### Appendix A: Literature Review of “Online Planning for F1 Race Strategy Identification”

The paper by Piccinotti et al. looks at planning race strategies for Formula 1. It discusses the challenges of modelling the race environment and the need for online planning. The authors propose an open-loop planning approach with QL updates. They evaluate their approach on a modified racing simulator and show that it outperforms other planning algorithms.

The paper presents a new method for planning Formula 1 race strategies. It is based on Monte Carlo Tree Search (MCTS), which is a technique for finding the best sequence of actions in a complex environment. The authors propose an open-loop planning approach that uses TD updates to address the high variance of the returns. They evaluate their approach on a racing simulator and show that it outperforms other planning algorithms and the real strategies applied by the racing teams.

In the Open-Loop Planning for Race Strategy section, the authors discuss the open-loop planning approach used to tackle the race strategy problem. They detail the algorithm employed, which performs multiple iterations of planning until the planning budget is reached.

A white text with black text

Description automatically generated with medium confidence A math equations and formulas

Description automatically generated with medium confidence

Algorithm 1: Q-Learning Open Loop Planning (Piccinotti et al., 2021)

At each search iteration, the UCB selection is performed at each node of the tree until a leaf-node is reached, from which a rollout is performed. The rollout gives an initial (noisy) estimation of the node value. Next, QL updates are recursively employed up the tree, updating the node and action values and counts. This means that the initial noisy back-up value given by the rollout, even though it is stored in the leaf node, might not make its way up to the root, since at each node the max operator is employed to define the target value. When all the children of a node have not been explored yet, for the Q-learning update, the max operator is only applied on the visited nodes, disregarding the unexplored actions.

In other words, the open-loop planning approach works by repeatedly searching for the best sequence of actions to take, given the current state of the race. The search is performed using a technique called Monte Carlo Tree Search (MCTS), which is a method for finding the best sequence of actions in a complex environment. At each step of the search, MCTS selects the action that is most likely to lead to a good outcome. The outcome of an action is estimated by performing a rollout, which is a simulation of the race from the current state to the end.

The authors also use a technique called Q-learning (QL) updates to learn from experience. QL updates allow the algorithm to learn which actions are more likely to lead to good outcomes.

The open-loop planning approach was able to find race strategies that were faster than the strategies used by the racing teams in the 2020 Formula 1 season. The method was also able to find strategies that were more consistent, meaning that they were less likely to lead to a retirement.

In the Rollout and Opponent Policies section of the paper, the authors discuss the rollout policies and opponent policies used in their open-loop planning approach.

* Rollout policies: The rollout policies are used to simulate the race from the current state to the end. The authors propose two rollout policies:
  + Random rollout policy: The random rollout policy selects actions randomly.
  + Opponent-aware rollout policy: The opponent-aware rollout policy uses a model of the opponent's behaviour to select actions.

Piccinotti et al. show that the opponent-aware rollout policy outperforms the random rollout policy in terms of race time.

* Opponent policies: The opponent policies are used to model the behaviour of the opponents. The authors propose two opponent policies:
  + Static opponent policy: The static opponent policy assumes that the opponents will follow the same strategy throughout the race.
  + Dynamic opponent policy: The dynamic opponent policy updates its model of the opponents' behaviours based on their actions during the race.

Piccinotti et al. show that the dynamic opponent policy outperforms the static opponent policy and the random rollout policy in terms of race time. This suggests that it is important to model the behaviour of the opponents when planning a race strategy. Overall, the rollout policies and opponent policies play an important role in the open-loop planning approach. The rollout policies are used to simulate the race from the current state to the end, and the opponent policies are used to model the behaviour of the opponents.

Table 1. Race Time Improvement over Original Team Strategies (recorded by ESPN) by using QL-OL UCT

Specifically, the results show that the proposed method was able to find race strategies that were faster than the strategies used by the racing teams in the 2020 Formula 1 season. The method was also able to find strategies that were more consistent, meaning that they were less likely to lead to a retirement. Overall, the results of the paper suggest that the proposed method is a promising new approach for planning Formula 1 race strategies.

#### Appendix B: Results of MCTS, Rigorously Flexible Q-Learning Open Loop Planning

A graph of a number of blue bars

Description automatically generated with medium confidenceA screenshot of a black screen

Description automatically generated

Diagram 6: Histogram of Race Time Improvement and General Descriptive statistics

From the above histogram plot, generally, we can visualize what is the distribution of improvement that RF-QLOP Planning achieves over the single-action agent. Our model is able to reduce the race time with a median of 825 seconds across all types of conditions with a constraint of radius from 600m to 1200m. We can see that distribution is slightly skewed to the right and has the most occurrences of race time reduction by 750 seconds to 775 seconds.

A screenshot of a computer

Description automatically generated

A table of numbers and letters

Description automatically generated with medium confidence

**A table with numbers and letters

Description automatically generated**

**A table with numbers and letters

Description automatically generated**

**A screenshot of a table

Description automatically generated**

Diagram 7: Breakdown of MCTS Results by Weather Condition

From the breakdown in the above diagram, we can see that the model is generalizable across all types of weathers, and does better in more distinct weather types, such as Dry or 100% Wet conditions. In less distinct weather conditions, such as 20% Wet, 40% Wet and 80% Wet, we see more variance in performance.

However, the performance begin to drop substantially across the board as the track radius increases. The expected range of improvement across all possible conditions would be from 600 seconds to 1100 seconds for track radius of 600 to 1200m.

# Appendix C: Rollout Policies and their Details

**Random rollout policy:** The random rollout policy simply selects actions randomly. This policy is simple to implement, but it is not very effective.

**Opponent-aware rollout policy:** The opponent-aware rollout policy uses a model of the opponent's behavior to select actions. This model is trained on data from previous races. The opponent-aware rollout policy is more effective than the random rollout policy, but it is more difficult to implement.

**Static opponent policy:** The static opponent policy assumes that the opponents will follow the same strategy throughout the race. This policy is simple to implement, but it is not very effective, as the opponents are likely to adjust their strategy during the race.

**Dynamic opponent policy:** The dynamic opponent policy updates its model of the opponents' behavior based on their actions during the race. This policy is more effective than the static opponent policy, but it is more difficult to implement.

# Appendix D: Monte Carlo Tree Search Results – Time Comparison with Single Action

A screenshot of a computer

Description automatically generated

## Appendix E: Monte Carlo Tree Search Results: Time comparison with Single Action - SORTED

A screenshot of a computer

Description automatically generated

## Appendix F: Monte Carlo Tree Search: Tyre changes over 5 iterations

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA graph of a bar code

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

**Tyre Change over Time – 600m Radius with more frequent weather changes**

A screenshot of a graph

Description automatically generated A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

**Tyre Change over Time – 700m Radius with more frequent weather changes**A line graph with numbers

Description automatically generated with medium confidenceA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

**Tyre Change over Time – 800m Radius with more frequent weather changes**A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

**Tyre Change over Time – 900m Radius with more frequent weather changes**A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

**Tyre Change over Time – 1000m Radius with more frequent weather changes**A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

**Tyre Change over Time – 1100m Radius with more frequent weather changes**A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

**Tyre Change over Time – 1200m Radius with more frequent weather changes**A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A graph of a tree

Description automatically generated with medium confidenceA graph of a barcode

Description automatically generated with medium confidence