Predictive Tyre Degradation Analysis and Strategy Optimization in Formula 1 Using Deep Learning and Telemetry Data

Eduard Ruiz Munné

Abstract. Accurately predicting tyre degradation is vital for Formula 1 teams' strategic planning. By determining the degradation rate, teams can identify the optimal tyre compound combination for achieving the best results. In this project, we developed an initial exploratory data analysis (EDA) and subsequently normalized fuel consumption. We then applied a second order polynomial function fitting. Additionally, a feature engineering process generated 86 features, which were utilized in predicting the three coefficients defining the degradation rate. Both Random Forest and Neural Network models were employed for the prediction task.

Keywords: Formula 1 · Random Forest · Feature Engineering.

1 Introduction

Formula 1 has evolved into a highly strategic competition that demands precise management of tyre degradation by drivers. The driving style and weather conditions significantly impact tyre degradation, making accurate predictions before a race is a differentiating factor among drivers. The objective of this project is to fit a model that accurately represents the tyre degradation rate using telemetry and meteorological information, providing valuable support to F1 engineers in determining optimal strategies and predicting the duration of selected tyre compounds.

This tool can result particularly useful after conducting various training sessions and testing different strategies during the free training sessions. In this project, we will evaluate both a Random Forest Regressor and a Deep Learning solution to predict the tyre degradation rate.

Tyre degradation refers to the thermal performance of the car's tyres, which is influenced by their internal temperature. Maintaining the tyres within the appropriate temperature range is crucial for optimal performance. If tyres are too cold, they become hard and brittle, leading to reduced grip, stability, and speed, and surface can brake in a process known as graining. On the other hand, if the tyres are too warm, they become elastic and rubbery, resulting in the removal of large pieces of rubber from the track. Achieving the perfect grip and contact between the tyre and the track requires maintaining the tyres in a specific temperature range. Deviating from this range can negatively impact tyre performance and lap times [6]. Understanding and managing tyre temperature is essential to plan strategies, going out makes degradation unpredictable [1].

Tyre degradation is influenced by various factors, including the temperature of the track, the tyre compound, and the driving strategy (avoiding aggressive movements) [5]. The choice of tyre compound is a critical part of a race strategy. Soft tyres are prone to overheating, which reduces their performance and durability. Conversely, hard tyres provide less grip but experience less overheating due to their harder rubber composition, allowing them to last longer. Medium tyres stay in the middle, offering a balance between performance and durability.

Selecting the optimal combination of tyre compounds, aiming for good performance with minimal degradation (considering that a pit stop can take around 20 seconds), is crucial for gaining an advantage over the rest of the cars. By accurately predicting tyre degradation, teams can make informed decisions to maximize performance and minimize the need of pit stops.

2 Methods

The dataset used for this project was obtained from a public API, Fast F1[1]. Data from races from 2021 until 2023 was considered. In addition to lap times and tyre compounds, this API provides detailed data such as the car's telemetry, containing information for velocity, brake and throttle usage at each moment. Furthermore, it supplies positional data, which allowing to track the car's position on the track, and event data, including yellow or red flags and pit stop information. Meteorological data encompassing air temperature, humidity and track temperature among others, is also available. However, for this project, we will primarily focus on the car's telemetry data and meteorological data. These two aspects are known to have the a remarkable impact on tyre degradation.

The project implementation entailed four steps:

1. Exploratory Data Analysis (EDA): We conducted a thorough analysis of the API data, extracting valuable insights and gaining a comprehensive understanding of the data available.
2. Normalization of lap times: To ensure fairness and consistency in the data, we applied a normalization technique to the lap times, considering factors such as fuel consumption which lead to a weight reduction, hence reducing the lap time.
3. Feature engineering: This step involved leveraging our domain knowledge from Formula 1 and the available data to generate informative features that would enhance the predictive capabilities of our models. This process aims to capture meaningful patterns and relationships to express car's telemetry, meteorological conditions, and tyre degradation.
4. Model prediction: Leveraging lap time normalization and feature engineering, we developed predictive models that could forecast tyre degradation based on the developed features.

[1] Fast F1 API for Python: https://docs.fastf1.dev

By following these four steps, we aim to extract valuable insights from the API data, enhance the reliability of our analysis through normalization, and employ advanced modeling techniques to predict tyre degradation.

2.1 Exploratory Data Analysis

First, we conducted a data analysis to explore the available data and determine the relevant information that can be extracted. Our objective was to develop a fitting for the degradation curve. Initially, we examined the strategies employed by different drivers within the same race. It is crucial to acknowledge that each driver adopts distinct approaches based on team instructions and individual driving styles. The degradation rate significantly varies between driving a Ferrari and driving a Red Bull, emphasizing the importance of considering the specific car being utilized.

Consequently, we recognized the significance of accounting for both the car and the circuit. Each circuit contains unique characteristics, encompassing different turns, velocities, and race conditions. The driving experience in Bahrain, for instance, differs substantially from the one in Canada. Consequently, we evaluated each race independently to comprehend the degradation rate in each specific context.

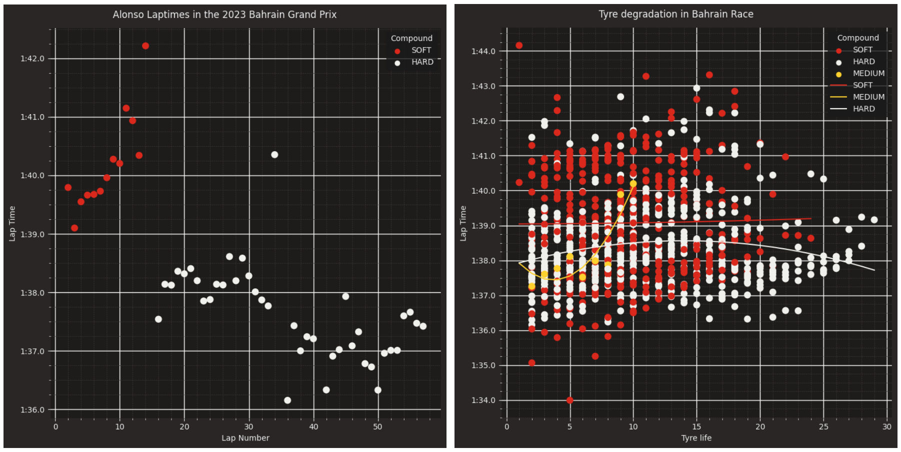
This Exploratory Data Analysis (EDA) process played a pivotal role in comprehending the data and filtering out the most meaningful insights. Throughout a race, drivers may encounter outlier lap times, typically resulting from incidents such as crashes or pit stops. These irregular values can introduce errors into our model. To mitigate this, we exclusively focused on quick laps that where contained within 107% of the fastest lap time. We consider these laps as reliable since they are unaffected by external conditions and they are considered quick laps.

Furthermore, we specifically focused on dry conditions and limited our analysis to the three most commonly used tyre compounds: soft, medium, and hard. This deliberate choice helped streamline the complexity of our problem, allowing us to concentrate on other factors. For further studies, wet conditions can be considered.

2.2 Lap time normalization

By analyzing our dataset, we noticed a strange behavior in the lap times. Contrary to the anticipated tyre degradation process, we discovered a consistent decrease in lap times over time. However, this unexpected trend can be attributed to fuel consumption (Figure 1a). At the start of an F1 race, all cars carry a weight of 110kg of fuel, which is gradually depleted during the race [2]. As cars become lighter, their speed increases, hence reducing lap times.

During the EDA phase, we uncovered that the hard tyres exhibited superior performance compared to the soft tyres (Figure 1b). This disparity arose because drivers initially used the soft tyres, when the car was heavier, and later switched



(a) Lap time evolution with the number of laps. Fuel consumption shows a linear decrease.

(b) Tyre degradation in Bahrain. Hard tyres have better performance before normalization.

Fig. 1: Each color represents a compound. Red: soft, yellow: medium and white: hard.

to the hard tyres, when consuming fuel. Thus, the improved performance of the hard tyres was better due to their usage with a lighter car.

To address this, we delved into the existing literature and discovered that the most prevalent normalization method involves reducing the lap time by 0.3 seconds after each lap [3]. This reduction accounts for the fuel combustion and subsequent weight loss of the car.

Following the lap time normalization process, we employed a polynomial function to fit our data points. Considering other studies that utilized this function order, we selected a second-order polynomial function as a suitable approach [4].

2.3 Feature engineering

Following EDA and lap time normalization steps, our next objective was to generate new features to serve as inputs for our model. The desired outcome was to predict tyre degradation for each combination of driver, compound, city, and year, utilizing a second-order polynomial function with three coefficients as the output.

To ensure comprehensive coverage of the entire race, our focus was on generating features that captured temporal information. We successfully created a total of 34 features (86 after one hot encoding), encompassing essential details such as tyre compound (soft, medium, or hard), city of the race, year of the race (considering potential changes in the race layout), and driver.

In addition to the basic information, we incorporated meteorological data, including air temperature, humidity, wind speed, and track temperature. For each of these metrics, we generated four features per value, such as the mean, standard deviation, minimum and maximum values across the race duration.

Turning to the telemetry data, we designed specific features for two boolean values: brake and DRS. These features included:

- Percentage activated: Represents the percentage of race time during which the brake or DRS was activated, contributing to increased peak car velocity.
- Number of times activated per lap: Indicates the frequency of peaks in the brake or DRS signal, normalized by the total number of laps.
- Width of the peaks: Measures the length of the peaks using the original sample frequency, providing insights into the duration of brake or DRS activations.

Regarding car speed, we employed the same four features (mean, standard deviation, minimum, and maximum) as utilized for the meteorological data. As for the throttle, which ranges from 0 to 100, we devised the following features:

- Hard throttle: Represents the percentage of time during which the throttle exceeded 75.
- Soft throttle: Indicates the percentage of time during which the throttle value fell within the range $(0, 30)$, excluding these values.
- Mean throttle: Calculates the average intensity of the throttle.
- Standard deviation: Captures the variation in throttle intensity.

It is worth mentioning that all these features can be obtained in advance by simulating various scenarios. For example, different velocities or braking intensities during turns can be simulated. Since Formula 1 is an individual sport, there is low affections from other racers and driving strategies can be simulated in advance to generate the desired plots. Furthermore, modern systems are available today to predict meteorological conditions with high precision some hours in advance. Teams have access to this information prior to the race, making it feasible to incorporate meteorological data as an input without any complications.

2.4 Model prediction

Finally, we implemented two distinct models to predict the three coefficients of our second-order degradation function.

Firstly, we employed a random forest model to gain explainability in our predictions. Understanding which features are most relevant in determining the degradation rate can provide valuable insights for a team when planning their strategy. We utilized a multi-output regressor that generated three outputs representing the three factors. The random forest consisted of 2,000 estimators with a maximum tree depth of 200.

In order to further enhance our results, we developed a simple neural network. This network incorporated all 86 features as inputs and was trained for 10,000 epochs using a batch size of 32. The hidden layers consisted of 128, 256, 512, 128, 64, and 16 units, consecutively.

Both models were trained and evaluated using a test set comprising 20% of the total 1,663 samples. The Mean Absolute Error was the metric to evaluate the performance for a tyre life of 20 laps. The remaining samples were used for training.

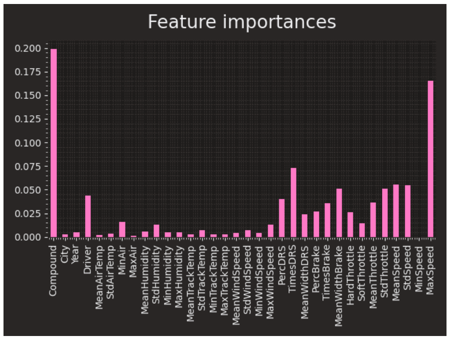
3 Results

The Multi Output Random Forest Regressor was trained to predict three outputs, resulting in a Mean Absolute Error of 13.538 seconds. This performance is considered extremely poor in the context of Formula 1, where races can be decided by few milliseconds. However, in an attempt to enhance our approximation, we made an adjustment by setting the zero-order coefficient (value when x is 0) equal to the actual data fitting. This adjustment can be justified since teams can determine the lap time for a Tyre Life of 1 during Free Practice sessions. Incorporating this correction improved our fitting slightly, resulting in a reduced MAE of 12.368 seconds. Although this slight improvement is marginal, the error remains significantly high.

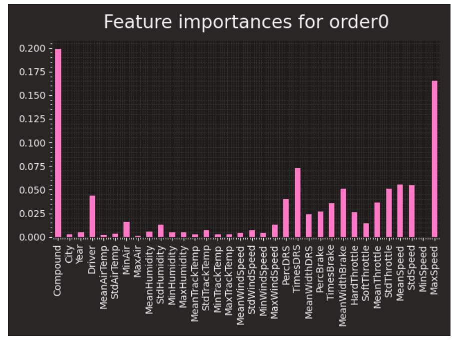
Furthermore, we investigated the most influential features utilized by the Random Forest for classification. The average feature importance across the three second-order function coefficients is displayed in Figure 2a. As expected, Tyre Compound emerges as the most significant feature, followed by maximum speed and time using DRS. However, the significance of the error can be attributed to the negligible importance of the City feature. Considering that each race entails distinct curves and lap times, this feature should play a crucial role in predicting tyre degradation. To gain further insights, we specifically examined the relevance of the zero-order coefficient, which the findings are presented in Figure 2b. These results confirm that the model is incorrectly learning and assigning excessive importance to certain features that should be less influential.

The Neural Network was trained for 10,000 epochs with 80% of the total dataset. The mean absolute error was of 20.173 seconds for the test data. After correcting the absolute error we got a value of 13.886 seconds. As a result, the model performs worse than the Random Forest, without capturing the relevant information despite the high complexity of the network.

However, it is important to consider that the ground truth against which the model was evaluated contains certain errors, which means that the model's performance is actually better than the reported values suggest. One scenario that contributed to this difference is when the same compound was used multiple times in a race, resulting in distinct curves instead of a single one. Consequently, the ground truth fitting becomes inaccurate, and the error becomes significantly higher than what the reality is. This case is illustrated in Figure 3a, where the predicted line outperforms the actual fitting, leading to a higher MAE and subsequently increasing the average error. On the other hand, there were other



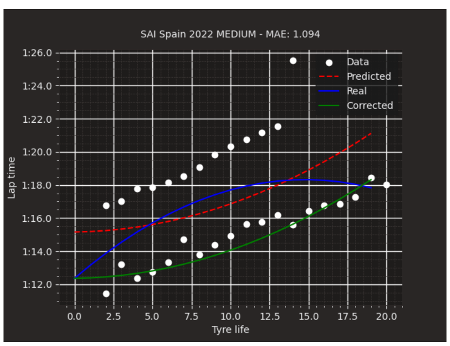
(a) Average feature importance of the three coefficients.



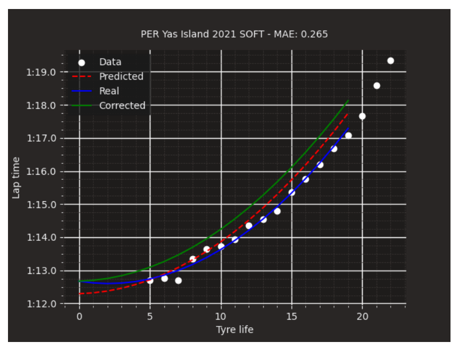
(b) Feature importance of the 0-order coefficient.

Fig. 2: Feature importance for the Random Forest Regressor model.

predictions that exhibited greater accuracy, as evidenced by their low MAE, as shown in Figure 3b.



(a) Predicted fitting (green) is better than ground truth fitting (red).



(b) Very accurate prediction (green) close to the ground truth (red).

Fig. 3: Both images show a good prediction from our model, however the first will lead to a high MAE and the second to a low MAE. Both of them are equally good though.

4 Discussion

The stability of our results is the main drawback that needs to be addressed, as a difference of more than 10 seconds does not provide meaningful information to a team. Nevertheless, in some cases accurate predictions of tyre degradation can be achieved by using telemetry and meteorological data, which can be known before a race.

It is important to note that we are working with publicly available data, whereas teams have access to additional sensor data that can significantly enhance the accuracy of predictions. In particular, tyre temperature can be measured accurately, playing a crucial role in determining the level of degradation, as previously explained. External factors also play a significant role in the resulting lap times. Crashes can affect different parts of the car, the temperature of the track can change, residues from other drivers' rubber can alter tyre performance, and track abrasion can vary. These factors make predicting tyre degradation a challenging task, even for Formula 1 teams.

The results obtained are not satisfactory, considering the substantial difference that a few seconds can make in Formula 1. While some predictions have been reasonably accurate, there is still ample room for improvement in the model. It is evident that there is noise in the features, with many being irrelevant and others lacking in relevance.

One potential improvement could involve incorporating information from Free Practice sessions (three sessions before the race) as input for our model. Fitting a second-order function to this data and using the resulting three coefficients as input for the model can significantly enhance the accuracy of the model, as it would consider the current track conditions and car characteristics. However, due to time constraints and the significant effort spent to develop new features, these improvements were not incorporated in this project.

Another area that requires attention is the decision making step to determine the optimal strategy. Once the degradation rates of each tyre compound are known, finding the best combination of tyres that minimizes lap times is necessary. Given that this is an optimization problem, solutions such as Evolutionary Algorithms can be effective in identifying the optimal strategy.

5 Conclusion

Our developed solution did not yield satisfactory results, as indicated by the large error obtained. This lack of accuracy becomes problematic in a sport where differences of milliseconds can make the difference. Consequently, the current state of the project becomes unusable for Formula 1 teams. Nonetheless, it provides a foundation to build upon, particularly in a field where limited public information is available.

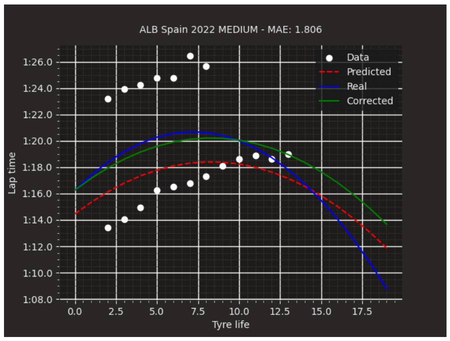
Tyre degradation is an basic but non-trivial aspect of Formula 1, and most teams face challenges in determining the optimal strategy due to the inherent uncertainty. To enhance the current results, further efforts in feature engineering are necessary. By refining the features and improving the model, this solution has the potential to become a valuable asset for Formula 1 teams.

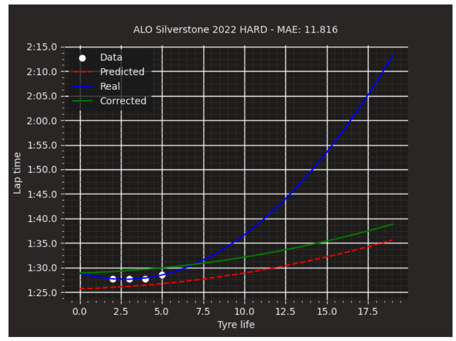
References

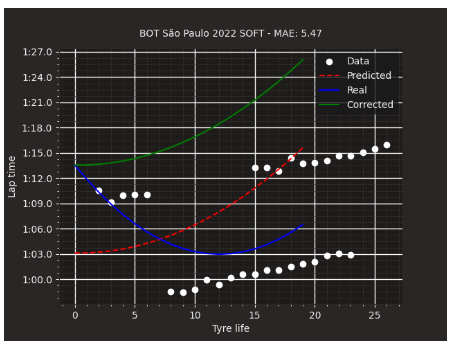
1. How tyre degradation affects race strategy, https://sbgsportssoftware.com/blog/how-tyre-degradation-affects-race-strategy
2. Duxbury, A.: How much fuel does a formula 1 car use? (Nov 2021), https://www.autosport.com/f1/news/how-much-fuel-does-a-formula-1-car-use-f1-nascar-more-compared-4980266/4980266/#:~:text=F1%20cars%20can%20use%20a,more%20lap%20time%20it%20costs.
3. Insights, T.: I've looked at 14,000,605 scenarios and max wins every time (Sep 2022), https://tracinginsights.substack.com/p/ive-looked-at-14000605-scenarios
4. Insights, T.: Why hamilton lost out to george in miami and tyre degradation analysis (May 2022), https://tracinginsights.substack.com/p/why-hamilton-lost-out-to-george-in?sd=fs
5. Rehman, M.u.: What is tyre degradation (deg) in f1? (Feb 2023), https://formulapedia.com/tyre-degradation-deg-in-f1/
6. de Technologie Michelin, S.: The tyre grip, http://docenti.ing.unipi.it/guiggiani-m/Michelin_Tire_Grip.pdf

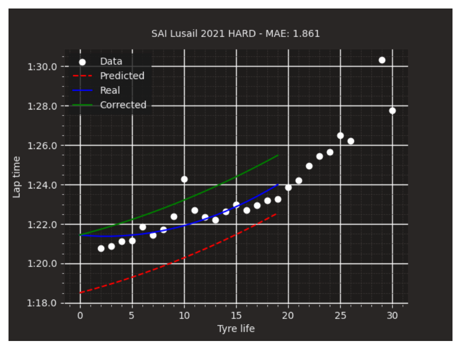
A Results from the Random Forest model

Some random results of the Random Forest model.









B Results from the Neural Network

Some random results of the Neural Network model.