin to evolve based on the new compound characteristics and reduced fuel load.

Table 6.2: Driver-specific (Leclerc) lap-by-lap table capturing key race metrics

Driver	Circuit	Lap	Lap Time [s]	Race Time [s]	Position	Compound	Tire Age	Fuel Level [%]	In-Lap	Out-Lap
Leclerc	Italian GP	1	88.140	88.140	2	MEDIUM	1	98.08	No	No
Leclerc	Italian GP	2	87.162	175.30	2	MEDIUM	2	96.15	No	No
Leclerc	Italian GP	3	86.571	261.87	2	MEDIUM	3	94.23	No	No
…	…	…	…	…	…	…	…	…	…	…
Leclerc	Italian GP	13	87.978	1133.4	3	MEDIUM	14	73.08	Yes	No
Leclerc	Italian GP	14	109.20	1242.6	11	HARD	1	71.15	No	Yes
Leclerc	Italian GP	15	85.513	1328.1	9	HARD	2	69.23	No	No
…	…	…	…	…	…	…	…	…	…	…
Leclerc	Italian GP	52	85.21	4515.8	2	HARD	14	0	No	No

Another useful table is represented by the overview of the final race results for all the drivers. For each participant, it lists the final position, total race time, time gap to the leader, time gap to the driver in front, pit stop strategy and the compound used across the race (S stands for SOFT, M stands for MEDIUM, H stands for HARD). Table 6.3 is useful for comparing overall performance and understanding each driver's strategic decisions.

Table 6.3: Final race results for all the drivers, including several key metrics

Circuit	Driver	Position	Race Time [s]	Gap To Leader [s]	Gap To Driver Ahead [s]	Pit Stop Strategy	Used Compounds
Italian GP	Sainz	1	4511.1	[-]	[-]	2 Stops	M → H → H
Italian GP	Leclerc	2	4515.8	+4.7	+4.7	2 Stops	M → H → H
Italian GP	Verstappen	3	4519.9	+8.8	+4.1	2 Stops	M → H → H
Italian GP	Russell	4	4529.0	+17.9	+9.2	2 Stops	M → H → H
Italian GP	Perez	5	4530.0	+18.9	+1.0	2 Stops	M → H → H
Italian GP	Albon	6	4536.7	+25.6	+6.7	2 Stops	M → H → H
Italian GP	Hamilton	7	4537.3	+26.2	+0.6	2 Stops	M → H → H
Italian GP	Norris	8	4553.0	+41.9	+15.7	2 Stops	M → H → H
Italian GP	Piastri	9	4554.7	+43.6	+1.7	2 Stops	M → H → H
Italian GP	Alonso	10	4572.1	+61.0	+17.4	2 Stops	M → H → H
Italian GP	Lawson	11	4592.6	+81.5	+20.5	2 Stops	M → H → H
Italian GP	Tsunoda	12	4593.1	+82.0	+0.5	2 Stops	M → H → H
Italian GP	Hulkenberg	13	4602.6	+1 Lap	+9.5	2 Stops	M → H → H
Italian GP	Sargeant	14	4605.2	+1 Lap	+2.6	2 Stops	H → M → H

Italian GP	Bottas	15	4607.4	+1 Lap	+2.3	2 Stops	M → H → H
Italian GP	Zhou	16	4628.7	+1 Lap	+21.2	2 Stops	M → M → H
Italian GP	Magnussen	17	4637.9	+1 Lap	+9.2	2 Stops	M → H → H
Italian GP	Ocon	18	4639.5	+1 Lap	+1.6	2 Stops	M → H → H
Italian GP	Gasly	19	4644.2	+1 Lap	+4.7	2 Stops	M → H → M
Italian GP	Stroll	20	4651.8	+1 Lap	+7.6	2 Stops	M → H → H

To complement the tabular outputs, several graphical representations are used to provide a clearer and more intuitive understanding of the race dynamics. One of the most informative among them is the Race Position Evolution plot in Figure 6.2, which traces the position of each driver throughout the race on a lap-by-lap basis. This figure enables a visual reconstruction of the race order and shows how positions changed over time.

A key feature of this plot is that every position change is governed by the overtaking logic described in detail in Chapter 5.9. Only the overtaking driver is actively tracked in this logic, since the overtaken driver is passively repositioned as a consequence. However, not all position changes stem from overtaking attempts. In particular, reordering due to pit stops, where a driver momentarily loses positions while in the pit lane, is not processed through the overtaking logic and thus does not count toward the overtaking statistics. Similarly, positional reshuffling caused by lapping scenarios, where a slower driver is overtaken due to being behind on race distance, is excluded from the overtaking framework.

At the end of each driver's line in the plot, a label reports the driver's abbreviation followed by their final race gap relative to the winner. This summary allows for a rapid visual interpretation of both finishing position and race performance. If a driver finishes the race more than one lap behind the winner, this is denoted explicitly as a lapped status. In such cases, the number of laps by which the driver has been lapped is calculated by dividing the final time gap by the base lap time. For example, a driver with a time gap equivalent to two base lap times will be labeled as "+2 Laps". This provides a simple yet realistic reflection of race distance deficits in a consistent manner with real-world classification standards.

Overall, this plot not only reflects the competitive flow of the race but also helps to isolate critical phases such as mid-race overtakes, pit strategy effects and late-stage performance drops or gains.

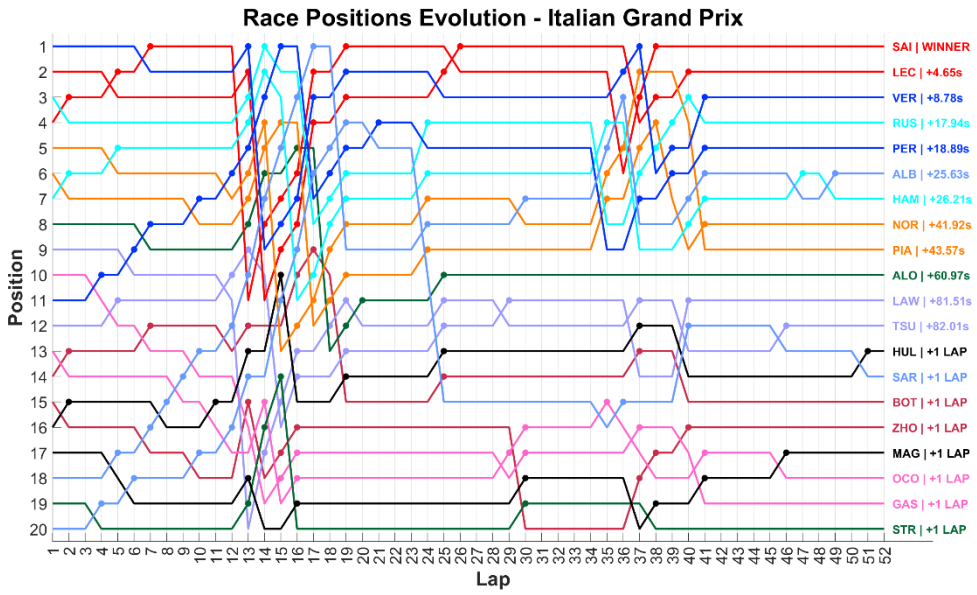

Figure 6.2: Lap-by-lap position changes across the whole race

The Lap Times Race Progression plot in Figure 6.3 provides a visual summary of how each driver's lap times evolve over the course of this race. This figure is particularly helpful for interpreting tire degradation patterns, assessing overall performance consistency and identifying the timing and consequences of pit stops.

Typically, lap times exhibit a gradual upward trend within a stint, reflecting the cumulative effect of tire wear. As the tire compound degrades, grip levels decrease and lap times progressively increase, even in the presence of a reducing fuel load. This trend is especially noticeable in softer compounds.

Superimposed on these trends are sharp spikes (sudden increase in lap time) that correspond to pit stop phases. These spikes a generally twofold: the first occurs during the in-lap, where the driver decelerates and enters the pit lane and the second, larger spike takes place during the out-lap, when the full duration of pit lane travel, tire change and re-acceleration to full speed is absorbed into the lap time. This pattern creates a characteristic signature that visually distinguishes pit stop events from normal racing conditions.

By analyzing these curves, one can easily detect when each driver initiated and completed their pit strategy, how their pace evolved across tire stints and how performance varied in response to compound choice and track conditions. As such, this plot serves as a dynamic tool for evaluating strategic effectiveness and performance resilience throughout the race.

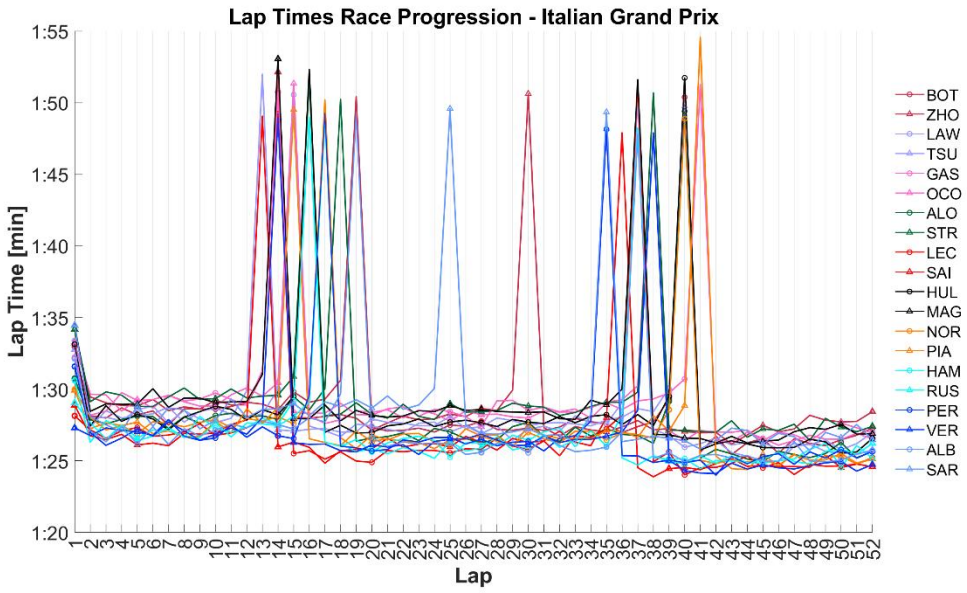

Figure 6.3: Lap times across the whole race for all the drivers

Another helpful figure for evaluating overall race performance is the Cumulative Race Time Summary plot in Figure 6.4. This graph provides a compact yet comprehensive overview of the total race time accumulated by each driver by the end of the simulation. Each driver is represented by a horizontal bar, whose length corresponds to his/her final cumulative race time, allowing for direct visual comparison across the grid.

At the top of the chart, the race winner is clearly identified by having the shortest race time, setting the benchmark for all subsequent gaps. Each remaining driver's bar is accompanied by two reference values: the time gap to the winner and the time gap to the driver immediately ahead in the final classification. This dual gap reporting allows for an intuitive understanding of both the driver's competitiveness against the front of the field and their closeness to adjacent rivals.

In addition to timing metrics, the plot also integrates key strategic information by indicating the number of pit stops performed by each driver. This provides immediate context for understanding driver and team performance and how different strategies may have contributed to overall race time.

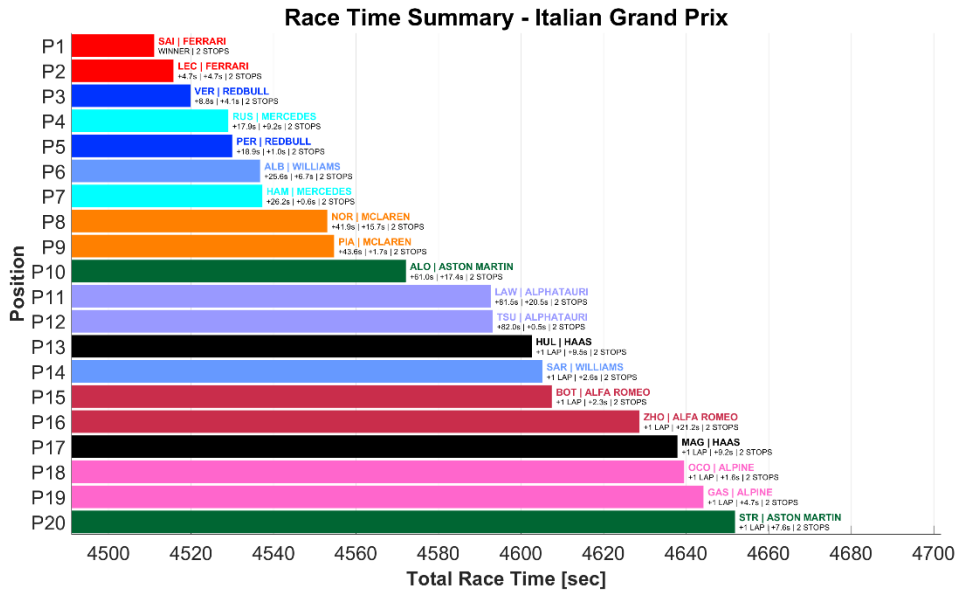

Figure 6.4: Final race summary for all the drivers regarding the time gaps to the leader and to the driver in front

The Race Strategy Summary plot in Figure 6.5 illustrates the tire strategies adopted by each driver throughout the race. Each horizontal bar represents a driver's full race stint, segmented into colored blocks that denote the compound used during each phase: red for SOFT, yellow for MEDIUM and grey for HARD. This visual format allows for a clear and immediate understanding of when tire changes occurred, how long each stint lasted and which compound choices were made.

This plot is particularly valuable for comparing the diversity in race strategies across the field. Not all drivers followed the same tire sequence and the variation in strategy often influenced their final classification. A notable example is Sargeant, who concluded the opening lap in the last position. However, by the end of the race, he had progressed to P14, an improvement largely attributed to his alternative tire strategy. Unlike most of the grid, which opted for a MEDIUM → HARD → HARD sequence, Sargeant ran a HARD → MEDIUM → HARD strategy. This approach allowed him to extend his initial stint, preserve track position during early pit phases and capitalize on fresher tires when others were managing degradation.

In contrast, Gasly employed a MEDIUM → HARD → MEDIUM strategy, which ultimately yielded less favorable results. His middle stint on the HARD compound appeared to stretch too long likely resulting in a drop in performance due to excessive tire wear. Although he returned to the MEDIUM compound for the final laps, the performance advantage it offered was insufficient to recover the time lost earlier in the race. This example highlights how the effectiveness of a strategy is not merely defined by compound selection, but also by the timing and duration of each stint.

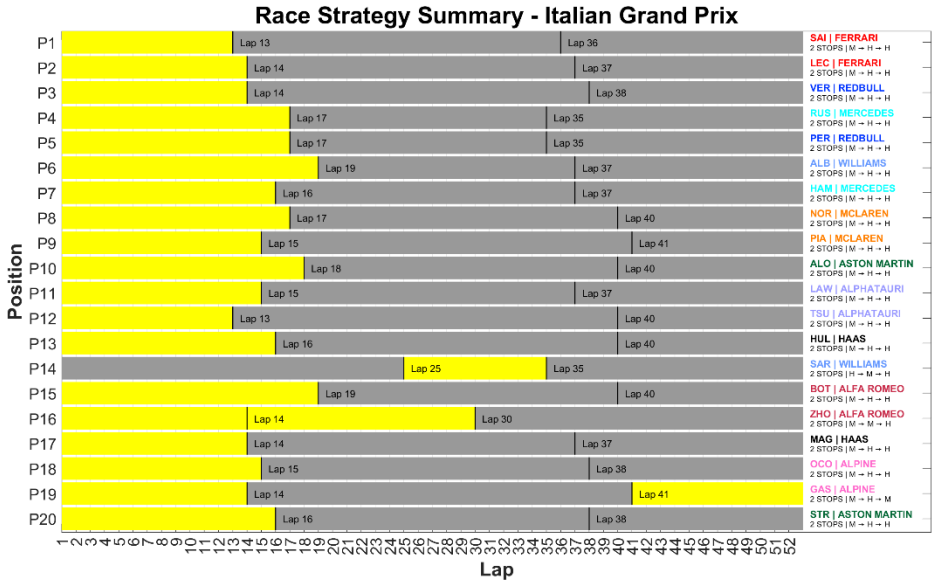

Figure 6.5: Race strategy summary across the whole race for all the drivers

In addition, overtaking activity is displayed through specific diagrams. The Overtaking Summary diagram in Figure 6.6 provides a concise visual record of all overtaking attempts across the entire race, including whether each attempt was successful or not. This plot is relatively straightforward in its interpretation, offering a clear overview of which drivers were more active in challenging rivals and how often those moves resulted in a change of position.

The Lap-Wise On-Track Overtaking Summary plot in Figure 6.7 offers a more detailed breakdown, summarizing overtaking attempts and successful maneuvers on a per-lap basis. This plot is particularly insightful for identifying the most dynamic phases of the race. It is evident from the chart that cars are more closely packed in the early stages, increasing the opportunity for drivers to engage in on-track battles before strategies and pace differentials begin to spread out the field. As the race progresses and gaps stabilizes, both the number of attempts and successes naturally decline.

It is important to note that, in both plots, overtakes resulting from pit stop reordering or from lapping significantly slower drivers are explicitly excluded. This ensures that the data reflects only genuine on-track overtaking scenarios governed by the overtaking logic described in Chapter 5.9. Together, these diagrams offer a clear and quantifiable picture of race engagement, making it easier to assess how competitive the on-track battles were throughout the event.

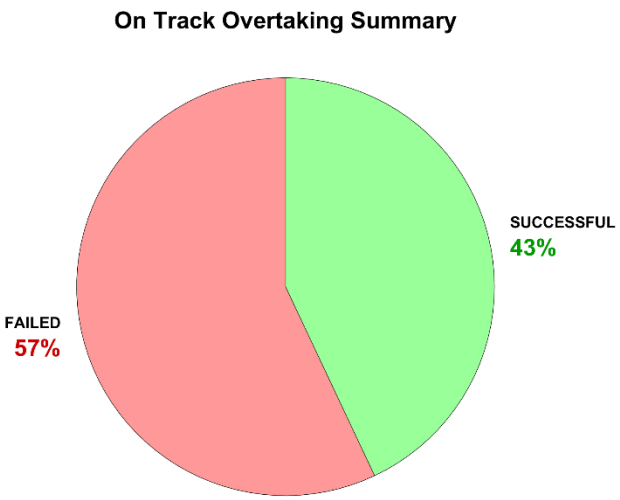

Figure 6.6: On track overtaking summary collecting the successful overtakes against the failed ones

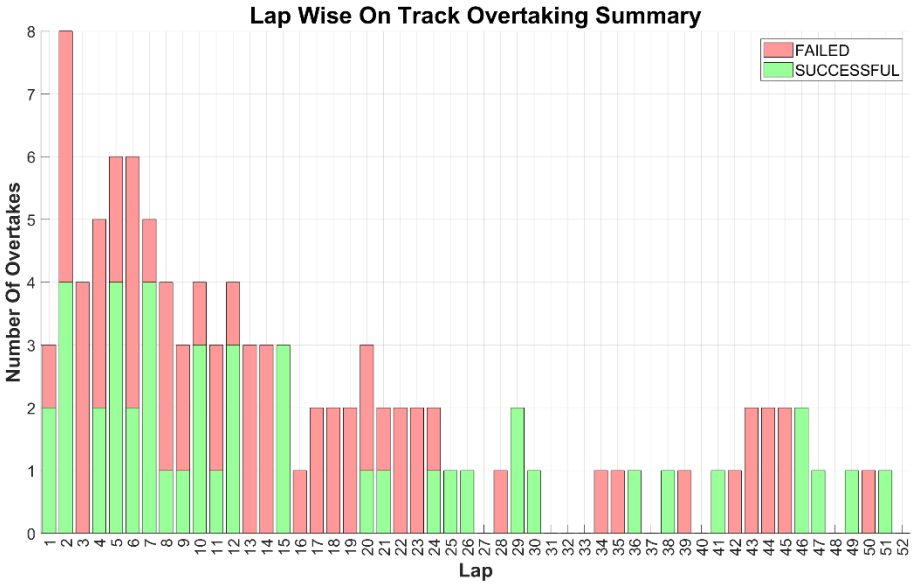

Figure 6.7: Lap-by-lap on track overtaking summary comparing the successful overtakes and the failed ones

Together, these tables and plots give a complete picture of the simulated race evolution, combining detailed numbers with visual clarity. They contribute to explaining how race strategies were executed and how driver performance evolved throughout the race, thereby offering both a quantitative technical assessment and a coherent narrative of the simulation's progression for a single-event scenario.

6.4 Visual Representation of a Single Simulation with FCY Phases

Following the execution of a single race simulation, again under dry conditions but this time involving the deployment of a FCY phase, a series of tables and visual outputs are produced to represent the resulting race dynamics. The presence of an FCY event, whether in the form of a SC or VSC, plays a critical role in shaping the progression and final outcome of the race. These neutralization phases momentarily freeze the relative gaps between drivers, often eliminating carefully built time advantages and compressing the field into tighter proximity. As a result, cumulative race times tend to increase and strategy execution may be significantly altered. The present chapter highlights how such events affect not only the raw performance figures but also the strategic unfolding and competitive landscape of the race.

Table 6.4 illustrates a selection of lap data from a single driver's race (selected driver is Leclerc), offering insights into race dynamics under a SC phase. Between laps 4 and 6 (highlighted in yellow in Table 6.4), a notable increase in lap time is observed, corresponding to the deployment of the SC on lap 4. This slowdown reflects the mandatory pace reduction imposed by race control during neutralization, effectively bunching up the field. During this phase, although the fuel level continues to decrease, the rate of consumption is clearly reduced due to the lower energy demand associated with the slower pace. This highlights how the simulator realistically accounts for fuel-saving effects during neutralized conditions, integrating external race events like the SC with internal performance parameters.

Table 6.4: Driver-specific (Leclerc) lap-by-lap table capturing key race metrics, highlighting SC period (yellow shade)

Driver	Circuit	Lap	Lap Time [s]	Race Time [s]	Position	Compound	Tire Age	Fuel Level [%]	In-Lap	Out-Lap
Leclerc	Italian GP	1	87.834	87.834	2	MEDIUM	1	98.08	No	No
Leclerc	Italian GP	2	86.938	174.77	2	MEDIUM	2	96.15	No	No
Leclerc	Italian GP	3	86.972	261.74	2	MEDIUM	3	94.23	No	No
Leclerc	Italian GP	4	93.702	355.45	2	MEDIUM	4	93.75	No	No
Leclerc	Italian GP	5	119.85	475.29	2	MEDIUM	5	93.27	No	No
Leclerc	Italian GP	6	119.97	595.27	2	MEDIUM	6	92.79	No	No
Leclerc	Italian GP	7	86.132	682.22	2	MEDIUM	7	90.77	No	No
…	…	…	…	…	…	…	…	…	…	…
Leclerc	Italian GP	17	89.038	1553.4	3	MEDIUM	17	70.60	Yes	No
Leclerc	Italian GP	18	108.45	1661.8	11	HARD	1	68.58	No	Yes
Leclerc	Italian GP	19	85.920	1747.7	9	HARD	2	66.57	No	No
…	…	…	…	…	…	…	…	…	…	…
Leclerc	Italian GP	52	84.621	4589.9	2	HARD	14	0	No	No

Another useful output is the final race results overview presented in table 6.5. Beyond comparing overall performance and strategies, this tale reveals the broader impact of race events. The leader's total race time shows a 1.7% (+76 s) increase compared to uninterrupted runs, reflecting the effect of

the SC phase between laps 4 and 6. Additionally, Piastri is marked as retired (on lap 4), illustrating how simulated failures are captured and influence both race outcome and flow.

Table 6.5: Final race results for all the drivers, including several key metrics

Circuit	Driver	Position	Race Time [s]	Gap To Leader [s]	Gap To Driver Ahead [s]	Pit Stop Strategy	Used Compounds
Italian GP	Sainz	1	4587.1	[-]	[-]	2 Stops	M → H → H
Italian GP	Leclerc	2	4589.9	+2.9	+2.9	2 Stops	M → H → H
Italian GP	Verstappen	3	4597.9	+1.0	+7.9	2 Stops	M → H → H
Italian GP	Perez	4	4602.9	+10.8	+5.0	2 Stops	M → H → H
Italian GP	Hamilton	5	4613.2	+15.9	+10.3	2 Stops	M → H → H
Italian GP	Albon	6	4613.6	+26.1	+0.4	2 Stops	M → H → H
Italian GP	Russell	7	4615.9	+28.9	+2.4	2 Stops	M → H → H
Italian GP	Norris	8	4626.2	+39.1	+10.3	2 Stops	M → H → H
Italian GP	Alonso	9	4649.3	+62.2	+23.1	2 Stops	M → H → H
Italian GP	Tsunoda	10	4663.0	+76.0	+13.8	2 Stops	M → H → H
Italian GP	Lawson	11	4666.6	+79.5	+3.5	2 Stops	M → H → H
Italian GP	Hulkenberg	12	4670.0	+82.9	+3.5	2 Stops	M → H → H
Italian GP	Sargeant	13	4679.4	+1 Lap	+9.4	2 Stops	H → M → H
Italian GP	Bottas	14	4683.1	+1 Lap	+3.7	2 Stops	M → H → H
Italian GP	Zhou	15	4705.5	+1 Lap	+22.5	2 Stops	M → M → H
Italian GP	Gasly	16	4706.5	+1 Lap	+0.9	2 Stops	M → H → M
Italian GP	Magnussen	17	4711.4	+1 Lap	+4.9	2 Stops	M → H → H
Italian GP	Ocon	18	4712.8	+1 Lap	+1.4	2 Stops	M → H → H
Italian GP	Stroll	19	4722.5	+1 Lap	+9.7	2 Stops	M → H → H
Italian GP	Piastri	20		DNF			

To complement the tables, several diagrams are used to provide a visual interpretation of the race. The Race Position Evolution plot in Figure 6.8 illustrates how driver positions evolved lap by lap. Between laps 4 and 6, the effect of the SC phase is clearly visible, with position lines stabilizing and no overtakes occurring during this period, as overtaking is not permitted under neutralized conditions. The plot also highlights Piastri's retirement on lap 4: from that point onward, his position is fixed at the bottom of the chart, represented by a dashed line to indicate that he is no longer actively participating in the race. This visual summary provides a clear and intuitive overview of race flow, including strategy moments, retirements and the structural impact of the SC.

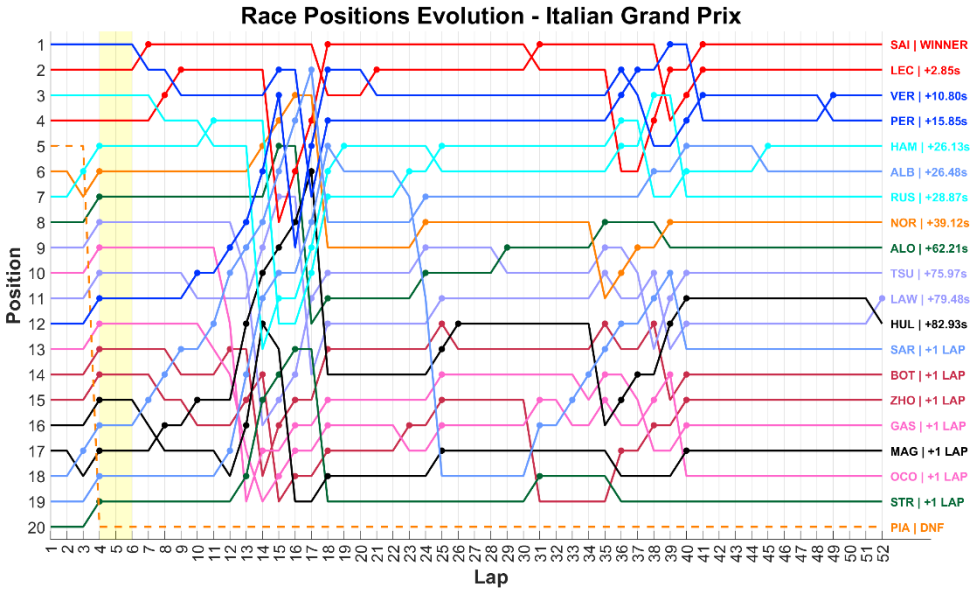

Figure 6.8: Lap-by-lap position changes across the whole race, featuring a SC period (yellow shade)

The Lap Times Race Progression plot in Figure 6.9 displays the evolution of each driver's lap times throughout the race. While it is typically used to observe tire degradation patterns, performance trends and the impact of pit stops, the plot also clearly reflects the dynamics of the SC phase. During laps 4 to 6, lap times increase significantly for all drivers, corresponding to the mandatory speed reduction imposed during the neutralized period. This behavior is consistent with the simulator's enforcement of SC constraints, where drivers slow down to follow the SC and maintain a controlled pace before race conditions resume.

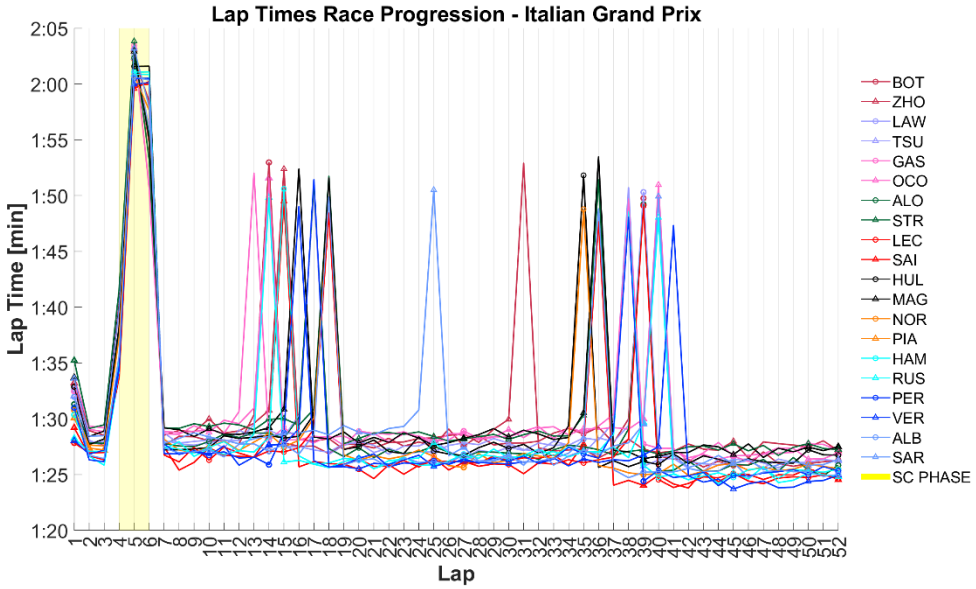

Figure 6.9: Lap times across the whole race for all the drivers, featuring a SC period (yellow shade)

Another helpful figure is the Cumulative Race Time Summary plot in Figure 6.10, which displays the total race time accumulated by each driver by the end of the simulation. This visualization allows for quick comparison of finishing times and relative performance across the field. Notably, the driver who did not finish the race, Piastri, is represented at the bottom of the plot, separated from the rest of the

classified drivers. His placement visually reinforces his early retirement and exclusion from final timing comparison, while still acknowledging his participation in the race event.

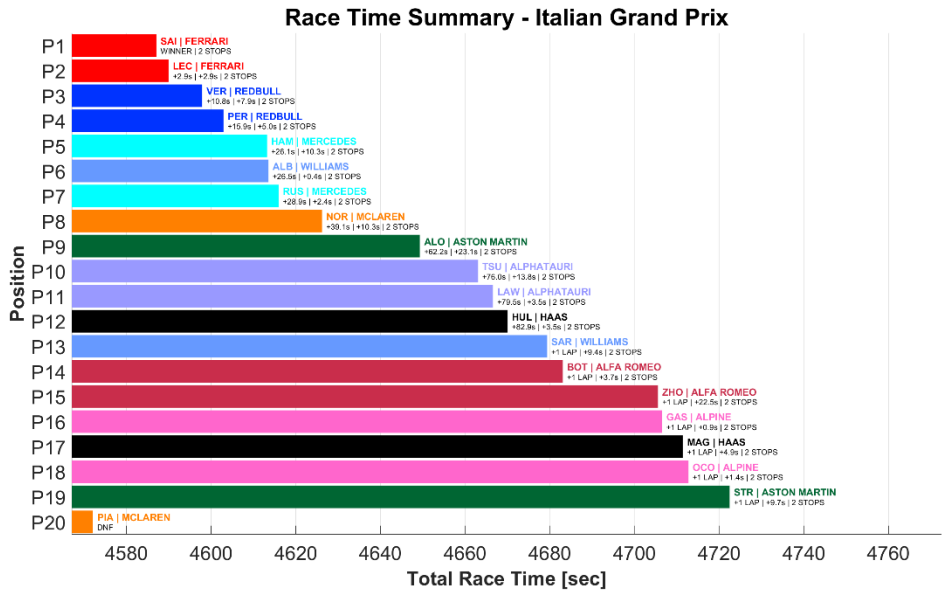

Figure 6.10: Final race summary for all the drivers regarding the time gaps to the leader and to the driver in front

The Race Strategy Summary plot in Figure 6.11 presents how each driver used different tire compounds over the race distance. Each color segment represents a different compound used during a stint (SOFT compound tire is red, MEDIUM compound tire is yellow, HARD compound tire is grey). This visual provides an immediate overview of when drivers made pit stops and how their compound choices evolved across the race. Piastri, who retired on lap 4, is placed at the bottom of the chart. His strategy appears truncated, reflecting that no further stints were executed beyond his retirement. This graphical treatment ensures consistency with his starting compound and initial race participation.

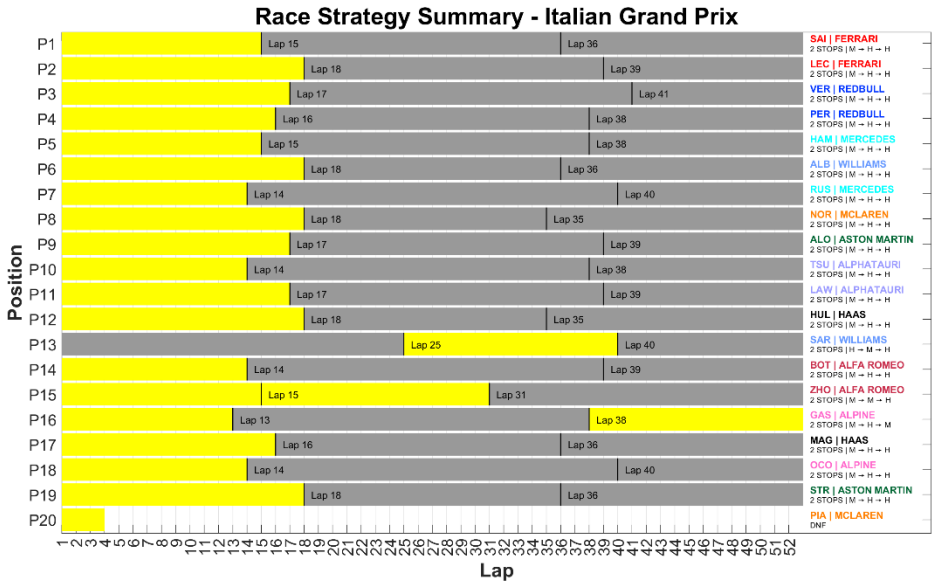

Figure 6.11: Race strategy summary across the whole race for all the drivers

In addition, overtaking activity is displayed through specific diagrams. The Overtaking Summary diagram (Figure 6.12) shows all overtaking attempts across the race and whether they were successful. The Lap-Wise On-Track Overtaking Summary plot (Figure 6.13) illustrates the number of overtaking

attempts and successful moves for each lap, providing a clear view of where most of the race action occurred. However, it is important to note that during laps 4, 5 and 6, when the SC phase is active, no overtaking is permitted. As a result, no overtaking logic is evaluated during this neutralized window and the corresponding section of the plot shows zero overtaking activity, in line with regulatory constraints.

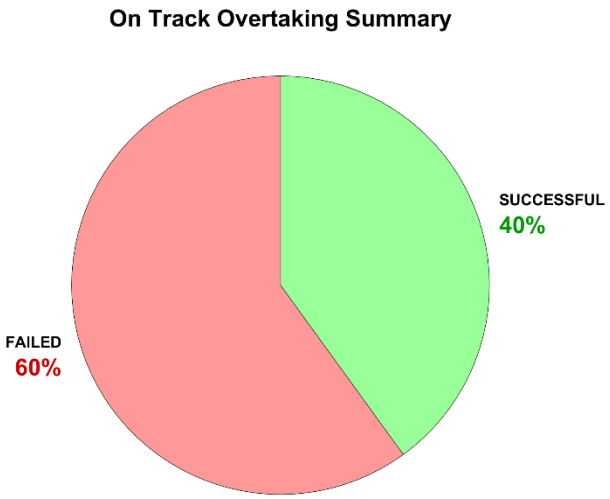

Figure 6.12: On track overtaking summary collecting the successful overtakes against the failed ones

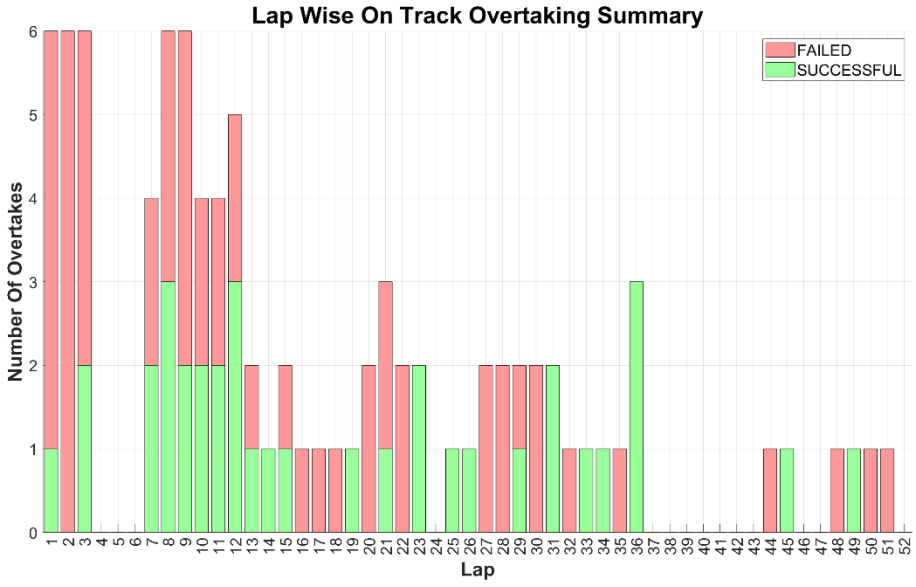

Figure 6.13: Lap-by-lap on track overtaking summary comparing the successful overtakes and the failed ones

In conclusion, the described process applies to any simulation scenario in which a FCY phase is introduced. The simulator is capable of handling multiple FCY phases during the same race, whether triggered by an accident (initiating a SC) or by a mechanical failure (leading to a VSC). Each time a driver retires, they are immediately placed at the bottom of the classification relative to the remaining participants, according to the lap of retirement. Concurrently, the lap times of all active drivers are adjusted to comply with the pace restrictions imposed by the FCY regulations, during which no overtaking is permitted. As a result, the total race time increases significantly, reflecting the

neutralization of race pace and the regrouping of the field. This mechanism ensures that the simulator realistically replicates the effects of safety interruptions on the race outcome, maintaining both strategic coherence and competitive integrity.

7 Monte Carlo Method: Simulator Application and Results

As detailed in the previous chapters, this thesis has developed a race simulation framework capable of reproducing the progress of a Formula One race lap by lap. The earlier sections explained how key phenomena, including tire degradation, fuel mass reduction, pit stop timing, FCY phases and overtakes, are modelled through deterministic equation and rules. However, as highlighted throughout those chapters, even a single race simulation run in this framework is not purely deterministic. Several parameters and mechanisms within the simulation run are inherently stochastic, such as lap time variability due to driver performance fluctuations, random variations in pit stop durations or the probabilistic success of overtaking attempts. These stochastic components introduce randomness into each individual simulation and make every run unique, even if initial conditions remain the same.

Nevertheless, a single stochastic simulation still represents only one possible realization of how a race might unfold. In reality, Formula One races are shaped not only by planned strategies and technical parameters but also by the immense variability of random events, like accidents, mechanical failures, fluctuations in driver performance and the unpredictable timing and frequency of race incidents.

To fully capture this inherent uncertainty, this work extends the simulation framework using a Monte Carlo Simulation (MCS) approach, building directly upon the deterministic and stochastic modelling described in the earlier chapters. The Monte Carlo Method is a numerical technique based in random sampling, used to approximate solutions for complex or analytically intractable problems. As described by Chen [30], this method has been widely adopted since the 1940s across disciplines such as physics, engineering, finance and computer science, precisely because of its power in dealing with problems where exact solutions are difficult or impossible to obtain analytically.

The fundamental idea of Monte Carlo Methods lies in probability theory, specifically the Law of Large Numbers and the Central Limit Theorem [30]. These principles ensure that, as the number of simulation trials increases, the average result converges towards the true expected behavior of the system. The general procedure involves defining the problem and its uncertain variables, generating large numbers of random samples for these variables, computing the outcome for each set of random inputs and finally analyzing the statistical properties of all results, such as averages, variances and probability distributions. Unlike traditional numerical methods, the computational cost of MCS depends primarily on the number of samples rather than the problem's dimensionality [30]. This makes them exceptionally effective for tackling high-dimensional systems or irregular problem domains that are difficult to handle with classical techniques.

As emphasized by Heilmeier, one of the primary advantages of MCS in motorsport is the possibility to evaluate the robustness of pre-race strategies under varying stochastic conditions, allowing teams to move beyond single deterministic predictions and assess the range and likelihood of potential outcomes [4]. Importantly, each simulated race run in the MCS is independent of the others, enabling straightforward parallelization to exploit multi-core computational resources effectively, which is a significant benefit given the complexity and runtime of modern race simulations.

Moreover, the aggregation of thousands of simulations provides distributions of possible race outcomes rather than a single deterministic forecast. This probabilistic insight allows strategists to quantify risks associated with various strategic options, to estimate probabilities of achieving certain race positions and to understand how random events might influence the success or failure of specific strategies. Such insights are invaluable not only for pre-race planning but also for making rapid, data-informed decisions in real time during a race [2].

To ensure that the insights derived from this Monte Carlo approach are statistically robust, this thesis performs the race simulation 10000 times. The choice of 10000 simulations reflects a balance between computational feasibility and the reliability of the results. As highlighted by Heimeier, the Law of

Large Number ensures that the more often a random experiment is repeated, the closer the average result will approach the true expected value of the system [4]. In his work, the deviation between batches of simulation runs diminishes noticeably as the number of runs per batch increases, as it can be seen in Figure 7.1. In most cases, 10000 simulations represent an effective compromise, delivering stable statistical indicators while remaining manageable in terms of computation time. However, it is crucial to emphasize that the reliability of MCS outcomes also depends on the accurate parametrization of the race simulation model itself.

Simulation Runs Per Batch	Mean Position	Deviation (95%)
100	2.115	±0.309
1,000	2.126	±0.082
10,000	2.107	±0.029
100,000	2.107	±0.011

Figure 7.1: Mean rank positions and deviations (95 % confidence) for Hamilton inside Heilmeier's work [4]

In the following sections, the results of the MCS are explored in two parts. The first one provides an overview of race-level statistics, such as the frequency of SC periods, gaps between drivers, overtaking figures and variations in race time for the leader across different race scenarios. The second one focuses on driver-specific analysis, comparing race strategies and examining how these influence finishing positions.

In summary, MCS enables a far deeper understanding of how race strategies might perform under the myriad uncertainties inherent to motorsport, providing critical insights for both pre-race planning and real-time decision-making.

7.1 Race Craft Analysis

This section presents a comprehensive analysis of race craft and leader-centric metrics across the 10000 simulations conducted for this thesis. The goal is to evaluate how realistically the simulation replicates the dynamics of a real Formula One race and to create a solid foundation for comparison with actual race results. Each result presented here is a window into how random factors, driver performance and race events shape competitive narrative of a race.

7.1.1 MCS Analysis of Leader's Median Cumulative Race Time

A fundamental starting point for the analysis is the examination of the median cumulative race time of the leading driver across all simulations. This metric is first computed as an overall median and standard deviation taken across the entire set of 10000 simulations, without distinguishing whether each individual simulation did or did not experience SC phases. As a result, this median value provides a central estimate of race duration as it emerges from the full range of race scenarios generated by the stochastic model, capturing both race conditions and the variability introduced by potential race interruptions and other unpredictable events. In this case, the MCS produces a median race time for the winning driver at the Italian Grand Prix of 4427.162 ± 65.203 s. It is important to note that the real 2023 Italian Grand Prix was shortened to 51 laps due to several interruptions during the formation lap, which slightly reduced the typical total race distance and consequently the overall race time. The simulation has been, therefore, adjusted according to this modification to capture the same race distance as the real one. For comparison, the actual winning time recorded by Max Verstappen in the 2023 Italian Grand Prix was 4421.143 s, according to official Formula One data [31]. According to this result, the leader's race time of the simulation is only 6.019 s longer than the real one, corresponding to a very modest relative error of 0.14%. Such a close agreement suggests that the simulation is highly effective in replicating realistic race durations, even when accounting for the complexities and stochastic nature of race events. A clear visual comparison is presented in Table 7.1.

Table 7.1: Representation of simulated median race time and standard deviation over 10000 simulations at the Italian Grand Prix relative to the winners, compared to the actual race time of the 2023 season winner at the same circuit. Also the relative error between the simulated and the actual race time value is represented

Winner Simulated Race Time [s]	Simulated Standard Deviation [s]	Winner Actual Race Time [s]	Relative Error from Actual Value [%]
4427.162	65.203	4421.143	+0.14 %

7.1.2 MCS Analysis of FCY Related Median Cumulative Race Time

To gain a clearer picture of how neutralization specifically affect race time, a second analysis is performed in which the simulations are categorized based on the number of SC phases observed: zero, one, two, three or more. For each of these categories, the median and standard deviation of leader's cumulative race time are calculated separately. This distinction allows to quantify how the occurrence of neutralization typically shifts the race duration and how variable race times become under differing race conditions. By comparing these detailed simulated statistics with real race timing data, it becomes possible to evaluate whether the simulation realistically reproduces both the mean effects and variability introduced by neutralized phases.

The results demonstrate a trivial trend. Races with no SC phases have the shortest median race time, measured at 4425.042 ± 11.845 s, reflecting a smooth, uninterrupted race under pure racing conditions. When one SC occurs, the race time increases to 4521.134 ± 24.730 s, suggesting that even a single neutralization event can add approximately 96 seconds to the overall race duration due to slower laps, potential pit stops taken under SC or subsequent race restarts. In case of two SC phases, the race time extends further to 4607.995 ± 29.944 s, while three or more SC phases record the longest duration at 4701.837 ± 29.926 s, demonstrating the significant impact of repeated neutralizations on overall race length.

It is important to note that these medians conceal significant underlying variability driven by the stochastic duration of each SC period. In real races, as well as in the simulations, the length of a SC phase can fluctuate considerably, typically lasting anywhere from 2 to 8 laps depending on the nature and severity of the incident that triggers it, as modelled in previous chapters. A short neutralization might only minimally disrupt race dynamics, while a prolonged SC period can substantially extend the race duration and compress the gaps between drivers, dramatically influencing strategy and final results. Consequently, while the medians give valuable insight into general trends, the true impact of FCY phases on race time is inherently probabilistic and must be interpreted within the context of this variability.

These findings are critical because they show how sensitive the total race time is to the frequency of FCY appearances and they align with the intuitive understanding that more interruptions lead to longer overall race durations. Table 7.2 summarizes these results, presenting the median and standard deviation of the leader's race time for each SC appearance. This table serves as a reference point for validating and assessing how well the model captures the time-altering effects of neutralized phases.

Table 7.2: Median cumulative race time of the leader across simulations, classified by the number of SC phases, with incremental differences between classes

SC Appearances	Median Race Time [s]	Standard Deviation [s]	Margin to Previous [s]	Percentage Increase [%]
0	4425.042	11.845	-	-
1	4521.134	24.730	+96.092	+2.17 %
2	4607.995	29.944	+86.861	+1.92 %
≥ 3	4701.837	29.926	+93.842	+2.04 %

Building upon this, the analysis further explores the distribution of the cumulative race time of the leader across all simulations, separated by the number of SC phases that occurred during each race. The overlapping histograms represented in Figure 7.2 illustrates how the presence or absence of SC periods influences the likely duration of the race. Such insights are invaluable because they highlight the stochastic impact of race incidents on the leader's finishing time. For instance, as explained before, races with two or more SC deployments tend to finish significantly slower than clean races. This plot exploits the comparison with real-world race times, ensuring the model not only replicates average behaviors but also captures the range of possible outcomes under different race scenarios.

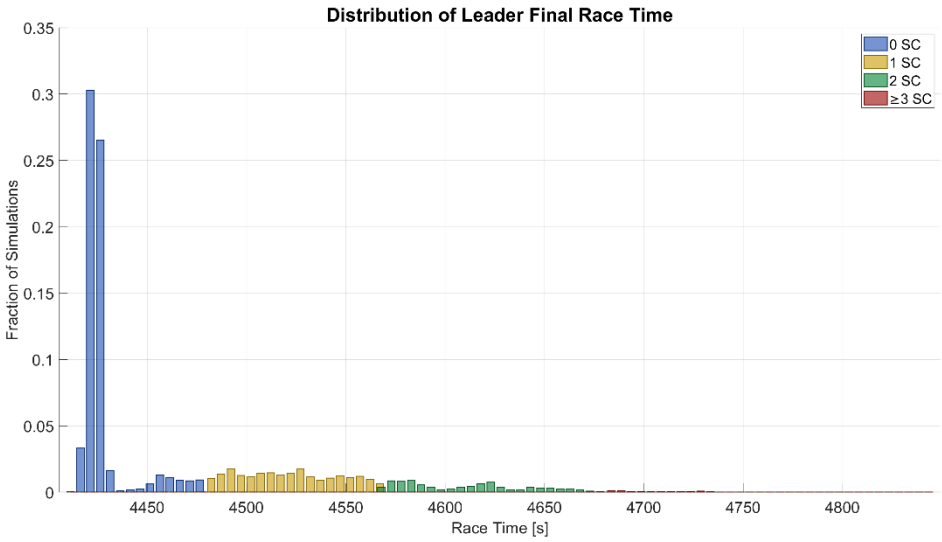

Figure 7.2: Distribution of the leaders' final race time across 10000 simulations on Monte Carlo approach, classified by the number of SC appearances. Each color represents races with a different count of SC phases (0, 1, 2, ≥3). The plot highlights the increasing spread and upward shift in race time as the number of SC deployment grows, reflecting the time impact of race neutralizations on overall performance

Closely connected to race duration is the analysis of the frequency of SC phases across the simulated races. Reporting these frequencies serves both as validation check against real race statistics and as essential context for understanding the variability in simulated race outcomes. Table 7.3 reports the appearances of SCs across simulations and compares it with the historical percentage registered by Heilmeier [4] and previously reported in Table 5.1.

Table 7.3: Comparison between simulated and historical frequencies of SC phases across 10,000 simulations

SC Appearances	Simulated Counts	Simulated Probability [%]	Historical Probability [%]	Relative Error [%]
0	6770	67.70 %	45.50 %	+48.79 %
1	2308	23.08 %	41.30 %	-44.10 %
2	856	8.56 %	9.90 %	-13.54 %
≥ 3	66	0.66 %	3.30 %	-80.00 %

The comparison reveals notable discrepancies between the simulation outputs and historical data regarding SC occurrences. The simulations predict a significantly higher proportion of races without any SC phases (67.70 %) compared to the historical rate (45.50 %). Conversely, the occurrence of one or more SC phases is markedly underestimated in the simulations, severely underpredicting the appearance of 3 or more SC periods (simulated 0.66 % compared to historical 3.30 %), resulting in a 80.00 % error. These differences suggest that the simulation, in its current parametrization, tends to model races as too clean and uninterrupted, underestimating the frequency of events that lead to neutralization. It should be noted, however, that the historical data on SC frequencies are computed over a span of five Formula One seasons across multiple circuits [4], whereas the simulated results in this work are derived from a single circuit simulation. This inherent limitation means that circuit-specific characteristics, such as track layout, run-off areas and historical incident rates, are not fully

reflected in the current simulation's frequency estimates. Furthermore, it is possible that the sample size of 10000 simulations, while substantial, may still be insufficient to fully capture the rarer events and match the precise historical percentages observed in real races. In the simulation logic implemented in the previous chapters, every accident directly triggers a SC period. Therefore, the lower frequency of SC phases observed in the simulations might also suggest that the model is generating too few accidents overall, possibly due to conservative assumptions in the probability settings for race incidents. These factors together may explain why the simulation shows fewer race disruptions than what actually happens in real races.

7.1.3 MCS Analysis of Position Related Mean Final Race Gap to Leader

An equally important aspect of understanding race dynamics is the examination of mean time gaps between the leader and each other classified driver. For every simulated race, the finishing time of the leader is subtracted from the times of all other drivers and these gaps are averaged over the entire Monte Carlo dataset. The results are presented in both tabular and graphical formats, offering a detailed perspective on the typical spread of the field. Such analysis is crucial for identifying whether certain time gaps could suggest dominance by particular drivers or cars or a performance drop-off beyond certain places in the order. For instance, if the simulation consistently shows a significant time jump between several positions, it could point to a natural split between the top teams and the midfield. This allows for direct comparison with real race data to assess whether the model accurately reflects the competitive landscape for the sport.

The analysis is complemented by Table 7.4 and Figure 7.3, which report the average time gaps between the leader and each finishing position across all the 10000 simulations. The gap grows steadily moving down the field and, notably, there's a significant jump in the gaps starting from around the fifth position onwards. For example, the gap between fifth and sixth is smaller (about 3 seconds), but then gaps gradually widen, reaching over 63 seconds for the tenth-place driver and exceeding 142 seconds for the last-place finisher.

This pattern suggests that the simulation captures a clear spread in performance between the top cars and the midfield, as well as between the midfield and the back markers. In real races, similar gaps are often seen, especially in circuits where overtaking is difficult or where race pace differences between teams are pronounced. However, the steadily increasing gaps might also indicate that the simulation slightly exaggerates how quickly time differences accumulate down the field, particularly in the lower positions where the gaps surpass a minute. Comparing these simulated values to real race results helps verify whether the model realistically portrays the competitive landscape in Formula One or whether it may be overestimating performance disparities among teams.

Table 7.4: Simulated mean time gaps between the leader and each classified position across 10,000 race simulations at the Italian Grand Prix

Position	Mean Gap to Leader [s]
1	-
2	3.859
3	8.231
4	18.150
5	21.843
6	25.108
7	28.417
8	32.882
9	42.829
10	62.948
11	72.802
12	79.835
13	84.694
14	90.003
15	99.993
16	118.685
17	124.711
18	129.438
19	136.795
20	142.773

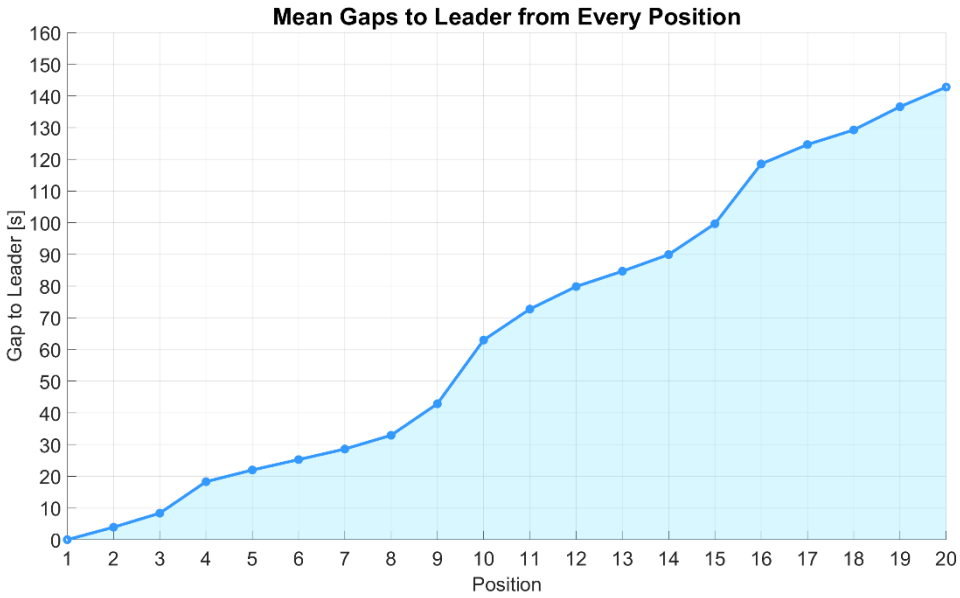

Figure 7.3: Simulated mean time gaps to the race leader for each classified finishing position at the Italian Grand Prix, computed across 10000 simulations of Monte Carlo approach. The plot shows how the average gap increases progressively through the field, reflecting typical performance differentials between the front runners, midfield and back markers in simulated race scenarios

Alongside the analysis of gaps between each finishing position and the leader, a specific focus is placed on the time gap between the leader and the last classified driver. To capture how much of an advantage the leader tends to build over the slowest car in the field, the simulation calculated this gap for each of the 10000 simulations, creating a statistical distribution that reflects the range of possible outcomes.

This distribution is illustrated in Figure 7.4, where the horizontal axis shows the size of the time gap in seconds and the vertical axis indicates how often gaps of various sizes occur across all simulations. The figure highlights that, in the majority of cases, the gap between the leader and the last classified driver falls within a relatively narrow range, clustering around a central value. Specifically, the simulation reports a mean gap between the leader and the last finisher of 142.503 ± 9.377 s. In the real 2023 Italian Grand Prix, Verstappen finished first and Magnussen was the last classified driver in P18, officially listed as +1 lap behind. However, according to detailed post-race analysis, the actual time gap between them was 100.149 s [32], even though Magnussen was one lap down in the final classification (comparison summary can be seen in Table 7.5). Comparing this real-world value to the simulation suggests that the model tends to slightly overestimate the gap to back markers, since the simulated mean gap of 142.773 s corresponds to about +1.6 laps, given the base lap time of roughly 88 seconds at Monza. Nevertheless, while the simulator predicts a larger spread than occurred in this particular race, it remains a valuable tool for capturing the general range of possible outcomes, especially considering the gaps can vary substantially depending on race circumstances.

Table 7.5: Comparison between the simulated mean race time gap from leader to the last classified driver and the actual gap observed between Verstappen and Magnussen (last place) in the 2023 Italian Grand Prix

Type	First-to-Last Gap [s]	Difference [s]
Actual	100.149	-
Simulated	142.503	+42.354

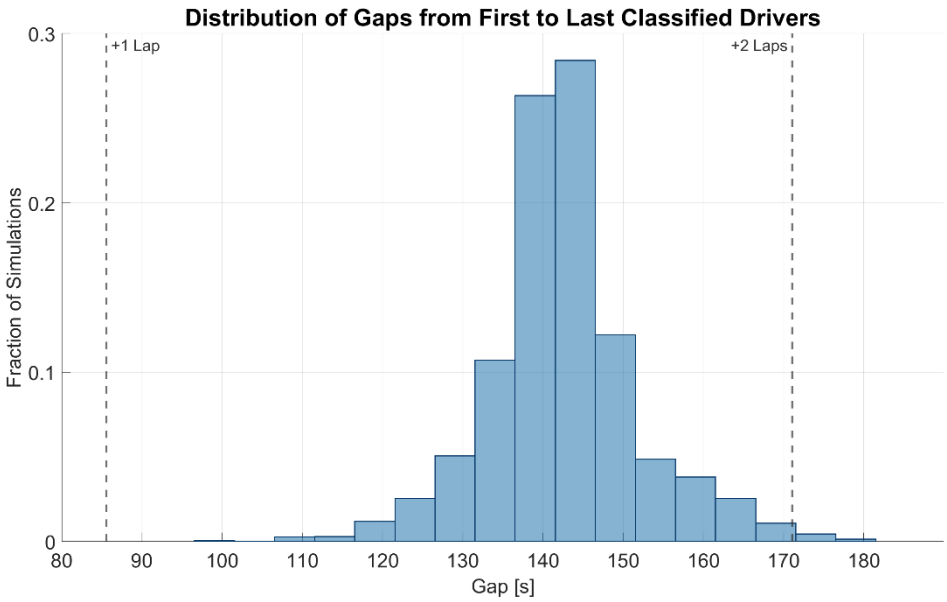

Figure 7.4:Distribution of simulated total race time gaps between the first and last classified drivers at the Italian Grand Prix, computed across 10000 simulations of Monte Carlo approach. Vertical dashed lines indicate the time thresholds corresponding to +1 lap and +2 laps relative to the leader's base lap time. The histogram shows that the most frequent gap between the leader and the last classified driver centers around approximately +1.6 laps, illustrating the likelihood of lapped traffic under simulated conditions

This analysis is significant because it reveals the extent of the leader's dominance over the slowest competitor. In real Formula One events, some races conclude with a tightly packed field, while others see the leader finish one or even two laps ahead of the back marker. By comparing this simulated distribution to actual race data, it is possible to assess whether the model realistically reflects how large, or small, these differences typically are. If the simulated distribution is too narrow, it might underestimate the variability of real race outcomes, whereas a distribution that is too wide could suggest the model is overestimating differences in car performance.

7.1.4 MCS Analysis on Overtaking Statistics

Finally, the analysis turns to overtaking statistics, which are critical for evaluating the realism of racing interactions in the simulation. Across the 10000 simulated races, the number of overtaking attempts and the number of successful overtakes are tracked and averaged to produce mean values. These figures are compared to historical race data to gauge the simulation's accuracy in reproducing on-track battles. Overtaking is central to Formula One's spectacle and significantly influences strategic decision-making, as races with frequent overtakes are more dynamic and less predictable.

Specifically, MCS reports an average of 111.49 overtaking attempts per race, resulting in an average of 38.39 successful overtakes. These figures are compared to the historical data from the 2023 season Italian Grand Prix, where the official count of real overtakes was 31 [17]. A summary report is illustrated in Table 7.6. This comparison reveals that the simulation predicts a slightly higher number of successful overtakes than was observed in the real event, suggesting that the model may slightly overestimate overtaking opportunities under the race conditions typical of Monza circuit. Such differences could arise from circuit-specific factors not fully captured by the empirical overtaking thresholds, from the inherent variability introduced in the simulation's overtaking model or from the possibility that the simulator generates an excessive difference in pace between drivers, making overtakes easier to perform than they might be in reality.

Table 7.6: Comparison of overtaking statistics between the MCS results and historical data from the 2023 Italian Grand Prix. The table reports the mean number of overtaking attempts and successful overtakes across 10000 simulated races at Monza. Relative error is calculated for successful overtakes as the percentage difference between the simulated and the actual number of overtakes observed during the race

Simulated Attempts	Simulated Overtakes	Actual Overtakes	Relative Error [%]
111.49	38.39	31	+23.84 %

If the simulation produces consistently higher or lower overtaking figures than real races, it might signal that certain parameters, such as the probability of successful overtakes or the impact of DRS and tire wear, need recalibration. Thus, overtaking statistics serves both as a direct measure of race craft realism and as a diagnostic tool for refining the model.

Taken together, the analyses in this chapter construct a detailed picture of how races unfold within the MCS framework. They allow for rigorous comparison with real race data, providing critical validation for the model and ensuring its practical utility for strategic analysis in Formula One racing.

7.1.5 MCS Analysis of Selected Drivers' Final Race Times

As a further step in validating the simulation framework, this section presents a targeted comparison of the simulated final race times for selected drivers against the actual times recorded in the 2023 Italian Grand Prix. Unlike previous comparisons focusing on the race leader, this analysis extends the validation of the model to other competitors, aiming to confirm that the simulation realistically predicts a plausible race time for each individual driver. By directly comparing the simulated times with the actual results, the analysis assesses how closely the simulator replicates real race performances across different drivers.

Three drivers have been selected to perform the analysis: Charles Leclerc, George Russell and Lando Norris. Their respective race times at the end of the 2023 Italian Grand Prix are 4432.520 s for Leclerc, 4444.171 s for Russell and 44466.592 s for Norris [31]. It is important to note that the race was shortened to 51 laps due to issues during the formation lap and that no SC deployments occurred during the entire race. Therefore, the simulated race times of these drivers have been filtered in order to include only those scenarios without any SC deployments, ensuring a fair comparison with the actual race conditions.

In the MCS runs performed for this study, the simulator produces median race times that remain close to their real-world counterparts. Specifically, the median simulated time for Leclerc is 4439.943 ± 9.852 s, while Russell's simulated time is 4453.375 ± 10.427 s and Norris's result stands at 4473.325 ± 11.298 s. These simulated results exceed the actual race times by only modest margin, corresponding to relative errors of just 0.168 %, 0.207 % and 0.151 % respectively. These details are summarized in Table 7.7.

These small discrepancies indicate that the simulator slightly overestimates the total race times for these drivers. However, the close alignment between the simulated and the actual results demonstrates that the model effectively reproduces not only the absolute performance of the race leader but also the relative race dynamics among other competitors, with only minor deviations. This reinforces confidence in the simulator's predictive capability.

Importantly, this analysis also confirms that the relative gaps between drivers are realistically represented within the simulation model. While absolute times are slightly longer, the proportional differences closely mirror those observed in the real race, providing strong evidence of the simulator's

value as a tool for evaluating not just individual race outcomes but also competitive relationships within the field.

Table 7.7: Comparison between actual race times from the 2023 Italian Grand Prix and median simulated race times obtained from MCS for selected drivers. The table shows the absolute difference between simulated and actual race times and the corresponding relative error, indicating how closely the simulation reflects real-world performance.

Driver	**Actual Race Time [s]**	**Median Race Time [s]**	**Standard Deviation [s]**	**Difference [s]**	**Relative Error [%]**
Leclerc	4432.520	4439.943	9.852	+7.423	+0.168 %
Russell	4444.171	4453.375	10.427	+9.204	+0.207 %
Norris	4466.592	4473.325	11.298	+6.733	+0.151 %

7.2 Single Driver Strategy Analysis

The previous analyses of MCS results have shown how race variability and stochastic events can significantly influence overall race outcomes. In particular, the distributions of leader race times and the spread of gaps across the field have highlighted that even small differences in race circumstances can lead to substantial variability in race duration and finishing order. This underscores the reality that a single deterministic prediction cannot capture the full complexity of a Formula One race.

Building upon this, it becomes essential to examine how different race strategies affect the performance of an individual driver, particularly in the context of the variability captured by the MCS framework. Strategy decisions, such as tire selection and pit stop timing, play a critical role in determining not only the absolute race time but also the likelihood of achieving a strong finishing position amidst the uncertainty of racing events. While a particular strategy might appear optimal under ideal circumstances, it may expose the driver to greater risks from tire degradation, SC interventions or time lost in traffic, potentially leading to poorer race results overall.

To explore this dimension, the present section focuses on the strategic evaluation of a single driver, simulating how different combinations of tire compounds and pit stops influence both race time and race classification across a wide range of simulated scenarios. The analysis leverages the MCS approach, conducting 10000 race simulations for each strategy to assess not only the expected performance but also the statistical variability and robustness of different tactical choices.

In this study, six distinct strategies were examined, covering both one-stop and two-stop configurations. In all strategies, the first compound type denotes the starting tire used for the opening stint of the race. Table 7.8 summarizes the strategies analyzed:

Table 7.8: Overview of the six pit stop strategies evaluated for a single driver in the Italian Grand Prix simulations. Each strategy indicates the number of pit stops and the sequence of tire compounds used, with the first compound representing the starting tire for the race

Strategy	Number of Stops	Tire Sequence
A	1	MEDIUM → HARD
B	1	HARD → MEDIUM
C	2	MEDIUM → HARD → HARD
D	2	MEDIUM → HARD → MEDIUM
E	2	HARD → MEDIUM → HARD
F	2	MEDIUM → HARD → SOFT

These strategies were chosen to reflect realistic race possibilities often observed at the Italian Grand Prix in Monza, a circuit characterized by long straights, high speeds ad relatively low tire degradation compared to other tracks. The inclusion of strategy F, which ends on the SOFT compound, was specifically designed to test whether a late switch to a softer tire could provide a decisive performance advantage in the final phase of the race.

The following sections present two different perspectives on strategy evaluation, focusing on both identifying the strategy yielding the fastest average race time and examining which strategy delivers the best mean finishing position, reflecting strategic consistency and reliability under the variable conditions captured in the simulations.

7.2.1 Fastest Strategy

The first aspect of this analysis addresses the purely time-focused perspective, seeking to determine which of the six strategies results in the lowest average race time across the 10000 simulations of the MCS approach. Figure 7.5 presents the race time distributions for each strategy, while Table 7.9 summarizes the mean race times achieved. The driver of choice for the various simulations is Max Verstappen.

The simulations revealed that among all tested configurations, strategy A (MEDIUM → HARD, 1 stop) emerged as the fastest on average, achieving a mean race time of approximately 75.91 minutes. This outcome reflects the nature of the Monza circuit, where high speeds and log straights make pit stops particularly costly in terms of time lost relative to maintaining track position. Consequently, a one-stop approach minimizes time spent in the pit lane and proves advantageous on such a fast track.

Interestingly, the comparison between two-stop strategies that conclude either MEDIUM or SOFT tires shows minimal differences in overall time. The data indicate that finishing on the SOFT compound does not provide a decisive advantage over ending the race on MEDIUM tires. This result likely stems from the significant degradation suffered by the SOFT tires in the latter stages of the race, preventing them from fully exploiting their higher performance potential. Meanwhile, the MEDIUM tire appears to strike a better balance between performance and durability, effectively compensating for the theoretical pace advantage of the softer compound over a long stint.

Among the one-stop strategies, starting on HARD tires and witching to MEDIUM, as seen in strategy B, proved less effective. This may be attributed to the combination of the car's high fuel load at the race start and the inherent limitations of the HARD compound, which together reduce the ability to extract competitive lap times in the early phases of the race. The HARD tires, while durable, seem less able to generate the necessary grip under heavy fuel conditions, impacting the overall pace.

The slowest strategy observed in the simulations was strategy C (MEDIUM → HARD → HARD, 2 stop). This approach appears to underperform partly because it relies entirely on the two hardest compounds, which, although more durable, may lack the necessary grip and thermal performance to deliver competitive lap times at Monza. Running both stints on HARD tires seems to diminish the car's potential, especially given the relatively low tire degradation characteristic of the Italian Grand Prix, which otherwise allows softer compounds to be effectively utilized.

It is important to note, however, that while differences exist among various strategies, these differences are relatively small, since approximately 0.1 min separates the fastest and slowest strategies. This limited spread likely reflects the competitive strength of Verstappen's car, which operates at the front of the field where overtaking variations produces only modest differences in total race time for a leading driver of this caliber.

The comparison between the simulation results and the real race data highlights a significant validation of the model's effectiveness. Notably, the simulation correctly identified strategy A (MEDIUM → HARD, 1 stop) as the fastest option, which aligns precisely with the actual strategy adopted by Verstappen in the 2023 Italian Grand Prix. This correspondence underscores the ability of the simulation to capture the strategic dynamics of a real race. However, a closer examination of race times reveals that the mean simulated race time for strategy A is 75.91 min, whereas Verstappen's actual race time is 73.69 min [31], resulting in a relative error of approximately 3.02 % (result summary can be seen in Table 7.10). This modest difference suggests that while the simulation reliably predicts the optimal strategy choice, it tends to be slightly conservative in estimating the total race time for a leading driver and car combination. The small overestimation may arise from simplified assumptions within the model or from unique real-world factors such as precise driving performance, specific track conditions or race management decisions that are challenging to replicate

fully in simulation. Nonetheless, the close alignment in strategy ranking and overall timing demonstrates that the simulation framework provides valuable and credible insights for race planning and strategic analysis.

Table 7.9: Mean race times for each strategy applied to Max Verstappen computed across 10000 simulations of MCS approach for the Italian Grand Prix, sorted from fastest to slowest

Strategy	Mean Race Time [min]
A	75.91
D	75.92
F	75.94
E	75.95
B	75.97
C	75.99

Table 7.10: Comparison between the mean simulated race times for the fastest simulated strategy and Verstappen's actual race strategy during the 2023 Italian Grand Prix. The relative error quantifies how closely the simulation matches real-world performance

Strategy	Type	Mean Race Time [min]	Relative Error [%]
A	Actual	73.69	-
A	Simulated	75.91	+3.02 %

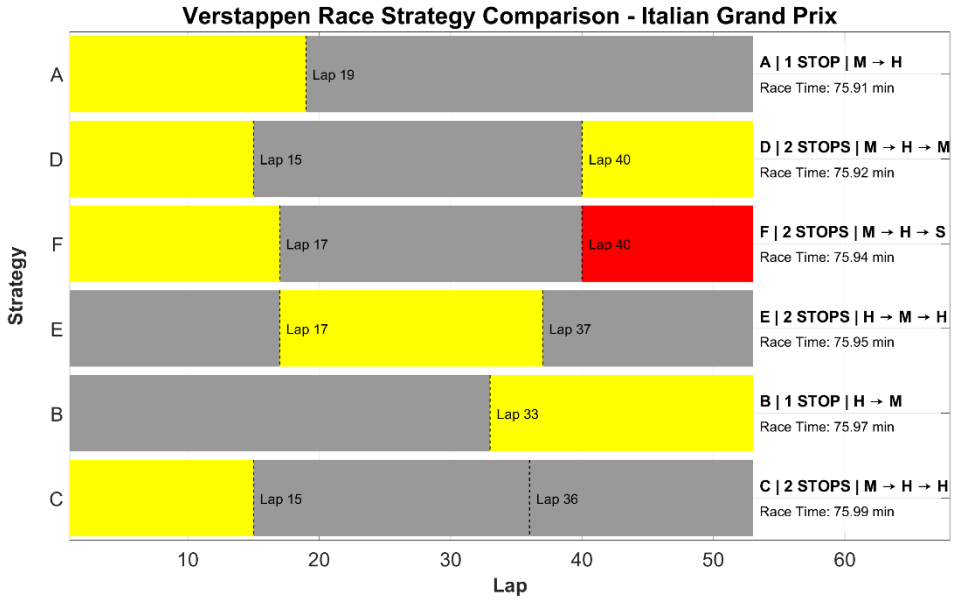

Figure 7.5: Simulated average race times for six different strategies applied to Max Verstappen across 10000 simulations of MCS approach for the Italian Grand Prix, sorted from fastest to slowest

7.2.2 Best Strategy for Mean Finishing Position

While the pursuit of the fastest possible race time is undoubtedly appealing, success in Formula One often hinges on securing the best possible finishing position across a wide range of race conditions. To address this dimension, the next analysis focuses on identifying the strategy that achieves the best mean finishing position across the simulated races.

Figure 7.6 illustrates the distribution of finishing positions for the driver under each of the six strategies, while Table 7.11 summarizes the average finishing position achieved for each approach.

The analysis of finishing positions across all simulated strategies reveals that there is relatively little variation in the mean finishing position achieved by each strategic approach. The distributions shown in Figure 7.6 demonstrate that, regardless of tire sequence or number of pit stops, Verstappen consistently finishes within the top positions, reflecting the strong baseline performance of his car and the reduced influence of overtaking challenges for a front-running driver.

Among the strategies, strategy A (MEDIUM → HARD, 1 stop) emerges as the one providing the slightly best average finishing position, confirming its effectiveness not only in terms of race time but also in securing reliable race outcomes. The combination of a single pit stop and the optimal exploitation of the performance of both MEDIUM and HARD compound tires appears to offer an optimal compromise between pace and race consistency, while minimizing the time loss associated with pit stops on a high-speed circuit like Monza.

Although other strategies, including two-stop approaches such as strategy E (HARD → MEDIUM → HARD, 2 stops), demonstrate solid performance with finishing positions close to those of strategy A, they do not yield a substantial advantage in terms of final classification. The analysis indicates that ending the race on softer compounds, as in strategy F (MEDIUM → HARD → SOFT, 2 stops), does not translate into better average positions, likely due to both the higher degradation of the SOFT tires in the latter stages of the race, which negates their potential performance benefit, and the trivial time loss relative to the occurrence of the second stop.

Overall, while differences exist, the spread in mean finishing positions across all strategies is minimal, typically varying by only a few tenths of position. This narrow range underscores the competitive strength of Verstappen's car and highlights that, for a leading driver, strategic choices tend to produce broadly similar outcomes under stable race conditions. Nonetheless, strategy A demonstrates a slight edge in delivering the most consistent and favorable race results within the variability captured by the MCS approach.

The real race saw Verstappen achieving first place [31], while the simulations estimate a mean finishing position of approximately 1.83 relative to the fastest strategy, strategy A (MEDIUM → HARD, 1 stop), across 10000 simulated races. This slight difference suggests that the model slightly underestimates the dominance of a top-performing driver and car combination like Verstappen's, possibly due to the inherent variability included in the simulation or real-world dynamics that are difficult to predict in the simulator. Nevertheless, the fact that strategy A emerges as the leading choice in both simulation and reality confirms the model's effectiveness in predicting strategic success, while the small gap in average finishing position underscores the challenges of perfectly replicating the exceptional consistency and race craft displayed by top drivers in real-world conditions (result summary can be seen in Table 7.12).

Table 7.11: Mean finishing positions for each strategy applied to Max Verstappen across the 10000 simulations of MCS approach for the Italian Grand Prix, sorted from best to worst

Strategy	Mean Finishing Position
A	1.83
C	1.84
D	1.85
B	1.88
E	1.88
F	1.88

Table 7.12: Comparison between the mean simulated finishing position and Verstappen's actual finishing result during the 2023 Italian Grand Prix

Strategy	Type	Mean Finishing Position
A	Actual	1
A	Simulated	1.83

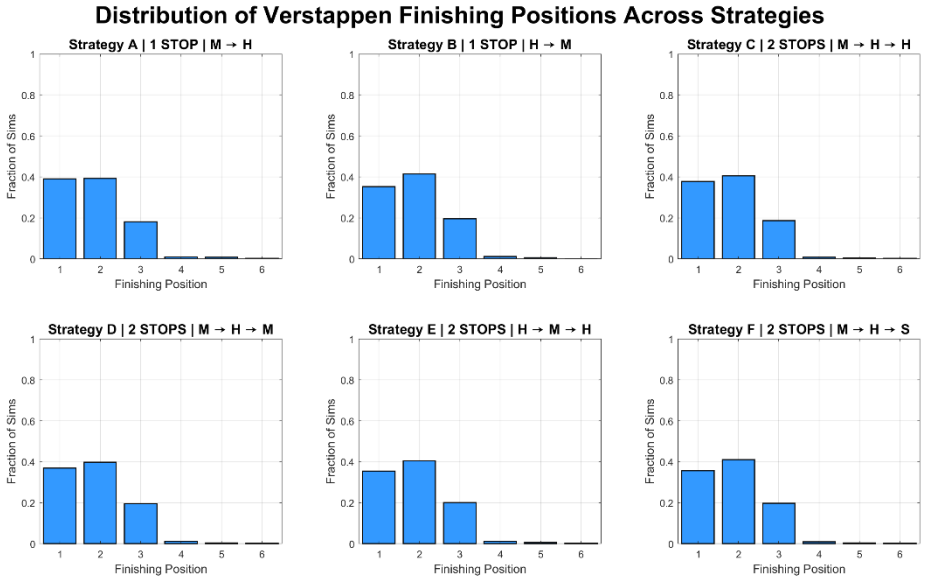

Figure 7.6: Distribution of finishing positions for Max Verstappen across six different strategies computed over 10000 simulations of MCS approach for the Italian Grand Prix

In conclusion, the analysis highlights how different race strategies can influence both total race time and finishing positions, even if the differences are often small for a top-performing driver like Verstappen. While strategy A (MEDIUM → HARD, 1 stop) proved to be the most effective overall, other strategies showed competitive results without offering significant advantages. These findings underline that, although strategy plays a key role, the performance of the car and driver remains crucial. Importantly, the use of the MCS approach provides valuable insights by showing how strategies might perform across many possible race scenarios, helping teams make more informed and confident decisions.

8 Future Developments

The simulator developed in this thesis has proven effective in capturing the essential dynamics of a Formula One race, offering a detailed view of race times, overtaking interactions and strategic consequences of pit stops and race events. Nonetheless, despite its robust framework and promising results, several areas remain open for further development to enhance the accuracy, realism and practical usefulness of the simulation tool.

The first important area for future improvement lies in refining the input parameters upon which the simulation is based. Currently, although the simulator handles numerous parameters effectively, there are cases where certain race outcomes, such as total race times or overtaking numbers, appear slightly overestimated when compared to actual race data. This is largely due to limitations in the availability of precise input data and the necessary simplifications adopted for some of the statistical models implemented. To address this, one avenue of development would be to introduce more detailed and physically based modelling for each parameter. For instance, tire degradation in the current simulator is modelled using empirical data fitted to simple quadratic or linear trends. However, tire performance in reality depends not only on the number of laps completed but also on the specific driving conditions experienced during those laps, such as longitudinal and lateral accelerations, cornering loads and surface temperatures. Incorporating such detailed telemetry-derived factors into the tire degradation model would allow for a more precise estimation of performance loss over time, leading to improved lap time predictions and more accurate strategic simulations.

This idea leads directly to the second area for future development, which is the inclusion of a vehicle dynamics model within the simulation architecture. At present, the simulator operates purely on statistical and empirical relationships, without considering the physical laws that govern the behavior of the vehicle on track. While the statistical approach has the advantage of computational simplicity and the ability to reproduce general race trends, it lacks the capacity to explain how the car physically interacts with different circuits or how vehicle setup changes might influence lap times and race outcomes. By integrating a vehicle dynamics model, whether in the form of a simplified quasi steady-state model or a more advanced transient dynamics simulation, it would be possible to simulate the car's behavior under various track conditions, including cornering forces, aerodynamic efficiency, braking capabilities and traction limitations. Such an enhancement would not only increase the fidelity of lap time predictions but would also open up the possibility to investigate the influence of engineering decisions, such as suspension tuning and aerodynamic setups, on race strategy and performance.

Closely linked to this is the third potential improvement, which involves the generation of the base lap time though a dedicated LTS. In its current form, the simulator calculates the base lap time by adjusting historical qualifying times with average race pace deltas derived from fastest laps during previous races. While this method provides a pragmatic solution, it is fundamentally dependent on the availability of historical data and therefore cannot easily accommodate scenarios involving new circuits, regulatory changes or significant shifts in vehicle performance characteristics. Developing a dedicated LTS, based on physics-driven models and incorporating track geometry, vehicle dynamics and tire characteristics, would enable the simulator to generate base lap times form first principles. This approach would reduce the dependency on past race data and enhance the flexibility of the simulator, allowing it to predict race performance under hypothetical or future conditions.

A fourth significant step forward would be the implementation of a strategy optimizer within the simulation framework. In this current version of the simulator, pit stop strategies are defined manually based on typical stint lengths and compound usage observed during past races. While this approach allows for realistic scenarios, it does not explore alternative strategies that might yield better outcomes under specific race conditions. A strategy optimization module could dynamically determine the most

effective pit stop laps and compound choices to minimize total race time, taking into account variables such as evolving tire degradation, traffic conditions, risk of SC deployments and competitors' strategies. This capability would transform the simulator from a tool used primarily for evaluating predefined strategies into an instrument capable of actively generating optimal strategies tailored to particular race circumstances.

Finally, a powerful expansion of the simulator would be the inclusion of wet-weather modelling. At present, the simulator is designed exclusively for dry weather conditions, as wet races introduce a level of variability and complexity that significantly exceeds the capabilities of the existing model. Wet weather affects almost every aspect of race performance, including tire compounds and grip levels, braking distances, fuel consumption and the probability of incidents such as aquaplaning or spins. Additionally, wet conditions often lead to more frequent SC deployments and dramatic shifts in race strategy. Developing models to capture these effects, including the behavior of intermediate and full wet tires, variable track drying rates and changing grip levels, would considerably increase the versatility and realism of the simulator. It would also enable the simulation of a broader range of race scenarios, reflecting the unpredictable nature of real Formula One seasons.

In summary, while the simulator already represents a significant step forward in modelling race dynamics and strategic outcomes, these future developments offer the possibility of transforming it into an even more powerful tool. Incorporating physically based vehicle modelling, advanced parameter calibration, strategy optimization, base lap time simulation and wet-weather capabilities would further align the simulator with the complex realities of modern Formula One racing. These improvements would not only enhance predictive accuracy but also provide deeper insights for engineers and strategists seeking to optimize performance in one of the most competitive and data-driven sports in the world.

9 Conclusions

Over the past two decades, Formula One has transformed into an increasingly data-driven sport, where simulation tools have become essential for both engineering development and strategic race planning. The motivation for this thesis arose from the need to build a race simulator capable of predicting race outcomes not only under ideal circumstances but also under the uncertainties and stochastic events that inevitably shape real races. As outlined in the introduction, this demand stems from regulatory constraints limiting on-track testing, the complexity of modern vehicles and the critical importance of race strategy in determining final results.

This thesis sets out to develop a probabilistic race simulator for Formula One, integrating telemetry-derived parameters and statistical models to replicate the dynamics of an entire race. The simulator was designed to operate at a lap-by-lap, driver-specific level, accounting for multiple influencing factors such as fuel consumption, tire degradation, driver variability, pit stop performance FCY phases and overtaking maneuvers. A key objective was to move beyond purely deterministic models and to embrace the inherent randomness of motorsport events, providing not just single predictions but distributions of possible outcomes.

The work presented in this thesis has achieved these goals through the construction of a modular simulator implemented in MATLAB, supported by extensive data acquisition using the FastF1 Python library and other statistical data sources. The simulator combines deterministic modelling, based on known physical and empirical relationships, with stochastic components driven by probability distributions calibrated from historical race and telemetry data. This hybrid approach allows the simulator to capture the average behavior of a race while also reflecting the variability that can dramatically influence results.

The effectiveness of the developed simulator was validated through two levels of analysis. First, single simulation runs were conducted for selected race events. These single simulations demonstrated the simulator's ability to generate realistic lap-by-lap evolutions of driver positions, tire usage strategies and cumulative race times. Visual outputs, including lap time trends, race position plots and overtaking summaries, confirmed that the simulator closely reproduces the typical patterns and dynamics of a real Grand Prix. For instance, in the simulation of the 2023 Italian Grand Prix, the lap times exhibited realistic degradation patterns, pit stop sequences were faithfully represented and the time gaps between drivers were consistent with observed race behaviors. Nevertheless, the single simulations also highlighted areas where the simulator tends to overestimate certain metrics, such as the number of overtakes, pointing to opportunities to further refinement.

Beyond single event analysis, the simulator was extended to perform large-scale MCS, running thousands of race iterations to analyze the statistical properties of race outcomes. This Monte Carlo approach proved highly valuable, allowing the evaluation of strategy robustness and the quantification of risks associated with various race scenarios. The results of the MCS revealed meaningful insights into how different variable, such as the occurrence of SC phases or stochastic driver variability, influence overall race dynamics. Additionally, the simulator successfully captured the distribution of time gaps between drivers across simulations, reflecting the spread typically observed in real races. However, the Monte Carlo analysis also revealed discrepancies in certain areas, such as an underestimation of number of SC deployments compared to historical data, suggesting that some probability parameters might require recalibration.

Despite these minor discrepancies, the results collectively demonstrate that the simulator effectively fulfils its primary purpose, modelling Formula One races in a way that combines event realism with stochastic variability. It enables detailed analysis of strategic decisions and provides a framework for exploring not only expected outcomes but also the full range of possible race scenarios. Such

capabilities are invaluable for engineers and strategists seeking to optimize race performance in a highly competitive environment where small margins can determine success or failure.

10 Appendix

Table 10.1: Circuit-specific quadratic and linear parameters regarding the compounds used during 2023 Formula One season

Circuit	Parameter	C0 tire	C1 tire	C2 tire	C3 tire	C4 tire	C5 tire
Bahrain Grand Prix	Quadratic	*No data*	0.0059	0.0384	0.0118	*No data*	*No data*
	Linear	*No data*	0.0613	0.0000	0.0593	*No data*	*No data*
Saudi Arabian Grand Prix	Quadratic	*No data*	*No data*	0.0006	0.0017	0.0014	*No data*
	Linear	*No data*	*No data*	0.0076	0.0044	0.0698	*No data*
Australian Grand Prix	Quadratic	*No data*	*No data*	0.0003	0.0012	0.0000	*No data*
	Linear	*No data*	*No data*	0.0001	0.0342	0.0000	*No data*
Azerbaijan Grand Prix	Quadratic	*No data*	*No data*	*No data*	0.0003	0.0027	*No data*
	Linear	*No data*	*No data*	*No data*	0.0074	0.0068	*No data*
Miami Grand Prix	Quadratic	*No data*	*No data*	0.0002	0.0005	*No data*	*No data*
	Linear	*No data*	*No data*	0.0208	0.0142	*No data*	*No data*
Monaco Grand Prix	Quadratic	*No data*	*No data*	*No data*	0.0005	0.0018	0.0028
	Linear	*No data*	*No data*	*No data*	0.0269	0.0008	0.0000
Spanish Grand Prix	Quadratic	*No data*	0.0003	0.0005	0.0046	*No data*	*No data*
	Linear	*No data*	0.0712	0.0466	0.0345	*No data*	*No data*
Canadian Grand Prix	Quadratic	*No data*	*No data*	*No data*	0.0016	0.0022	0.0029
	Linear	*No data*	*No data*	*No data*	0.0063	0.0074	0.0000
Austrian Grand Prix	Quadratic	*No data*	*No data*	*No data*	0.0013	0.0034	*No data*
	Linear	*No data*	*No data*	*No data*	0.0411	0.0327	*No data*
British Grand Prix	Quadratic	*No data*	0.0020	0.0004	0.0011	*No data*	*No data*
	Linear	*No data*	0.0530	0.0147	0.0190	*No data*	*No data*
Hungarian Grand Prix	Quadratic	*No data*	*No data*	*No data*	0.0016	0.0020	0.0030
	Linear	*No data*	*No data*	*No data*	0.0479	0.0147	0.0000
Belgian Grand Prix	Quadratic	*No data*	*No data*	0.0000	0.0067	0.0033	*No data*
	Linear	*No data*	*No data*	0.1714	0.1603	0.1116	*No data*
Dutch Grand Prix	Quadratic	*No data*	0.0000	0.0017	0.0021	*No data*	*No data*
	Linear	*No data*	0.0000	0.0031	0.0273	*No data*	*No data*

Italian Grand Prix	Quadratic	*No data*	*No data*	*No data*	0.0021	0.0052	*No data*
	Linear	*No data*	*No data*	*No data*	0.0376	0.0091	*No data*
Singapore Grand Prix	Quadratic	*No data*	*No data*	*No data*	0.0015	0.0036	0.0053
	Linear	*No data*	*No data*	*No data*	0.0090	0.0098	0.0000
Japanese Grand Prix	Quadratic	*No data*	0.0034	0.0040	0.0078	*No data*	*No data*
	Linear	*No data*	0.0733	0.0794	0.0631	*No data*	*No data*
Qatar Grand Prix	Quadratic	*No data*	0.0028	0.0018	0.0000	*No data*	*No data*
	Linear	*No data*	0.0179	0.0337	0.0000	*No data*	*No data*
United States Grand Prix	Quadratic	*No data*	*No data*	0.0022	0.0051	0.0000	*No data*
	Linear	*No data*	*No data*	0.0366	0.0276	0.4945	*No data*
Mexican Grand Prix	Quadratic	*No data*	*No data*	*No data*	0.0007	0.0014	*No data*
	Linear	*No data*	*No data*	*No data*	0.0826	0.0555	*No data*
San Paulo Grand Prix	Quadratic	*No data*	*No data*	*No data*	0.0020	0.0044	*No data*
	Linear	*No data*	*No data*	*No data*	0.0358	0.0138	*No data*
Las Vegas Grand Prix	Quadratic	*No data*	*No data*	*No data*	0.0026	0.0070	*No data*
	Linear	*No data*	*No data*	*No data*	0.0066	0.0008	*No data*
Abu Dhabi Grand Prix	Quadratic	*No data*	*No data*	*No data*	0.0027	0.0068	*No data*
	Linear	*No data*	*No data*	*No data*	0.0192	0.0195	*No data*

Table 10.2: Driver-specific mean variability values ε_{lap} (Equation 5.11) and probabilistic starting performance values $t_{start,performance}$ [4] (Equation 5.13). The table accounts for all the drivers who appeared in the 2023 Formula One season

Driver	**Team**	$\varepsilon_{\mathbf{lap}}$	$t_{start,performance}$
Alexander ALBON	Williams	$\mathcal{N}(0,\ 0.3278^2)$	$\mathcal{N}(0.044,\ 0.174^2)$
Fernando ALONSO	Aston Martin	$\mathcal{N}(0,\ 0.3622^2)$	$\mathcal{N}(-0.009,\ 0.0152^2)$
Valtteri BOTTAS	Alfa Romeo	$\mathcal{N}(0,\ 0.3547^2)$	$\mathcal{N}(0.088,\ 0.245^2)$
Nick DE VRIES	AlphaTauri	$\mathcal{N}(0,\ 0.3221^2)$	$\mathcal{N}(0.027,\ 0.146^2)$
Pierre GASLY	AlphaTauri	$\mathcal{N}(0,\ 0.3795^2)$	$\mathcal{N}(-0.047,\ 0.112^2)$
Lewis HAMILTON	Mercedes	$\mathcal{N}(0,\ 0.3509^2)$	$\mathcal{N}(-0.052,\ 0.098^2)$
Nico HULKENBERG	Haas	$\mathcal{N}(0,\ 0.3982^2)$	$\mathcal{N}(0.102,\ 0.156^2)$
Liam LAWSON	AlphaTauri	$\mathcal{N}(0,\ 0.3852^2)$	$\mathcal{N}(0.027,\ 0.146^2)$
Charles LECLERC	Ferrari	$\mathcal{N}(0,\ 0.3543^2)$	$\mathcal{N}(-0.042,\ 0.125^2)$
Kevin MAGNUSSEN	Haas	$\mathcal{N}(0,\ 0.3804^2)$	$\mathcal{N}(-0.014,\ 0.171^2)$
Lando NORRIS	McLaren	$\mathcal{N}(0,\ 0.3839^2)$	$\mathcal{N}(0.003,\ 0.142^2)$
Esteban OCON	Alpine	$\mathcal{N}(0,\ 0.3351^2)$	$\mathcal{N}(-0.027,\ 0.115^2)$
Sergio PEREZ	RedBull	$\mathcal{N}(0,\ 0.3979^2)$	$\mathcal{N}(0.044,\ 0.115^2)$
Oscar PIASTRI	McLaren	$\mathcal{N}(0,\ 0.3969^2)$	$\mathcal{N}(-0.097,\ 0.126^2)$
Daniel RICCIARDO	AlphaTauri	$\mathcal{N}(0,\ 0.4054^2)$	$\mathcal{N}(0.027,\ 0.146^2)$
George RUSSELL	Mercedes	$\mathcal{N}(0,\ 0.3744^2)$	$\mathcal{N}(-0.028,\ 0.135^2)$
Carlos SAINZ	Ferrari	$\mathcal{N}(0,\ 0.3469^2)$	$\mathcal{N}(-0.050,\ 0.115^2)$
Logan SARGEANT	Williams	$\mathcal{N}(0,\ 0.3734^2)$	$\mathcal{N}(0.029,\ 0.128^2)$
Lance STROLL	Aston Martin	$\mathcal{N}(0,\ 0.3806^2)$	$\mathcal{N}(-0.095,\ 0.135^2)$
Yuki TSUNODA	AlphaTauri	$\mathcal{N}(0,\ 0.3348^2)$	$\mathcal{N}(0.050,\ 0.168^2)$
Max VERSTAPPEN	RedBull	$\mathcal{N}(0,\ 0.3270^2)$	$\mathcal{N}(-0.001,\ 0.171^2)$
Guanyu ZHOU	Alfa Romeo	$\mathcal{N}(0,\ 0.3913^2)$	$\mathcal{N}(0.006,\ 0.148^2)$

Table 10.3: Driver-specific data-driven retirement percentages [17] and estimated posterior distribution parameters

Driver	Team	Historic Retirement Percentage	Posterior Distribution
Alexander ALBON	Williams	0.115	$Beta(15, 126)$
Fernando ALONSO	Aston Martin	0.025	$Beta(13, 426)$
Valtteri BOTTAS	Alfa Romeo	0.065	$Beta(19, 264)$
Nick DE VRIES	AlphaTauri	0.039	$Beta(13,281)$
Pierre GASLY	AlphaTauri	0.105	$Beta(19, 171)$
Lewis HAMILTON	Mercedes	0.011	$Beta(7, 387)$
Nico HULKENBERG	Haas	0.022	$Beta(8, 256)$
Liam LAWSON	AlphaTauri	0.039	$Beta(13, 281)$
Charles LECLERC	Ferrari	0.095	$Beta(17, 167)$
Kevin MAGNUSSEN	Haas	0.108	$Beta(23, 199)$
Lando NORRIS	McLaren	0.086	$Beta(14, 151)$
Esteban OCON	Alpine	0.103	$Beta(19, 174)$
Sergio PEREZ	RedBull	0.043	$Beta(15, 303)$
Oscar PIASTRI	McLaren	0.065	$Beta(6, 77)$
Daniel RICCIARDO	AlphaTauri	0.039	$Beta(13, 281)$
George RUSSELL	Mercedes	0.132	$Beta(20, 145)$
Carlos SAINZ	Ferrari	0.068	$Beta(17, 226)$
Logan SARGEANT	Williams	0.194	$Beta(10, 63)$
Lance STROLL	Aston Martin	0.108	$Beta(21, 182)$
Yuki TSUNODA	AlphaTauri	0.115	$Beta(13, 111)$
Max VERSTAPPEN	RedBull	0.057	$Beta(15, 231)$
Guanyu ZHOU	Alfa Romeo	0.132	$Beta(12, 93)$

Table 10.4: Driver-specific probability values, computed by means of Bayesian inference, regarding accidents ($P_{accident}$), accidents on lap 1 ($P_{accident,lap1}$) and failures ($P_{failure}$ [4])

Driver	**Team**	$\boldsymbol{P_{accident}}$	$\boldsymbol{P_{accident,lap1}}$	$\boldsymbol{P_{failure}}$
Alexander ALBON	Williams	0.0859	0.859	0.0568
Fernando ALONSO	Aston Martin	0.0239	0.239	0.0448
Valtteri BOTTAS	Alfa Romeo	0.0541	0.541	0.0448
Nick DE VRIES	AlphaTauri	0.0357	0.357	0.0568
Pierre GASLY	AlphaTauri	0.0806	0.806	0.0688
Lewis HAMILTON	Mercedes	0.0144	0.144	0.0328
Nico HULKENBERG	Haas	0.0246	0.246	0.0808
Liam LAWSON	AlphaTauri	0.0357	0.357	0.0568
Charles LECLERC	Ferrari	0.0745	0.745	0.0688
Kevin MAGNUSSEN	Haas	0.0834	0.834	0.0808
Lando NORRIS	McLaren	0.0685	0.685	0.0936
Esteban OCON	Alpine	0.0793	0.793	0.0688
Sergio PEREZ	RedBull	0.0380	0.380	0.0568
Oscar PIASTRI	McLaren	0.0590	0.590	0.0936
Daniel RICCIARDO	AlphaTauri	0.0357	0.357	0.0568
George RUSSELL	Mercedes	0.0576	0.576	0.0328
Carlos SAINZ	Ferrari	0.0564	0.564	0.0688
Logan SARGEANT	Williams	0.0945	0.945	0.0568
Lance STROLL	Aston Martin	0.0833	0.833	0.0448
Yuki TSUNODA	AlphaTauri	0.0847	0.847	0.0568
Max VERSTAPPEN	RedBull	0.0492	0.492	0.0568
Guanyu ZHOU	Alfa Romeo	0.0925	0.925	0.0448

Table 10.5: Circuit-specific time loss relative to pit lane travel and circuit-specific overtaking data

Circuit	$t_{pitdrive}$ **[s]**	α_{overtake}	$P_{overtake}$
Bahrain Grand Prix	20.057	-0.353	0.355
Saudi Arabian Grand Prix	16.112	-0.305	0.432
Australian Grand Prix	12.798	-0.300	0.440
Azerbaijan Grand Prix	14.445	-0.283	0.467
Miami Grand Prix	19.814	-0.351	0.358
Monaco Grand Prix	18.984	-0.600	0.600
Spanish Grand Prix	17.415	-0.351	0.359
Canadian Grand Prix	18.914	-0.291	0.454
Austrian Grand Prix	18.644	-0.368	0.331
British Grand Prix	24.462	-0.294	0.45
Hungarian Grand Prix	17.222	-0.314	0.417
Belgian Grand Prix	21.222	-0.387	0.301
Dutch Grand Prix	15.021	-0.450	0.300
Italian Grand Prix	19.746	-0.326	0.398
Singapore Grand Prix	25.152	-0.595	0.448
Japanese Grand Prix	18.792	-0.294	0.45
Qatar Grand Prix	22.842	-0.334	0.386
United States Grand Prix	19.814	-0.370	0.328
Mexican Grand Prix	17.883	-0.309	0.425
San Paulo Grand Prix	23.297	-0.339	0.377
Las Vegas Grand Prix	19.814	-0.437	0.221
Abu Dhabi Grand Prix	17.019	-0.387	0.301

Table 10.6: Team-specific Fisk distribution parameters (taking into account that loc parameter is null for simplicity) and average pit stop time relative to 2023 Formula One season [4, 23]

Teams	**Shape**	**Scale**	**Average Pit Stop Time [s]**
RedBull	2.045	0.598	2.64
Ferrari	2.414	0.737	2.90
McLaren	2.639	0.98	2.91
Aston Martin	3.498	1.188	3.08
Mercedes	1.563	0.48	3.18
Alpine	3.876	1.433	3.19
AlphaTauri	4.29	1.606	3.23
Williams	2.562	0.966	3.28
Alfa Romeo	5.827	2.327	3.51
Haas	5.324	1.866	4.22

References

[1] J. Woodhouse, "All the major changes for the F1 2022 season," in *PlanetF1*, 2023.

[2] Catapult, "How Data Analysis Transforms F1 Race Performance," in *Catapult Sports*, Melbourne, 2024.

[3] A. Heilmeier, M. Graf and M. Lienkamp, "A Race Simulation for Stategy Decisions in Circuit Motorsports," in *IEEE 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, 2018.

[4] A. Heilmeier, M. Graf, J. Betz and M. Lienkamp, "Application of Monte Carlo Methods to Consider Probabilistic Effects in a Race Simulation for Circuit Motorsport," *Applied Science,* vol. 10, no. 12, 2020.

[5] C. Sulsters, "Simulating Formula One Race Strategies," in *Vrije Universiteit Amsterdam*, Amsterdam, 2018.

[6] Formula 1, "Everything you need to know about F1 - Drivers, teams, cars, circuits and more," in *formula1.com*, London, 2024.

[7] F1 Chronicle, "Formula 1 Fuel Tank Safety Explained," in *f1chronicle.com*, 2024.

[8] Red Bull Racing, "The Bulls' Guide to Tyres," in *RedBulRacing.com*, Milton Keynes, 2024.

[9] Formula 1, "Parc Fermé," in *Formula1.com*, 2016.

[10] G. Fieni, M. P. Neumann, F. Furia and A. Caucino, "Game Theory in Formula 1: Multi-agent Physical and Strategical Interactions," in *Cornell University: Systems and Control*, New York, 2025.

[11] J. Timings and D. Cole, "Tobust lap-time simulation," *Journal of Automobile Engineering,* vol. 228, no. 10, pp. 1200 - 1216, 2014.

[12] A. Reinero and A. Tonoli, "Lap time simulator for a single-seater electric car," in *Master's Thesis at Politecnico di Torino*, Torino, 2021.

[13] J. Bekker and W. Lotz, "Planning Formula One Race Strategies Using Discrete-Event Simulation," *The Journal of the Operational Research Society,* vol. 60, no. 7, pp. 952-961, 2009.

[14] F1Metrics, "Building a race simulator," in *Mathematical and statistical insights into Formula 1*, 2014.

[15] T. Salminem, "Race Simulator: Downloadable R Program Code," F1 Strategy Blog, 2020.

[16] GitHub, "FastF1 Python Library," 2023.

[17] StatsF1, "Statistics Drivers, Misc, Retirement," in *statsf1.com*, 2025.

[18] Federation Internationale de l'Automobile, "2018 F1 Sporting Regulations," in *fia.com*, 2011.

[19] Red Bull Racing, "Bulls' Guide To: Formula One Fuel," in *RedBullRacing.com*, Milton Keynes, 2024.

[20] McLaren Racing Limited Sports Technology, "Formula," in *The Royal Academy of Engineering*, Woking, 2012.

[21] formula1.com, "Pirelli clarify new "free choice" tyre regulations," in *formula1.com*, 2015.

[22] P. Fisk, "The Graduation of Income Distributions," *Econometrica,* vol. 29, no. 2, pp. 171 - 185, 1961.

[23] A. Dickinson, "F1 Pit Stop Championship 2023 - who comes out on top?," in *Total Motorsport*, 2023.

[24] M. Taboga, "Bayesian inference," in *Lectures on probability theory and mathematical statistics*, Kindle Direct Publishing. Online appendix, 2021.

[25] T. Thorns, "The F1 Safety Car - All You Need To Know," in *f1mix.com*, 2023.

[26] David, "F1 Car Length: Understanding Formula 1 Vehicle Dimensions," in *apexbite.com*, 2024.

[27] Mercedes AMG Petronas F1, "INSIGHT: Understanding F1 Safety Cars," in *mercedesamgf1.com*, 2018.

[28] Scuderia Fans, "Overtakes per race in 2022: Brazil on top, Imola and Monaco at bottom of rankings," in *scuderiafans.com*, 2022.

[29] M. Coleman, "What is DRS and why does the device raise criticism?," in *The Athletic*, 2024.

[30] Z. Chen, "A Review of Research Progress in Monte Carlo Methods," in *Theoretical and Natural Science*, 2025.

[31] formula1.com, "Formula 1 Pirelli Gran Premio d'Italia 2023," in *formula1.com*, 2023.

[32] K. Collantine, «Analysis: Why Ferrari didn't go aggressive on race strategy in tehir fight with Red Bull,» in *racefans.net*, 2023.