PhD-FSTM-2024-031
The Faculty of Science, Technology and Medicine

DISSERTATION

Defence held on 26/04/2024 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L’UNIVERSITÉ DU LUXEMBOURG

EN SCIENCES DE L’INGÉNIEUR

by

Federico BIGI
Born on 23 February 1995 in Rome, Italy

MAINTENANCE IMPACT IN FREIGHT TRAIN OPERATIONS

Dissertation defense committee

Prof. Francesco Viti, Dissertation Supervisor
Professor, University of Luxembourg, Faculty of Science, Technology and Medicine.
Prof. Joachim Arts, Chairman
Professor, University of Luxembourg, Faculty of Law, Economics and Finance.
Prof. Andrea D'Ariano
Professor, University of Roma Tre, Dipartimento di Ingegneria civile, informatica e delle tecnologie aeronautiche.
Prof. Paola Pellegrini
Professor, Université Gustave Eiffel, Département LÉOST (Laboratoire Électronique Ondes et Signaux pour les Transports) - COSYS (Composants et Systèmes).
Prof. Nikola Bešinović
Professor, Technical University of Dresden, Faculty of Transport and Traffic Sciences.

To The Skies From A Hillside,
Maybeshewill

Dissertation Defense Committee

Prof. Joachim Arts

Professor, University of Luxembourg, Faculty of Law, Economics and Finance.

Prof. Andrea D'Ariano

Professor, University of Roma Tre, Dipartimento di Ingegneria civile, informatica e delle tecnologie aeronautiche.

Prof. Paola Pellegrini

Professor, Université Gustave Eiffel, Département LÉOST (Laboratoire Électronique Ondes et Signaux pour les Transports) - COSYS (Composants et Systèmes).

Prof. Nikola Bešinović

Professor, Technical University of Dresden, Faculty of Transport and Traffic Sciences.

Prof. Francesco Viti, Dissertation Supervisor

Professor, University of Luxembourg, Faculty of Science, Technology and Medicine.

Summary

Europe aims to reach carbon neutrality by 2050, imposing a radical transformation of the continent in every aspect. Freight transport, in this context, accounts from a carbon footprint perspective for 27% of total transport CO2 emissions, with 95% of these produced by cars, vans, trucks, and buses, i.e., road transport. As the freight market is expected to grow in Europe by 50% by 2050 from the level of 1990, European Union seeks to double the use of freight rail transportation in the next 30 years. This is not a trivial task, as freight rail transportation comes with specific problems, such as the coexistence in the railway network with passenger trains which leads, in some cases, to average delays of up to 11 hours. Therefore, to ensure punctual departures, and optimize the management of the freight railway system, among the possible solutions, we have chosen to focus on the optimization of internal freight railway operations as our primary approach in this thesis. Among the different operations within the shunting yard, the specific stations where freight trains are parked, loaded, and dispatched to their destination, we deal with the problem of shunting optimization. Shunting operations are defined as the movement of rolling stock within a specific station, and are essential for ensuring the smooth operation of activities within the shunting yard.

The goal of this dissertation is to implement maintenance consideration within freight train operations, specifically focusing on the impact that maintenance has on shunting operations. The primary objective of this research is to explore how maintenance operations impact shunting operations and, by extension, the overall performance of freight rail systems. Shunting operations are expensive, time and resource-consuming. Still, maintenance scheduling and operations impact them, as maintenance creates unavailability for the demand to be fulfilled, leading to additional shunting, and potentially causing delays and cancellations. In this context, we identify the role of maintenance, framing the problem from both strategic and tactical dimensions, assessing the

impact's magnitude, and proposing solutions to improve freight rail management.

Our methodological approach includes the development of two Mixed Integer Linear Programming (MILP) models and Machine Learning (ML) techniques to include in our considerations both mileage and condition-based maintenance. These models are designed to incorporate maintenance requirements into the daily operational planning of freight trains, aiming to improve system reliability, cost efficiency, and rolling stock management, while reducing delays and cancellations.

Key findings of this research reveal that the lack of integration between maintenance planning and operational scheduling from a strategic point of view can lead to significant underestimations of operational needs and system performance. We demonstrate that shunting policies, defined as the criteria to choose which wagon to choose for a service, when accounting for mileage-based maintenance considerations can substantially improve the efficiency of shunting yards and fleet management. Furthermore, we developed a machine learning model for condition-based maintenance prediction, which coupled with a tactical model enables more informed decision-making based on risk assessment, leading to more resilient and efficient freight rail operations.

Acknowledgments

I always found the Acknowledgments the hardest part to make in between bachelor, master, and PhD theses, especially because when we write this, we try to make you think that the order does not matter, but in the end it always does! Therefore, for whoever might encounter this issue, I recommend this link, which can at least give you a structure to follow.

First, I would like to thank all my dissertation committee members for their insightful feedback and support. Your expert advice and constructive criticism have been instrumental in refining my research. Your guidance has not only shaped my work but also my perspective as a researcher. Thank you for your contributions to my academic growth.

I would also to personally thank all my CET members: my main supervisor, Prof. Francesco Viti, for his role as both a mentor, supervisor, and colleague with whom to share ideas, jokes, and amazing moments together. If it was not for your bravery to let me follow the path I wanted to follow, probably I would not be as happy as I am doing the things that I love every day for work. I would like to thank also my co-supervisors Prof. Andrea d'Ariano personally. He was my guide from the master thesis, and kept following me throughout my PhD. Without you, I would not be here today, as you opened me the roads that let me sit where I am standing now, and for this I would be never thankful enough. I would like to personally thank also to Prof. Joachim Arts, whose mentorship, feedback, and insights made me grow professionally in a way I could not have thought of at the beginning of this adventure. You taught me the importance of being rigorous, on point and clear (which is something I will still have to master), and I want to think that I am on the good road for this. I would like to thank Dr. Juan Pineda. Man, without your guidance and help, from both the professional and the personal side, I would be lost by now, but look, I did it![1]

[1]NB: this is a huge jinx, as today (12/03/2024) there is still 1 month and a half before the defense.

I would also like to thank (hopefully by now) Dr. Tommaso Bosi, my colleague and companion for this adventure. We cooperate in more than half of this thesis, and you know we're going to keep on going through this road. We finally achieved our goal, and to many more to go. A special thanks go to the CFL Team, whose invaluable support in this ANTOINE project made us reach goals that we could not have expected the moment we started. Finally, I would like to thank all the Mobilab Team, actual and alumni. These three years and a half were the best of my life professionally speaking, and if it was not also for you I am pretty sure it would have not ended like this!

Now, I would like to start by thanking all my friends and family members. First, I would like to thank my friends from high school. I always find it amazing that after 14 years we're still so close, and how every one of us is growing personally and professionally at a rate that we were only dreaming of when starting the first year of high school, but now I am officially and by Luxembourgish law the most intelligent in the group. Next, I would like to thank all my friends from my bachelor's and master's university period. We keep on acting like kids, but we're all achieving amazing goals. Then, I would like to thank all my friends from Luxembourg, who made me have an amazing time and fun, making me wish that we could do it forever.

Per quanto riguarda la famiglia, volevo ringraziarvi tutti insieme dato che siete, e sarete sempre la mia famiglia. Non vi cito personalmente perche' per me siete tutti fondamentali, e un ordine sarebbe difficile e sbagliato. Mi mancate un sacco quando sono qui, ma sapete bene quanto sono felice qui, e so quanto felici siete di sentirmi felice. Perfortuna Roma e' piu vicina di quello che si pensa. Una piccola digressione per Nonno, perche' tanto lo sai che lo devo fare sempre, come ho fatto in ogni tesi. Grazie per tutti i suggerimenti, consigli, discussioni che facciamo insieme. Ci copriamo le spalle a vicenda no?

Finally, I would like to thank the two most important people in my life, Simone and Simona. Fratello, ce l'ho fatta, ed e' l'unica cosa che sai che devo dirti. Grazie per questi 29 anni passati insieme, per ogni discussione e per avermi fatto diventare la persona paziente che sono, considerando che devo combattere con una persona testarda come te. Potrei farti un acknowledgment a parte e lo sai, ma come al solito non dobbiamo dirci niente io e te. And Lyubov Moya, you know as well that there will never be enough pages in a Latex file and in my vocabulary that could show how much I love you, so as every time in our life I will keep showing you my gratitude through my actions. You made me believe in myself and made me realize that I can really be who I am without any fear, and now I am the best version of myself, and all is thanks to you.

Who is left? Obviously myself. Thank you Federico of the past for always being smarter than Federico of the future.

Contents

List of Figures **XVII**

List of Tables **XIX**

1 Introduction **1**

1.1 Context and Motivation . . . 1

1.2 Objective and Scope . . . 8

1.3 Thesis Contribution . . . 10

1.4 Thesis Outline . . . 11

2 Literature Review **13**

2.1 Shunting Yards . . . 14

2.2 Shunting Operation . . . 15

2.2.1 RSP . . . 18

2.2.2 TUSP . . . 21

2.2.3 Integration of RSP and TUSP . . . 25

2.3 Maintenance in Railways . . . 26

2.3.1 Mileage and Condition-Based Maintenance . . . 26

2.3.2 Integration of Maintenance within the RSP and TUSP Problem . . . 29

2.4 Modelling of Shunting Yard and Shunting Operations in Freight Train Systems . . . 31

2.4.1 Types of Wagon Load . . . 31

2.4.2 Classification Problem in Shunting Yards . . . 32

2.4.3 Integrated Modeling of Yards and Network . . . 33

2.5 Research Challenges 35

3 Research Objectives **38**

3.1 RO1: Analysis and Assessment of Maintenance Integration in Shunting Yard Operations 38

3.2 RO2: Identification of the Most Relevant KPIs Required for Strategic and Tactical Planning 40

3.3 RO3: Development of a Hybrid Modeling Approach for Long-Term Assessment through MILP and Simulation 43

3.4 RO4: Data-Driven Modeling for Condition-Based Maintenance and Unplanned Disruption, and its Integration in MILP Modeling through a Risk-Management Approach 44

4 Simulation and MILP Modeling for the Shunt-In/Shunt-Out Problem (SISO) **47**

4.1 Introduction 47

4.2 The Shunt-In Shunt-Out Problem (SISO) 48

4.3 Assumptions 49

4.3.1 Mandatory Shunts 50

4.3.2 Computing Time to Shunt-In 51

4.4 Mathematical Model 53

4.4.1 Shunt-In/Shunt-Out (SISO) Model 56

4.5 Shunt-In Policies 61

4.5.1 MIN 62

4.5.2 AVG Long - Short (AVG L-S) 62

4.5.3 NCLD 63

4.5.4 Reserving 64

4.5.5 Random Policy 65

4.6 Simulation 65

4.6.1 Interaction between the Mathematical Model and the Simulation Environment 66

4.6.2 Input Requirements 67

4.6.3 Shunting Yard Areas 68

4.7 Case Study 69

4.7.1 Case Study Parameters 71

4.7.2 No - Maintenance Scenario . . . 71
4.8 Results and KPIs . . . 72
4.8.1 No Maintenance Scenario . . . 74
4.8.2 Maintenance Constraint Scenario . . . 75
4.9 Discussion . . . 76
4.9.1 Assessing the Impact of the Maintenance Constraint . . . 76
4.9.2 Addressing the Impact of the Maintenance Constraint . . . 79
4.10 Summary . . . 84

5 ML model and MILP Modelling with Condition-Based Maintenance Integration 85
5.1 Introduction . . . 85
5.2 ML Model . . . 87
5.2.1 Data Collection, Data Pre-Processing and Feature Engineering . . . 88
5.2.2 Selection of the Model . . . 92
5.2.3 Performance Assessment and Probability Threshold in the TDT model . . . 98
5.3 MILP Modelling with Condition-Based Maintenance Integration . . . 102
5.3.1 Risk Assessment . . . 103
5.3.2 Operational Research Model . . . 107
5.3.3 Comparison with the Classical Approach . . . 115
5.4 Case Study and Result . . . 115
5.4.1 Case Study . . . 115
5.4.2 Result . . . 117
5.5 Summary . . . 123

6 Conclusion 124
6.1 Answer to the Research Questions . . . 124
6.1.1 Main Findings . . . 127
6.2 Future Research . . . 128

7 Bibliography 129

A Shunt-In/Shunt-Out (SISO) Model Formulation 141

B Event-Based Simulation Framework (EBSF) Flowchart **144**

C Data Joining and Feature Engineering for the ML model - Flowcharts **147**

D MILP Model with Condition-Based Maintenance Integration **150**

List of Figures

1.1 Roadmap Europe Emission . . . 2
1.2 Modal Split Freight Europe 2021 . . . 3
1.3 CO2 and Trade values Freight Transport . . . 4

2.1 Scheme of a hump yard . . . 14
2.2 Railway operations planning taxonomy . . . 17
2.3 Taxonomy for the RSP problem . . . 20
2.4 TUSP Taxonomy . . . 24
2.5 Wheel damage example . . . 28

3.1 Framework RCs and ROs . . . 46

4.1 Simple/Double wagon . . . 50
4.2 Simulation and MILP integration . . . 66
4.3 Flowchart operations train simulator . . . 69
4.4 Bettemburg Eurohub Sud Terminal . . . 70
4.5 Annual Mileage Distribution - Case study MILP + Simulation . . . 77
4.6 AVG mileage available - Case study MILP + Simulation . . . 80
4.7 AVG mileage vs Train Dep - Case study MILP + Simulation . . . 82

5.1 Correlation matrix features ML . . . 89
5.2 Learning Curve DT model . . . 95
5.3 TDT learning curve and Feature Importance . . . 97
5.4 TPR vs TNR probability threshold . . . 101

5.5 Conceptual framework risk assessment 106
5.6 Risk (a) and Operational Cost (b) - Case study Cond. MILP 118
5.7 Cancellation Rate - Case study Cond. MILP 119
5.8 Cancellation Rate - Case study Cond. MILP 120
5.9 Cancellation Rate - Case study Cond. MILP 121
5.10 Average Dep. - Case study Cond. MILP 122

B.1 EBSF flowchart 145
B.2 EBSF flowchart - 2 146

C.1 Data joining phase - ML 148
C.2 Feature Engineering - ML 149

List of Tables

1.2 Potential Economic Impacts Shunting Yards 6

2.1 References and Respective Solving Methods 30

3.1 KPI Analysed in the Literature per Planning Phase 40

4.2 Comparison No-Maintenance and Maintenance Scenario 73

5.2 Gravity Accident Explanation 88
5.4 Explanatory Table of Selected Features 90
5.5 ML Model Tested 93
5.6 Initial DT Results 94
5.7 Tuned DT Results 96
5.8 Confusion Matrix Definition 98
5.9 Performance Metrics TDT 100
5.10 Nomenclature of the Deterministic Risk model. 110
5.11 Results - Case study Cond. MILP 117

Chapter 1

Introduction

When the war of the beasts brings about the world's end,
The goddess descends from the sky.

Loveless, Prologue

1.1 Context and Motivation

In March 2011, the European Commission published its *Roadmap for moving to a competitive low carbon economy in 2050* [30], a document defining the framework to implement a competitive low carbon footprint economy, whose aim is the reduction of all its *domestic* emission by 80% compared to 1990. Domestic is here defined as the real internal reductions of EU emissions and not offsetting through the carbon market. Figure 1.1 shows the projected reduction between different sectors of CO2 Emissions.

In the same month, the European Commission published an ambitious transport strategy named Transport 2050 [31], aimed at enhancing mobility, promoting growth and employment, while significantly reducing Europe's oil dependency and cutting transport-related carbon emissions by 60% by 2050. Key objectives of this comprehensive plan include:

- Eliminating conventionally-fuelled cars in urban areas by 2050.
- Increasing sustainable, low carbon fuels in aviation to 40%, and reducing shipping emissions

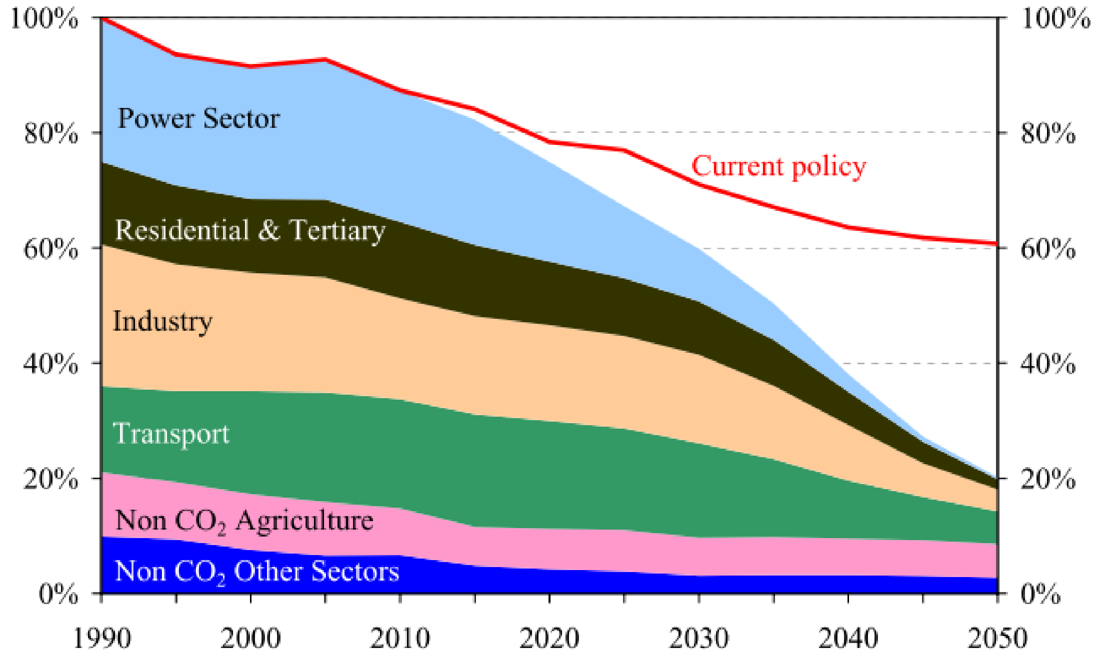


Figure 1.1: EU GHG emissions towards an 80% domestic reduction (100% =1990). Source: [30]

by at least 40%.

- Shifting 50% of medium-distance intercity passenger and freight transport from road to rail and waterborne transport.
- Implementing a Single European Transport Area to remove barriers and integrate different transport modes.

For intercity and urban travel, the plan emphasizes a significant move towards rail and waterborne transport, modernizing infrastructure, and enhancing multimodal connections. For long-distance and intercontinental freight, the strategy relies on more efficient and lower-emission air, passenger rail, maritime transport, and freight rail transport. This is also supported by huge investments in the railway infrastructure in Europe, which will have to account for the new expected growth in demand for both freight and passenger trains.

The Transport 2050 proposal was followed up by the European Green Deal (EGD) [33], presented on the 11th of December 2019, which upped the stakes bringing the proposed cut of transport emission to 90% by 2050. Looking specifically at freight transport, the European Commission's strategy aims to mitigate its environmental impact, as this sector is the backbone of the EU's Single Market. The freight transport keeps supermarkets, factories, and pharmacies stocked, enabling European companies to sell their products across the continent and beyond. In 2020, 6 million people worked in the EU freight sector in 2020, with freight transport in the EU responsible for an annual turnover of €938 billion. This market, from a carbon footprint perspective, accounts for 27% of total EU emissions of transport CO2 emissions, as per data from the European Environment Agency (EEA) [8]. Of this percentage, 95% of these are produced by cars, vans, trucks, and buses, i.e., road transport. This is an impressive number, considering that road freight transport accounts for only 25% of the overall tonne-km transported in 2021. Only 5.4% delegated to rail freight transportation, as depicted in Figure 1.2, [21].

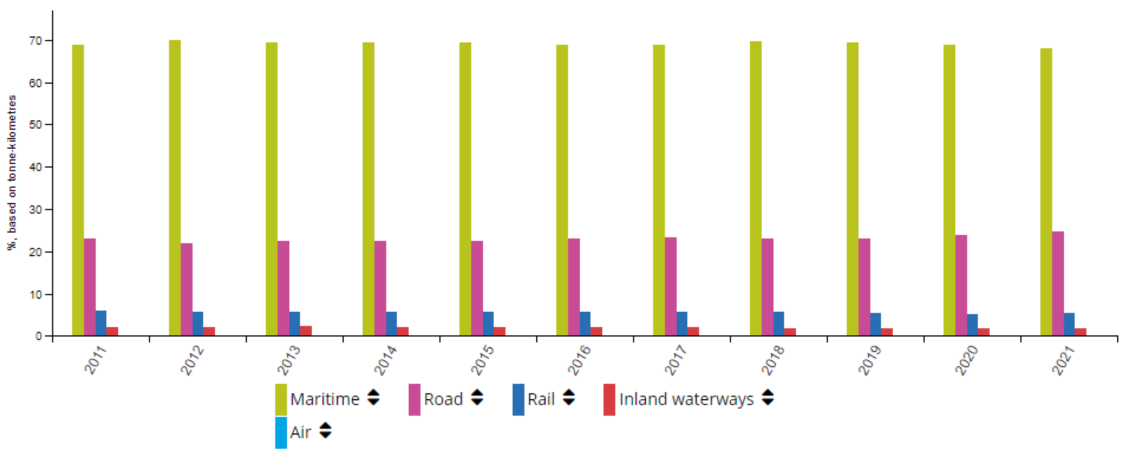
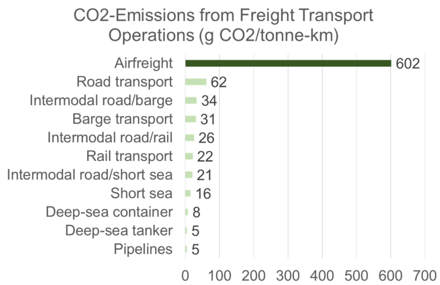


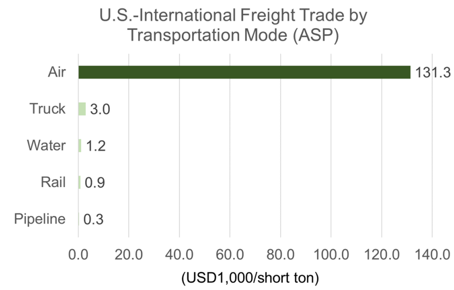
Figure 1.2: Modal Split Freight as per 2021. Source: [21]

With the freight market expected to grow in Europe by 25% by 2030 and 50% by 2050 from the level of 1990, and as part of its goal for reaching the declared carbon neutrality goals, the European Union aims to double the use of freight rail transportation in the next 30 years [19]. This sector can play a crucial role in supporting the achievement of these environmental challenges, reducing congestion on roads and costs related to this sector. This is because freight train is one of the least pollutant modes of transporting freight, Figure 1.3a, while having the second lowest cost of operations among the different options, Figure 1.3b.

Nonetheless, freight rail transportation also has its inherent challenges: costs, logistic complex-



(a) CO2 - Emissions from Freight Transport Operations (g CO2/tonne-km). Source: ECTA



(b) Unit trade value (import/export unit cost) for each means of transportation for USA in 2021. Source: Bureau of Transportation Statistics

Figure 1.3: Comparative Analysis of CO2 Emissions (a) and Trade Values in Freight Transport (b)

ities, and the coexistence with the passenger railway market on the same infrastructure, which are on a different scale compared to road transport.

The coexistence of the passenger railway, with its expected growth in traffic by 34% by 2030 and 51% by 2050 ([29]), becomes an important challenge for improving freight railway attractiveness.

Freight rail transport has one of the lowest priorities in the railway network [75], meaning that when there is a conflict in a railway section, meaning two trains arriving at the same time in the same section, train dispatchers decide which train is going to pass first to solve the conflict based on some predetermined priorities. This increases the likelihood that freight trains have to wait for the other train to pass, reducing the quality of service and decreasing its punctuality, bringing increased costs for this sector. Punctuality in trains can be improved through the increase of what is defined as the reserved times, defined as the time for which a section of the railways is booked for a specific service, whether it be passenger or freight. However, larger reserved times require first a larger capacity of the network, which would come with time with new investments, but reduces the capacity utilization of lines, which is a trade-off that must be considered [23]. With the expectation of doubling freight and passenger traffic by 2050 and the prioritization system in action in Europe, reducing late departures can be a key strategic lever for improving train punctuality. Reducing late departures, especially for freight, increases the probability of the trains meeting the requested appointments at each section, reducing the likelihood of a conflict and therefore of propagated delays.

This is not a trivial task, as delays on departure are influenced both by the network and the operations to be performed on the train. The Polish Office of Rail Transport states that for the period July-September 2023 the average punctuality rate of their freight trains was 51.89%, with an average delay of 679 minutes (around 11 hours), [78].

These delays, being propagated from the network to the station, increase the probability of incurring delays and potential cancellation and rescheduling of service, impacting the quality of the service provided and inducing increased costs. These delays create a vicious circle, where the delays in arrival create delays in starting the necessary operations that are needed to make the train available for the next service. The operations performed on the train itself can also be a source of delay, which usually are tightly scheduled to cope with the limited capacity of the stations.

Economic Impact	**Description**
Fleet Requirements	Late arrivals of a train may necessitate additional wagons or rolling stock to meet other scheduled services, increasing capital and operational costs.
Cancellation Fees	Delays could lead to cancellations of subsequent services, incurring fees and reducing customer satisfaction.
Idle Time	Rescheduling due to delays can lead to increased idle time for wagons and locomotives, reducing asset utilization and increasing costs.
Labor Costs	Additional workforce required to handle the backlog and rescheduling, leading to increased labor costs.
Service Quality	Delays can cascade through the system, affecting service quality and potentially leading to loss of business.
Handling Costs	Late arrivals might require extra shunting moves and handling, increasing operational costs.
Buffer Stocks	Customers may need to maintain higher inventory levels to accommodate unreliable service, which could lead to increased storage costs and reduced competitiveness.
Contractual Penalties	Customers may impose fines or seek damages for late deliveries impacting their operations.
Fuel Costs	Inefficiencies and additional movements in the yard can lead to higher fuel consumption.
Wear and Tear	Additional and unexpected operations can accelerate the wear and tear on wagons and infrastructure.
Network Capacity	Delays can affect the overall network capacity, leading to a broader economic impact on the train operating company's ability to transport goods.

Table 1.2: Table of potential economic impacts of a delayed train on shunting yard operations.

The impact of late-arriving trains can have multiple effects, such as the ones presented in Table 1.2: increased fleet requirements to costs related to fees, overtime labor, and increased consumption of fuel to cite some. While some operations in the yard that are capacity-constrained can be improved only by increasing the infrastructure of the station, such as the loading and unloading operations, others such as shunting operations can only be optimized and are demand-dependent. According to [17], the shunting operation refers to the movement of one or more rolling stocks within a shunting yard, which is a specific station where railway coaches are maneuvered. These are operations that lie in between strategic, tactical, and operational planning, as they are affected by long-term investment (shunting yard, fleet), middle-term management (assignment of wagons to a service), and short-term operations (physical routing of the wagons around the yard). CFL Multimodal, the Luxembourgish Railway Multimodal Freight Company, stated that 20% of their delays and cancellations are directly caused by shunting operations inefficiency. This is due to the complexity of this problem, which involves the assignment of wagons to a specific service, as well as routing. As stated by CFL, a single shunting operation can last up to 15 minutes, costing around 350 €. Given that trains typically undergo several shunting maneuvers as needed, it becomes clear that the efficient performance of these operations is a critical factor in reducing late departures of outbound freight trains.

When it comes to why shunting operations are performed, this happens for different reasons: arranging wagons to prepare them for a service, parking them inside the station waiting for the next service, and removing them from their position due to maintenance operations. Regarding the formers, the assignment of rolling stock without considering any wagon condition parameter (e.g. the mileage performed by each wagon, which is linked to contractual clauses and maintenance thresholds), can lead to an additional number of shunting operations to be performed in the long term. This inefficiency occurs under the assumption that one wagon or multiple wagons can be moved altogether while consuming the same time and cost [17]. As a result, a sub-optimal choice of wagons for a train may result in repeated, unnecessary shunting operations, affecting the overall performance of freight train management. Moreover, in terms of cost, wagon maintenance is itself a major cost factor for a freight rail company [48], even though its management is always focused on a short-term period rather than looking at the long-term impact of scheduling these operations. This can lead to unavailability as the fleet is being used in a sub-optimal way, leading to the creation of delays and cancellations to gather the wagons mentioned above. Furthermore, maintenance is usually not integrated into the train composition process but is instead solved as a separate problem [36], usually without considering how this may affect long-term operations.

1.2 Objective and Scope

In this thesis, we focus on understanding how we can optimize the shunting operations processes when including maintenance considerations within the problem. We focus on assessing how shunting is affected by the various maintenance operations to be performed, and how to optimize shunting operations such that the impact of maintenance is minimized, within an integrated framework. For this dissertation, we focus on both the strategic planning and the tactical planning of this problem. This choice comes from the intuition that, as maintenance is a periodical operation on a rolling stock, a short-term analysis could not grasp the proper impact on more strategic aspects such as fleet management. The main research question we want to answer in this thesis is as follows:

What is the impact of maintenance operations in shunting operations? How can we assess the integration of maintenance and shunting operations in freight rail management within a strategic and tactical vision?

The complexity of answering this question lies in the interdependence that the maintenance has with the shunting operation. The first requires the second, but as for the shunting definition from [17], where one or multiple rolling stocks can be moved with the same movement, the assessment of the maintenance impact becomes more complex to understand. Capturing the short-term impact of these maintenance operations on shunting, and therefore on the system performance, becomes convoluted, especially if we want to look at problems such as fleet management. Proposing instead a long-term vision can provide us with insight into how the management of the shunting operation should be directed. Another reason that motivated us to tackle this problem with this approach is the lack of literature on the strategic planning of freight train operations, which is further explored in Chapter 2.

To answer the main question, we propose 3 intermediate Research Questions (RQs) that will provide us with the methodological framework to assess, model, quantify, and address these dynamics.

RQ1: Can we assess the impact of the maintenance in freight train operations on shunting operation and on the system performance?

This first question concerns whether there is a direct correlation between maintenance in freight train operations and shunting operations, and how this affects the whole system. Specifically, we want to understand which are the dimensions that are worth exploring in terms of the impact of neglecting maintenance, and the underestimation that comes with it. It is rational to conclude that this correlation exists, but as explained in Chapter 2, usually the maintenance scheduling problem is solved separately and is not linked to the shunting operations.

RQ2: How can we model this problem in a way that we can quantify its impact?

The second research question looks at the availability of tools and models to assess the impact of maintenance considerations within freight train operations in an integrated way. To address it, the thesis aims to develop and evaluate models that can integrate maintenance considerations within freight train operations. This involves the development of a methodological framework that can capture the complexities of maintenance activities and their effects on shunting operations, fleet utilization, and service reliability. The core of the model should integrate maintenance requirements with daily shunting operations. This involves creating methodologies that can trigger maintenance when needed (based on mileage or other wear and tear indicators) and incorporating these maintenance activities into the shunting schedules without significantly disrupting operations. The models should aim to optimize multiple objectives, mainly strategic and tactical, such as minimizing delays and cancellations, reducing operational costs, and maximizing fleet utilization.

RQ3: How can we address the shunting operation optimization problem with maintenance considerations, therefore improving the performance of the system?

The third research question looks at techniques that can address the impact of maintenance considerations within freight train operations. The focus shifts towards implementing the insights and models developed in RQ2 to create strategies that enhance system performance under these considerations. This includes developing practical approaches that facilitate the integration of maintenance considerations. Many tools can be used in this matter:

- Developing strategies for operations that look at long-term approaches to exploit the shunting definition, such that we can reduce the number of operations performed, therefore improving the system mechanics.

- Use of optimization models to allocate resources (e.g., rolling stock) more efficiently and consider maintenance constraints, ensuring that maintenance activities are carried out without compromising the availability of resources for operational needs while making more informed decisions.

- Use of data-driven machine learning models and data analytics techniques to include more precise maintenance information within the models. This involves analyzing historical data on rolling stock performance, wear and tear patterns, and failure rates to anticipate maintenance requirements before they lead to operational disruptions.

1.3 Thesis Contribution

We can divide the contribution of this dissertation into theoretical and practical:

- Theoretical:
  - We developed two Mixed Integer Linear Programming (MILP) models whose aim is to optimize shunting operations performed and that allow the assessment of the impact of maintenance operations for tactical and strategic planning: the first MILP model implements mileage-based maintenance considerations and long-term propagation of effect through policies, with the second implementing both mileage-based and condition-based maintenance information.
  - We developed four shunting policies aimed at optimizing shunting operations long-term, under both maintenance and no-maintenance considerations.
  - We developed a simulation environment to assess the long-term impact of operations in a shunting yard.
  - We developed a Binary Classification Machine Learning (ML) model to detect unplanned disruptions for rolling stocks for condition-based maintenance.
  - We developed a risk assessment framework to assess the ML model's prediction against itself, which has been implemented in the second MILP model to measure the trustworthiness of the model and allow for a more informed decision on shunting optimization under maintenance considerations.

- Practical:
  - We demonstrate that the non-integration of maintenance in operations in the shunting yard can lead to underestimation of shunting operations, fleet requirements, and fleet performance.
  - We demonstrate that different shunting policies lead to different performances of the fleet, shunting yard, and operations under both no-maintenance and maintenance considerations.
  - We highlight that for our Binary Classification Machine Learning (ML) model, the most important features for predicting the disruption of the rolling stocks based on condition-based analysis are a monthly carried weight by the wagon, seen as the TEU (Twenty-foot Equivalent Unit) count, the Actual Mileage, Journey Destination, and average monthly absolute elevation performed.
  - We demonstrate that the application of a mild input of the ML within a risk-assessment framework improves the performance of the MILP model for shunting operations for wagon fleets with lower mileage.

1.4 Thesis Outline

The manuscript is organized into six chapters, with the current one being the first. Chapter 2 reviews the state of the art, with a particular focus on the existing methodologies and models to represent shunting operations. At the end of it, we highlight four Research Challenges (RCs) given the state of the art and the proposed Research Questions (RQs). Chapter 3 explains how we are going to address each RC with the Research Objectives (ROs), which create the theoretical framework of this thesis. Chapter 4 describes the first MILP model for optimizing shunting operations and the shunting policies, paired with the description of a simulation environment. We also present a case study in which we show the efficacy of this model, together with the assessment of the magnitude of error related to the implementation of the maintenance considerations. Chapter 5 describes the Binary Classification ML model, shows its development process, and introduces the model's performance. Then, it shows the second MILP model, together with the risk-assessment framework, to introduce condition-based maintenance consideration and further improve shunting optimization. In this Chapter, we also present a case study showcasing the improvement of the performance of the model compared to the traditional implementation of the ML prediction inside the model. Chapter 6 presents conclusive thoughts and highlights some possible future research directions.

Chapter 2

Literature Review

Infinite in mystery is the gift of the Goddess,
We seek it thus, and take to the sky.

Loveless Act I

This chapter aims to cover the relevant aspects concerning the state of the art of shunting operations and shunting yards, maintenance, and how these operations are modeled in the literature. Section 2.1 discusses the different types of shunting yards, the operations performed, together with the planning phases regarding railway operations, and how these are connected to the yards. Section 2.2 presents the theoretical aspects related to shunting operation. This section also discusses, within the context of shunting operations, how maintenance services are scheduled and up to which level. The latter is further discussed in Section 2.3, which covers the criteria for assessing when a rolling stock has to go to maintenance. Section 2.4 discusses the practical aspects of both shunting operations and yards, how these are modeled for freight trains, and discusses the state-of-the-art solutions for these problems. Finally, Section 2.5 highlights the shortcomings of the current literature and proposes four research challenges that this dissertation aims to address.

2.1 Shunting Yards

A shunting yard, also called a classification or marshaling yard, is defined as a large railway yard in which wagons are organized into trains. Inbound trains are sorted and reassembled to create the desired composition of outbound trains. These shunting yards can be of different types: flat yards, which function on a level terrain, where trains are manually maneuvered for the assembly and disassembly of wagons; hump yards and gravity yards, which utilize a hill or "hump" to separate and distribute the wagons, in which these are pushed up the hump and then roll down by gravity onto their designated tracks [1]. The typical layout of a shunting yard can be roughly separated into three main areas, which all consist of a set of parallel tracks [15]: the receiving area/tracks, the classification area/tracks, and the departure area/tracks. An example scheme from [24] of a hump yard is presented in Figure 2.1.

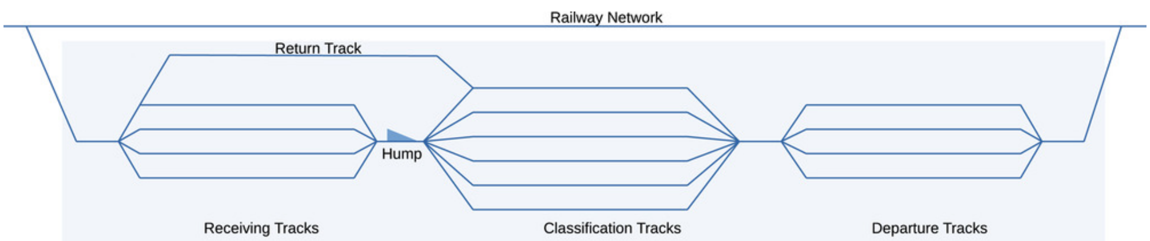


Figure 2.1: Scheme of a hump yard

In shunting yards, incoming trains are temporarily stationed for inspection and coded based on their destinations in the receiving area, while awaiting humping and disassembly. Depending on priority, these trains or subsets of their wagons are then maneuvered by shunting locomotives over the hump (in hump yards) or on a slope (in gravity yards) directed to specific tracks by switches, or moved directly through a shunting locomotive in the classification yard (in flat yards). Once the composition of the outbound train has been completed, these trains are then pulled to departure tracks to be inspected, road engines are attached, and ready for departure.

The management of shunting yards involves addressing a variety of tasks and challenges, each subject to specific constraints and objectives. While some tasks in the yard, like inspections and train assembly, can have relatively fixed durations, others significantly impact yard efficiency and require careful planning. Uncertainty in train arrivals adds complexity to decision-making. To streamline the process, some subproblems are tackled independently [14], while only in a few cases we see an integration of multiple problems together [50, 51].

To provide a more comprehensive outline of the different operations that can be performed on these yards, we present here the different phases into which these can be clustered. The research panorama is quite aligned on how to categorize the type of planning phases that exist in railway operations, as outlined by [14, 54]:

- **Strategic planning**: this phase focuses on the assessment of the current infrastructure for future train services. This is based on expected demand, rolling stock growth, environmental policies, and long-term mobility strategies. A key aspect is estimating shunting capacity needs across the network.

- **Tactical planning**: this phase involves capacity checks at individual stations, ensuring sufficient infrastructure for routing trains and parking rolling stock, adequate crew capacity for local shunting activities, and sufficient resources for cleaning rolling stock.

- **Operational planning**: this phase deals with the generation of a detailed plan for the near future, involving timetable matching of outbound and inbound trains for demand fitting, and actual routing of wagons around the shunting yard.

- **Real-time planning**: this phase deals with the re-planning for addressing disruptions that occur during operations. These disruptions can range from minor deviations, like slight delays in train services, to major issues such as infrastructure failures or train breakdowns. Both types of deviations can lead to significant disruptions in local shunting activities and overall network operations.

In this literature review, we will categorize the state of the art for the shunting operations within these phases.

2.2 Shunting Operation

While many definitions of shunting operations exist, we will adhere to the one proposed by [17]. We define shunting operations as the sorting, assembling, and disassembling of trains, and the movement of rolling stock for purposes such as maintenance, loading, or preparation for departure. As an example, assuming we have a flat yard, when rolling stock needs to be moved from the receiving yard into the shunting yards, the operation performed to move the wagon from the first area to the second area, using a shunting locomotive, is commonly referred to as a shunting operation. These operations, while being expensive and time-consuming, are usually performed for a specific reason, which is for example fulfilling a service for an outbound train that has to be created in the departure track (demand) from the available wagons parked in the classification tracks (supply). Once this matching has been performed, for each wagon or cluster of wagons that needs to be shunted, a specific routing is created within the tracks of the shunting yard to connect these rolling stocks to the actual outbound train. Thus, the shunting optimization problem, as for [41], is usually modeled through two interrelated and sequential sub-problems: the *Rolling Stock Problem* (RSP) and the *Train Unit Shunting Problem* (TUSP). The RSP addresses the planning of each wagon's service time and aims to optimize the management of rolling stocks and reduce costs, and is discussed further in detail in Section 2.2.1, while the TUSP is concerned with the routing of different rolling stocks throughout the shunting yard, parking of these in the shunting yard and all the operations that have to be performed on the rolling stock to get it ready for service, and discussed in detail in Section 2.2.2.

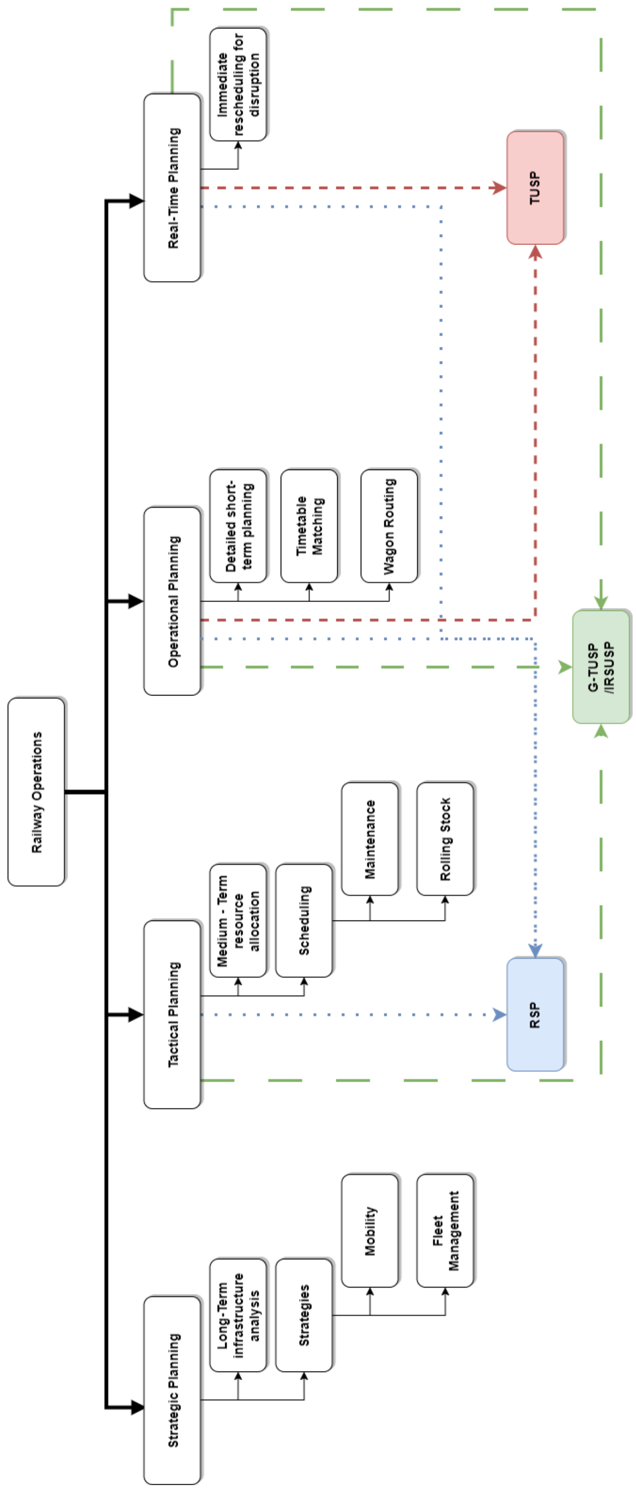


Figure 2.2: Diagram of Railway operations based on planning phases

2.2.1 RSP

The Rolling Stock Problem (RSP) is defined as the scheduling of a time of service for each rolling stock to best manage wagon and/or train units and in turn, reduce costs to supply the services or to cover the demand. It is a strategic problem that is tightly correlated with shunting operations. When the RSP assigns times of service to wagons or train units, it essentially schedules when and where these assets need to be. This schedule determines when a wagon or a train needs to be ready at a specific location to start its service. Following the scheduling and rostering decisions made during the RSP, shunting operations are required to physically arrange the rolling stock accordingly.

This problem has been divided by [18] into two sub-problems: the Rolling Stock Rostering (RSR), which focuses on the assignment of rotations for individual units of rolling stock and, simultaneously, to each train unit; the Rolling Stock Circulation (RSC), where once the roster is assigned, deals with the assignment of locomotives and carriages to the timetable services. The majority of the literature covers this problem from the passenger railways perspective, whether it be the passenger trains or metro lines. While being a more strategic problem, solution approaches for the RSP have been proposed for all the different planning phases, with the goal being to determine the fleet size or to find a feasible rolling stock schedule (or circulation) [61].

For the operational/real-time level, the literature focuses on real-time replanning. [66] deals with real-time disruption management, describing a generic framework for dealing with disruptions of railway rolling stock schedules. Their approach uses a rolling-horizon period iterative approach to deal with disruption and reassignment of rolling stock to outbound trains in passenger train railways, considering inventory status in the system in terms of wagon type and stations. [42] propose an extension of the model from [66] for real-time planning, considering the ordering of rolling stock constraints, using a MILP modeling approach to see the different train decisions throughout the horizon. [82] solves a real-time train rescheduling and rolling stock circulation problem under short-term disruption through a MILP model considering the maximum number of available rolling stock, turnaround constraints, and service connections in metro networks.

At the tactical and strategic level planning, [9] solves the rolling stock problem considering the compositions of inbound and outbound trains, solving it within the tactical planning horizon level. [89] solve the RSP for urban rail transit proposing a bi-level formulation model, solving it through a simulated-annealing-based heuristic. In their high-level model, the focus is on the trade-off between waiting time for passengers and frequency of the network, while in the lower-level model, the goal is to minimize the number of infeasible train paths. [91] deals with the tactical planning level of the rolling stock problem solving it for intercity high-speed railways (IHSR), including passenger demand for multiple stations with fixed train formation. [43] proposes a Mixed Integer Programming (MIP) modeling for solving the Departure Matching Problem (DMP), which is a subproblem of the Rolling Stock Unit Management on Railway Sites (RSUM). In their paper, they do not consider how to route, couple, and de-couple train units in the station when solving the DMP, assuming that the routing is performed in a subsequent step. [87] solves a more complex model, the rolling stock allocation and timetables problem (RATT), where for the part regarding the rolling stock they look at the inventory in the depot and the fleet size, optimizing how to allocate it. [20, 40, 55] propose a *placing-in/taking-out* approach for the train composition, which can be seen as a reformulation of the RSP problem. It was initially solved through a MILP model, and eventually evolved into a simulation-based approach with various shunting policies to reduce operating costs for the wagon entry and exit system. [92] solves the RSP problem integrating the Train Formation Plan (TFP) and Rolling Stock Scheduling (RSS) through MILP modeling, applying it for the Bejing Subway Network and minimizing both the shunting operations and the rolling stock used within the horizon considered. A schematic taxonomy for the RSP problem is provided in Figure 2.3.

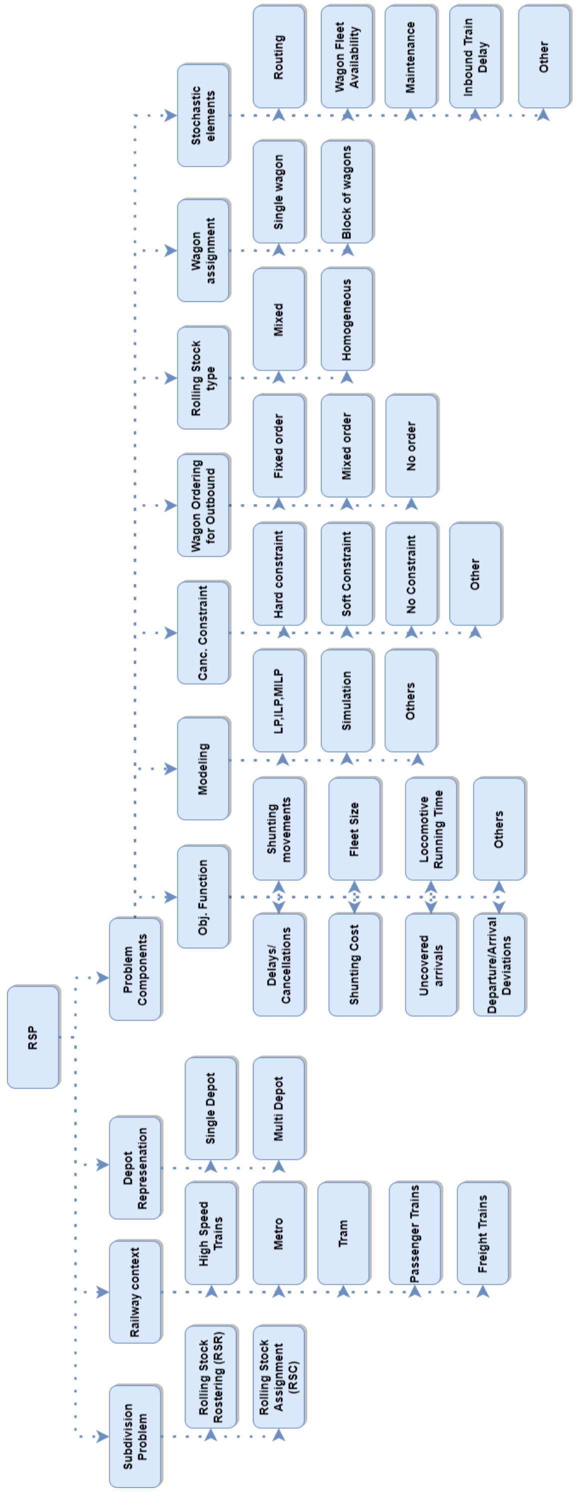


Figure 2.3: Taxonomy for the RSP problem.

2.2.2 TUSP

Once the rolling stock is scheduled for a specific service, the Train Unit Shunting Problem (TUSP) deals with the routing of the rolling stock inside the shunting yard, together with all the operations that have to be performed on the rolling stock to get it ready for service. To provide clarity given the diverse interpretations of the TUSP in the literature, our study adheres to the specific definition of TUSP as provided by [53, 54, 71]. [53, 71] defines *arriving* and *departing* shunt units as train units that respectively require to be moved (either parked or retrieved) from the shunt tracks. They formally define the Train Unit Shunting Problem (TUSP) as the problem of effectively pairing the arriving shunt units with the departing ones, and allocating these units on the shunt tracks. The goal of this problem is to minimize the total shunting costs while avoiding any occurrences of crossings. A viable solution for the TUSP consists of allocating specific tracks to the arriving shunt units and aligning them with the appropriate departing units. [54] proposes a more comprehensive modeling of the TUSP problem from a passenger train's point of view, extending the definition provided by [53] and dividing it into multiple subproblems:

- **Matching**: this is the subproblem in which the planner has to match the arriving units with the departing ones, considering position constraints and types, given a specific timetable provided in advance, with the matching coming theoretically from the RSP problem.

- **Parking**: the planner receives information about various train units, including their associated arrival or departure services. For each arriving train unit, the expected time of arrival and the specific arrival platform is predetermined. The second element of the input contains the shunt tracks and their characteristics as previously explained. The matching of arriving and departing services for train units is completed in advance as part of the previously mentioned subproblem (RSP or Matching, depending on the formulation). The output of this process is the allocation of train units to tracks, such that the rolling stock does not impair the movement of others, either during arrival or departure.

- **Routing**: the routing of train units involves their movement from platforms to the shunting yard and back, as well as transfers between arriving and departing platforms. Additional routing may be required for local tasks such as internal or external cleaning. These tasks, along with the parking of train units, generate the need for specific route planning over the station's infrastructure. If the planning of shunting operations for train units does not consider their routing to and from the shunt yards, the resulting plans are likely to be impractical, leading to unresolvable routes that can generate conflicts.

- **Cleaning**: in passenger railways, cleaning is a service that has to be performed on the rolling stocks such that it can perform a specific service. By examining the efficiency of cleaning procedures at various stations, it's possible to highlight bottlenecks in the system, from a practical point of view. Similarly, evaluating these processes in the context of different rolling stock types can also yield insights into performance variations.
- **Crew Planning**: in shunt planning, crew planning deals with the assignment of shunting crew to the tasks related to shunting, as these activities can be carried out only by specialized crews. These tasks include operating trains across railway infrastructure, coupling, and decoupling train units, and cleaning. Shunting drivers are qualified to route trains, while shunting assistants handle the coupling and train preparation. Cleaning crews, on the other hand, are tasked with maintaining train units. Although each task has a suggested start time, some, like routing train units locally, are flexible in scheduling. Shunting tasks vary in duration and are often combined into a single day's work schedule for each crew. The duration of these tasks can differ based on the station and the specific task at hand. For instance, the time taken to route a train is influenced by the route's characteristics and existing train reservations. However, if multiple units for the same train are parked in sequence on the same track, they can be coupled there instead.

Moreover, [54] provides for each subproblem a respective formulation, solving it for multiple instances and highlighting the NP-hardness of their models. [81] extends the TUSP creating Train Unit Shunting and Service (TUSS) problem, solving the model formulated in [54] through a heuristic approach. This model was extended later in [80] considering service scheduling and developing a local search approach for the train shunting and scheduling problem with the consideration of train matching, parking, and service tasks scheduling as well as train routing decisions. [85] considers a similar TUSP problem as [80] for high-speed railways including service scheduling daily maintenance, cleaning operation, and safety operational requirements, solving it through a minimum-cost multi-commodity network flow model (MCNF) formulation and Lagrangian relaxation heuristic. [47] proposes a TUSP mixed-integer formulation, demonstrating the NP-Hardness of the TUSP problem, and solving it via heuristic to minimize the weighted tardiness of each outbound train. A high-level formulation has been provided in the literature by [22]. They discuss the train marshaling problem, which is the problem of rearranging carriages in a freight train. They focused on grouping rolling stock by destination to reduce both shunting operations and the minimum number of tracks used for this rearrangement to fulfill the demand of a departing train. [84] solve special cases of the TUSP problem for dispatching trams in a depot by proposing multiple binary program models. [83] extends the TUSP approach considering length restrictions for the trains and mixed arrivals and departures, discussing also an application for the bus depot. [44] propose different algorithmic approaches for the solution of the TUSP problem using MIP modeling, solving it from the operational point of view. [37] deal with both LIFO (Last In First Out) and FIFO (First In First Out) tracks for the TUSP problem, proposing two Integer Linear Programming (ILP) formulations. The former includes arrival times, while the second one considers possible conflicts, proposing also a robust extension and a stochastic version to take into account possible delays. [65, 79] propose a formulation of the TUSP in which wagons have to be routed on a shunting yard such that maintenance tasks can be performed without collisions occurring. In this context, incoming trains have to be matched to outgoing trains, since trains of the same type can be used interchangeably. They propose a TUSP modeling based on a Multi-Agent Pathfinding (MAPF) to solve the routing problem, solving the model using Conflict-Bases Search (CBS). A schematic taxonomy for the TUSP problem is provided in Figure 2.4.

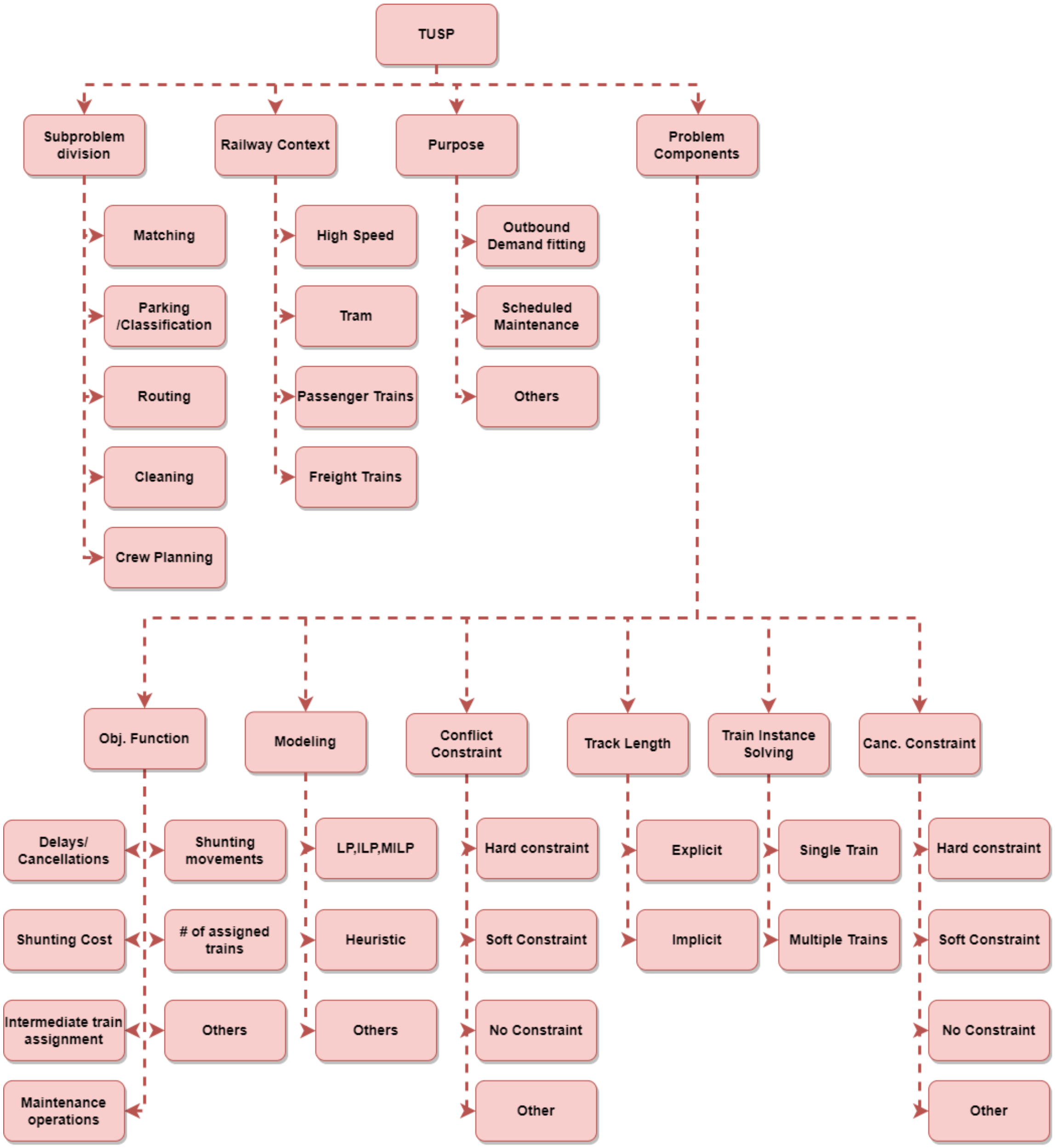


Figure 2.4: Taxonomy for the TUSP problem.

2.2.3 Integration of RSP and TUSP

Generally, the RSP is solved before the TUSP in both passenger and freight train operations. The sequencing is logical, as the TUSP depends on the outcomes of the RSP for essential inputs, particularly the alignment of rolling stock for inbound and outbound trains. Recognizing this distinction, various studies proposed methodologies to integrate these two problems. [50, 51] proposes the Generalized Train Unit Shunting problem (G-TUSP), which is composed of 4 subproblems:

- Train Matching Problem (TMP), the problem of matching arriving and departing train units, respecting constraints linked to type matching of rolling stock and schedule. This can be seen as a very similar problem to the RSP, and for this paper, it has been solved through the use of a MILP model.
- Track Allocation Problem (TAP), the problem of choosing train units location.
- Shunting Routing Problem (SRP), the problem of determining train units routing during shunting movement.
- Shunting Maintenance Problem (SMP), the problem of defining train units maintenance scheduling.

The authors state that the majority of research contributions, with the notable exception of [50], concentrate on addressing only specific portions of the four sub-problems discussed. This observation highlights a dominant trend in the literature, where a wide coverage of the entire problem scope remains fairly rare. A similar approach has been proposed by [41], which presents an integrated sequential framework approach for these RSP and TUSP, namely the *Integrated Rolling Stock and Unit Shunting Problem* (IRSUSP), highlighting the need for an integrated approach given that from their computational results the high-quality solutions for the integrated problem are obtained in instances where a conventional, sequential approach ends in infeasibility.

2.3 Maintenance in Railways

Wagons spend a significant portion of their downtime in maintenance and repair activities in the workshop, incurring overhead costs (e.g. leasing costs) and variable costs (e.g. storage costs). On average, the rolling stock remains idle 70% of the time within the shunting yard, resulting in additional storage costs, and implying potential overcapacity of the owned fleet [12]. Moreover, the maintenance operations designed for a larger fleet become inefficient, expecting to accommodate the servicing and inspection of numerous idle wagons. Maintenance scheduling is driven by either mileage, time, or condition monitoring [7, 74]. Specifically:

- Time-Based Maintenance: traditionally used method where functions like braking safety and wheel condition are evaluated on a set schedule. This approach may not always reflect the actual condition of the train, especially if the vehicle has been idle for significant periods.

- Mileage-Based Maintenance: this approach is used to perform maintenance based on the actual use of the vehicle, rather than a predetermined time frame. It helps address the wear and tear that reflects the actual usage of the train. Still, it can be challenging due to the extensive documentation required to accurately track the mileage of multiple vehicles.

- Condition-Based Monitoring: this is the most modern method, relying on automated data collection and sensors to trigger maintenance checks based on the real-world performance of the vehicle. This method uses data collection and software solutions to efficiently manage and schedule maintenance activities. These schedules must be coordinated with train utilization schedules to ensure efficient operations.

While time-based methods were traditionally used, mileage-based or condition-based maintenance is now preferred by many operators [2, 6]. This shift is supported by advances in cloud computing and enhanced data collection and storage methods, which have significantly evolved. For this thesis, our analysis will specifically cover mileage-based and condition-based maintenance approaches.

2.3.1 Mileage and Condition-Based Maintenance

Mileage-based maintenance involves defining a specific mileage threshold at which wagons have to be scheduled for maintenance. This approach is utilized for preventive maintenance scheduling purposes. Unfortunately, the literature regarding this topic is scarce, as maintenance is usually scheduled in advance. In the majority of the models analyzed in this literature review, the maintenance schedule is provided without distinguishing whether a wagon is due for maintenance based on mileage, time, or condition. This is a deliberate modeling choice, neither advantageous nor disadvantageous in itself.

[59] approaches the high-level maintenance scheduling problem study for high-speed trains with capacitated workshops. In their problem, they show different thresholds to be faced as a rolling stock goes throughout its life-cycle, focusing specifically on the high-level maintenance, which happens at 1.2 million and 2.4 million km specifically, or on average between 3 and 6 years, depending on which services the rolling stock is assigned. In their model, they consider the historical daily mileages as raw inputs, and the future daily mileages are considered as ranges to generate wider maintenance time windows for train sets, formulating the problem as a 0–1 Integer Linear Programming (ILP). In their respective studies, [56] and [58] focus on the synchronized scheduling of train operation plans in the context of mileage-based monthly maintenance. Their approach incorporates maintenance as a constraint by establishing maintenance cycles and capacity limits, where their goal was to minimize the total mileage loss for all trains over a week, treating maintenance as an integral factor in the scheduling process.

Following our discussion on mileage-based maintenance, we now shift our focus to condition-based maintenance. This approach relies on continuous monitoring and assessment of the rolling stock's condition to schedule maintenance tasks. This method is particularly effective in identifying potential breakdowns and preventing them before they occur, increasing the effectiveness of pre-emptive maintenance. To understand how condition-based maintenance is performed, we present quickly which are the most common causes of rolling-stock disruption from a physical perspective. [63] define the railway wheelsets based on three main components; the *wheel*, *axle* and *axle bearing*. Faults can develop on any of the aforementioned components, but the most common are related to wheel and axle bearing damages. For wheel disruption, [77] proposes a neural network approach for multiple-defect detection of rail wheel images. They provide a list of usual defects that were reported from the maintenance workshop of the London Underground Northern Line fleet, some of which are listed below:

- Indentation: Superficial dent caused by the wheels running over a hard object on the track.

- Rolling Contact Fatigue: Cracks caused by repeated contact stress during the rolling motion of the wheels.
- Wheel Flat (FLT) Rash that appears on both wheels caused by the wheelset skidding on the rail.
- Crazing (CRAZ) Also known as thermal cracking, it is a fracture that occurs with repeated heating and cooling of the wheel tread surface caused by traction and braking actions.

One of the most common causes of wheel damage is due to severe braking. This activity includes sudden braking, braking on steep gradients, and braking with high-weight loads, as explained in [28, 67]. [25] assess the correlation between the mechanical deterioration of wheels and winter conditions. An extreme example of wheel damage is presented in Figure 2.5.

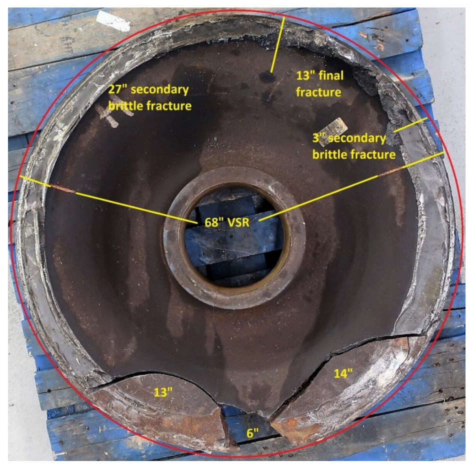


Figure 2.5: Example of a wheel damage consequence of a derailment of a freight train near Rennie, Manitoba. A minor fracture went undetected during automated wayside inspections and several inspections at terminals. The fracture expanded in opposite directions until parts of the wheel broke away, the report says. Submitted by the Transportation Safety Board.

The axle bearing is the component that transmits part of the weight of the carriage directly to the wheelset. A severe axle-bearing fault will lead to an increase in the temperature of the components due to additional heat produced by frictional interactions during rotation. [35, 68] list down some of the reasons for the failure of train axle bearings, among which fatigue is reported to be the leading cause. [62] presents a basic framework for the evaluation method, proposing an approximation of the S-N curve (number of cycles to failure, $N(S)$, when a material is repeatedly cycled through a given stress range S) highlighting the importance of the frequency of the stress to which the rolling stock is put onto. [10] propose a procedure for the railway axle risk of fatigue failure under service loading, proposing different methodologies for the fatigue assessment, highlighting many factors that can contribute to the overall load applied on the rolling stock axle bearing. An example is the train running on curved rails or in a snaking motion, which makes the bearings receive axial loads due to lateral movement.

2.3.2 Integration of Maintenance within the RSP and TUSP Problem

[64] states that maintenance scheduling in the rolling stock problem is often ignored and integrated models that simultaneously schedule maintenance tasks and railway operations are scarce in the literature. They propose in their research an integration of the maintenance scheduling within the RSP problem, specifically including the shunting operations to be performed for maintenance reasons. [93] presents a similar approach, solving RSP under maintenance requirements, proposing an extension of the MILP model presented in [42]. In their study, they address the problem of improving the integration of passenger timetabling with track maintenance scheduling at a microscopic level. Rolling stock management has been identified also by [36] as a significant cost factor for railway companies and a key determinant of service quality. To address this, they propose a two-step approach that incorporates scheduling tasks related to train services, short-term maintenance operations, and empty runs into the solution of the Rolling Stock Problem. [45] addresses both the RSP and maintenance scheduling for passenger trains, including not only preventive maintenance but also degradation based on the distance traveled, to maximize the functional life of each train. [70] propose a discrete-time model that integrates maintenance and the TUSP, solving it through a heuristic approach.

Practitioners need to consider two major issues when organizing wagon maintenance: which wagons in the inbound train require shunting due to maintenance regulations and which wagon should be replaced; and how to implement a traffic schedule that ensures traffic safety [90]. Our study focuses on the first decision problem, as most literature assumes the number, position, and type of wagons to be replaced as an input [20]. Table 2.1 provides an overview of how the shunting operations problem, between the TUSP and RSP, has been dealt with in the literature.

Name	Year	Problem	Planning Phase	Methodology	Maintenance	Note	Reference
Haahr et al.	2017	IRSUSP	Tactical	MIP	No	-	[41]
Kamenga et al.	2019	IRSUSP	Tactical	MILP	Yes	Separated, scheduled	[50]
Kamenga et al.	2021	IRSUSP	Tactical	MILP	Yes	Separated, scheduled	[51]
Alfieri et al.	2006	RSP	Tactical	LP	No	-	[9]
Nielsen et al.	2012	RSP	Real Time	Rolling Horizon	No	-	[66]
Giacco et al.	2014	RSP	Tactical	MILP	Yes	Separated, scheduled	[36]
Guo et al.	2014	RSP	Tactical	MILP	No	-	[39]
Haahr et al.	2015	RSP	Tactical	MIP	No	-	[43]
Haahr et al.	2016	RSP	Operational	MILP	No	-	[42]
Yue et al.	2017	RSP	Tactical	Bi-level model	No	-	[89]
Herr et al.	2017	RSP	Tactical	LP	Yes	Separated, scheduled	[45]
Zhong et al.	2019	RSP	Tactical	MILP	Yes	Integrated, scheduled	[93]
Zhao et al.	2020	RSP	Tactical	MILP	No	-	[91]
Li et al.	2020	RSP	Tactical	Simulation	No	-	[55]
Mira et al.	2020	RSP	Tactical	ILP	Yes	Integrated, scheduled	[64]
Wang et al.	2021	RSP	Real Time	MILP	No	-	[82]
Zhao et al.	2023	RSP	Operational	MILP	No	-	[92]
Yin et al.	2023	RSP	Tactical	MILP	No	-	[87]
Winter et al.	2000	TUSP	Operational	Binary Model	No	-	[84]
Freling	2002	TUSP	Operational	MILP	No	-	[71]
Lentik	2006	TUSP	Operational	LP	No	-	[54]
Kroon et al.	2008	TUSP	Operational	OR Model	No	-	[53]
Van Den Broek	2016	TUSP	Operational	Heuristic	No	-	[81]
Haahr et al.	2017	TUSP	Operational	MIP	No	-	[44]
Qi et al.	2017	TUSP	Operational	Heuristic	Yes	Integrated, scheduled	[70]
Gilg et al.	2018	TUSP	Operational	ILP	No	-	[37]
Van Cuilenborg	2020	TUSP	Operational	MAPF	No	-	[79]
Van Den Broek et al.	2021	TUSP	Operational	Heuristic	Yes	Integrated, scheduled	[80]
Xu et al	2022	TUSP	Operational	MCNF	Yes	Integrated, scheduled	[85]

Table 2.1: Table of references with respective solving methods.

2.4 Modelling of Shunting Yard and Shunting Operations in Freight Train Systems

2.4.1 Types of Wagon Load

To assess the impact on shunting yard delays of freight train operations, we need to first clarify which methods of transport of goods are used usually, and how these affect the operations. Among the different freight railway methods of shipping goods, the two that we will cover in this literature review are:

- Single Wagon Load (SWL): this method involves grouping railway wagons headed in the same direction at various nodes. These are then combined to form trains that travel to the next node in a hub-and-spoke network. At each of these nodes, the wagons often need to be sorted, grouped, and regrouped based on their destinations and the composition of outbound trains [72]. The shunting process in SWL is more complex as it requires careful planning to group wagons efficiently, as this is essential to form trains headed to the same or nearby destinations, making the operation less time-consuming and labor-intensive, reducing shunting operations.

- Intermodal Wagon Load (IWL): this method uses wagons that carry goods in standardized containers, allowing an easy switch between different modes of transportation, like ships, trains, and trucks. Usually, intermodal wagon loads are used for large volumes of goods that can fill entire containers, such that these containers can protect the goods and facilitate easy handling and transfer. These are typically loaded and unloaded at dedicated facilities designed for intermodal transfers and can be directly loaded onto or off the railcars without the need for breaking down or reconfiguring the train compositions extensively, meaning less shunting of individual wagons within a specific yard, [34]. Analyzing the freight train composition in terms of wagons, IWL on average presents a smaller number of wagons with usually longer rolling stocks compared to the SWL.

When integrating maintenance considerations into the problem of inbound train logistics, efficient grouping of wagons becomes more critical no matter the method of transporting goods. If a train has several wagons that require removal, whether it be for a change in the demand between inbound and outbound trains or maintenance, strategically grouping these wagons can significantly simplify the process while minimizing the operations to be performed. This reduces the shunting operations to be performed, and the time required for operation, reducing delays and cancellations. Looking specifically in the context of the freight train, this approach of clustering applies intuitively and beneficially in the context of SWL, where multiple sorting, grouping, and removing needs to be performed. Nonetheless, also IWL, and other rail freight transportation methods, can benefit from this approach, especially when they involve mixed cargo or wagons with varying maintenance needs.

2.4.2 Classification Problem in Shunting Yards

[24] states that we can classify the types of classification of a yard into three categories:

- Single-stage classification;
- Multistage classification with mixing tracks;
- Multistage classification with car ordering.

In a single-stage classification yard, railcars are moved just once, going directly from the receiving tracks to the classification tracks. There, they are immediately assembled into the outbound trains they are designated for. In contrast, multistage classification yards with mixing tracks don't have enough formation tracks to immediately start forming all outbound trains. So, in these yards, cars that cannot be immediately added to an outbound train are temporarily stored on mixing tracks. These cars must be moved back to the receiving tracks at least once more before they can be added to their outbound train. Lastly, in multistage classification yards with car ordering, cars are promptly used for their intended outbound trains, but they need to be arranged in a specific sequence, [24]. The TUSP problem, specifically the *parking* step as for [54], is strongly correlated with the classification problem in the shunting yard, which is key for the optimization of these operations. Imagining the structure as presented in Figure 2.1, the arrangement of the wagon parked on each track affects how long it will take for the composition of an arbitrary outbound train. A good portion of the research related to shunting yard operations is focused on departure prediction and formalized under the classification problem [24]. This problem is formulated usually into finding, for hump and gravity yards, the humping sequence of the arrival of the trains, together with their respective assignment on the classification tracks [14]. For this problem, the proposed solutions usually found in the literature consist of the minimization of departing trains' delay as the main objective of the function. [52] proposes a model to minimize the lateness of all outbound trains using arrival, departure, and processing times, highlighting that the congestion level of the shunting yard impacts directly the delayed departure. They suggest that the trains should be handled as soon as possible to prevent delay propagation on the operations. Minimization of weighted departures of the train is presented in [48, 49], by optimizing the humping sequences.

2.4.3 Integrated Modeling of Yards and Network

To finalize the modeling of the shunting yard, it is important to cover how the shunting yard and the network impact each other. The *yard and the network-integrated* modeling is on how these two elements impact each other. As the yard can generate delayed trains, impacting the network and creating disruptions in the schedule, likewise, the railway network can lead to delays in train arrivals at the yard, complicating the scheduling and organization of shunting operations. Therefore, while a properly functioning network would ensure the on-time arrival to the yards, a flawless functioning yard should ensure on-time departure to the network. This connection can ensure that the whole system would perform as expected.

When modeling shunting yards, [24] states that regardless of the specific problem being addressed, the basic input and output elements tend to remain consistent. Commonly, the input for these problems includes:

- The estimated arrival time of incoming trains (ETA).
- The expected departure time of outgoing trains (ETD).
- The order and makeup of railcars in both incoming and outgoing trains.
- Details about the yard's layout and the rolling stock, such as the number of tracks, their lengths, and the types and sizes of railcars.
- Expected time for each shunting operation.

As for the typical outputs provided by the solution, these usually encompass:

- A detailed schedule of various operations, including when to leave the receiving track, when to move to the classification track, and the timing for pullback operations.

- A timetable for locomotive usage.
- The planned departure times for all trains leaving the yard.

Nonetheless, the complexity of the railway operations, as well as uncertainties and variations leads to deviation in train runs [88, 94]. The study of this integration evolved into a new research topic when the focus shifted from the railway network importance to the shunting yard optimization, considering it not as a separate entity in a network but as an object that constantly interacts with it. Therefore, how the trains are dispatched into and received from the network, as well as the delay given and obtained by it, became of utter importance. The studies around this topic are branching towards two directions: the impact of the network on yards through a parametric approach with assumption on the yard modeling, and the connection of one or multiple separate yards and network models. [27] tried to analyze the network impact on yards through arrival variability, with [26] showing that an increasing train arrival time variability impacts yard performance by increasing the proportion of wagons missing their planned connection, increasing overall dwell times and the variability in volume/length of departing trains, therefore requiring more shunting operations. [27] explains that the influence of volume variation, train arrival variability, arriving block volume variability, and departing train distribution affects yard performance. The paper suggests also that, based on the different types of yards, different metrics should be analyzed.

The most relevant example of modeling interactions between yards is presented as a conceptual dispatching decision support system (DSS) in the OPTIYARD European Shift2Rail project [57], presenting a novel approach that combines an optimization model with a simulation environment. The core component of the DSS is the optimization algorithm that evaluates the yard operations several hours ahead through a simplified yard model. Based on the yard condition, the optimization algorithm makes decisions to generate a yard operating plan, which is later communicated to the yard simulator representing the real yard. The latter module simulates the operating plan, and if any perturbation happens during the execution, the optimization model is run again to provide a backup plan. If the plan is executed successfully, then another module micro-simulates the surrounding network to detect any perturbation that might impact the yard and vice versa.

2.5 Research Challenges

Despite the considerable advances in shunting yard operations and rolling stock management, the literature reveals distinct areas where current methodologies and practices remain unexplored. We therefore identify several critical gaps within the shunting operation domain:

- Maintenance operations on rolling stocks are usually solved as a separate, distinct problem from the RSP and TUSP, with rare cases in which the integration is presented, as underscored by [64]. As highlighted in this literature review though, these operations usually constrain the shunting movements inside the shunting yard and therefore can impact both costs and departure delays. The separate resolutions of the two problems might lead the whole system state to the unavailability of the fleet requirements and, therefore, to train cancellations, which cannot be foreseen by the classical, non-integrated models.

- Maintenance operations are usually scheduled in advance in each of the presented models, omitting if this happens due to mileage, time or condition-based constraints. While this modeling choice is completely legitimate, as explained in 2.3.1, when proposing an integrated mileage a similar framework has to be integrated as well inside the models.

- The existing literature primarily deals with disruption management in terms of delays, lacking a detailed focus on fleet management in terms of fleet status, which for this thesis is the integration of the maintenance using either a mileage or condition-based approach. This omission extends to the integrated impact of maintenance on the wagon fleet, as this might generate unavailability of specific wagons for a specific time, or increase the number of operations to be performed to fulfill a specific service. For instance, [42] and [66] do not consider the maintenance-related status of the fleet, such as feasibility due to maintenance constraints or parameters indicating the health of the fleet. Additionally, while [43] recognizes the need to consider the number of wagons in shunting operations, their approach overlooks the shunting time for adjacent wagons and does not account for the actual status of the inventory before proceeding with shunting, coupling, and decoupling activities.

- Most studies focus on maximizing the feasibility of timetables without specifically addressing the impact of shunting operations in yards. While these papers consider moving blocks of wagons, they do not explore strategies to optimize the movement of individual wagons or maximize the efficiency of moving wagon blocks, which could significantly reduce operational delays. As for [18], moving 1 or multiple adjacent wagons will require the same time. This becomes crucial when looking at the classification in the shunting yard for the Parking step performed in the TUSP, as with one shunt we could compose a higher portion of the train with fewer operations. Neglecting the adjacency of wagons in the shunting yards when it comes to retrieval might lead to unrealistic optimal solutions, as maybe wagons that were close to each other might have reduced the shunting time to fulfill an outbound demand.

- In the context of freight train operations, none of the papers mentioned above address the RSP in the freight train context, which requires different modeling approaches, such as those for SWL and IWL. This aspect is crucial and warrants further discussion, as outlined in Section 2.4.1.

These gaps highlight the need for a more comprehensive approach that incorporates both fleet status and maintenance considerations into the shunting operation environment, particularly within the freight train context. This approach should not only focus on the strategic level, involving fleet requirements and long-term costs but also integrate maintenance as an essential component of fleet management and shunting policies. Moreover, as for the state of the art, within this integrated vision, no condition-based maintenance model has been developed to see its impact on shunting operations. Another problem arises when not considering an integrated model that combines the RSP, the TUSP, and the maintenance within a strategic planning point of view: while tactical planning involves more analysis of delays and cancellations within the considered horizon, looking at the strategic point of view involves more choices that are by definition long-term, such as investment in the rolling stock fleet, workshop capacity, maintenance costs, and revenue, which are overlooked by the current literature.

We identify four primary challenges that our methodology aims to address in this thesis, with innovative approaches that integrate maintenance within the combined RSP and TUSP in the freight train context under SWL constraints for the strategic and tactical planning level and could easily be applied under IWL constraints as well. The following research challenges (RCs) have been summarized and addressed in this doctoral thesis:

- RC1: Lack of maintenance integration in the shunting operation from a system perspective.

- RC2: Lack of approaches to tackle the RSP, TUSP and mileage-based maintenance within an integrated problem from a strategic and tactical planning point.

- RC3: Lack of an hybrid model to assess the impact of the maintenance consideration modelling shunting operation for strategic assessment in freight rail.
- RC4: Lack of complete understanding of the impact of unplanned maintenance/disruption for condition-based maintenance, and how these affect the normal shunting operation.

Chapter 3

Research Objectives

There is no hate, only joy
For you are beloved by the goddess.

Loveless Act II

This chapter presents the formulated research objectives based on the order of research challenges identified in Section 2.5.

3.1 RO1: Analysis and Assessment of Maintenance Integration in Shunting Yard Operations

Current literature predominantly treats maintenance scheduling as a distinct entity, separate from the critical operational processes within shunting yards, assuming that a maintenance plan is already in place. Most of the research focuses on solving this maintenance scheduling separately, overlooking the influence that planned maintenance might have on shunting yard operations. As stated in Section 2.3, as mileage-based or condition-based maintenance are the preferred approaches for maintenance scheduling, this needs to be formalized and integrated into the model.

In this dissertation, we will integrate both types of maintenance. Specifically, in both the models presented in Sections 4.4 and 5.3, we propose formulations for the RSP model with TUSP considerations in which we include the mileage-based maintenance following the guideline from

[59]. In their work, they define mileage-based maintenance as a threshold-specific mileage at which a wagon has to be removed from the system for a specific amount of time and sent to the workshop. For the condition-based maintenance, we integrate it in the model presented in Section 5.3 through the use of a data-driven model to predict the condition of the rolling stock, integrated within the improved MILP model.

The reason for the necessity of this integration, in this dissertation, is tightly connected to the definition of shunting operation. In the context of the SWL, the number of operations required on a train changes significantly based on the position of the wagons in both the shunting yard and the inbound train. In the worst-case scenario, wagons in the inbound trains are positioned in an alternating sequence where one requires maintenance and the adjacent one does not. If then every wagon in the shunting yard that we choose to place in the outbound train is positioned in such a way that each wagon requires one shunting, the number of shunting operations to be performed is N, where N is the length of the train in terms of wagons. This is because $N/2$ operations are required from both the inbound train removal and outbound train composition respectively. When integrating maintenance, information regarding the mileage in the model, or some metric of similarity between wagons, can promote the creation of clusters of the wagons within the train. These can be exploited to reduce the amount of shunting operations to be performed for maintenance operations. When clustering wagons to enhance the performance of the shunting operations by solving the RSP problem, it is important to analyze the fleet status under different lenses. Integrating mileage information in the model in a strategic vision provides the ability to control fleet status in terms of mileage, which is necessary to ensure that everything is running smoothly and we can meet the maintenance thresholds. A non-controlled fleet in the long term might present skewed distributions of mileage towards the maximum mileage allowed, leading to unbalanced trains that might end up not benefiting from being able to create clusters when it comes to maintenance. This is highly dependent on the supply given by the shunting yard, where ideally the wagons parked there should have an aligned mileage with the trains going around in the system. This can improve the exchangeability of the wagons in case of disruption. The exchangeability of wagons for our problem is defined as the capability of the shunting yard to replace a wagon with another wagon of similar characteristics in terms of mileage.

3.2 RO2: Identification of the Most Relevant KPIs Required for Strategic and Tactical Planning

KPI	Planning Phase	KPI Class	Type of railway	Reference	Maintenance
Waiting time for passenger	O	Demand	P	[42]	No
Number of travelling passenger	T	Demand	P	[87], [92]	Yes
Num. Train	T/S	Demand	P	[84]	Yes
Planned daily mileage	S	Fleet	P	[59]	No
Rolling stock trajectories	T	Fleet	P	[9], [50]	No
Num. of allocated Rolling stock	T	Fleet	P	[87]	Yes
Degradation Level	T	Fleet	P	[45]	No
Fleet Size	T/S	Fleet	P	[9],[16],[87], [36], [93]	No
Number of blocks	O	Yard	P	[71]	No
Capacity rolling stock parked	O/T/S	Yard	P	[9], [41]	No
Maintenace operations	T	Yard	P/F	[51],[36],[64]	Yes
Shunting ops.	T	Yard	P/F	[51], [93], [64]	Yes
Shunting time	T	Yard	P/F	[51]	No
Delay	T	Yard	P/F	[50],[51]	No
Detention Time	T	Yard	F	[55]	Yes
Maintenance Time	T	Yard	P	[64]	No
Mileage distribution of the fleet	T	Yard	P	[93]	No
Operational costs	T/S	Yard	P	[61], [93]	No
Cancellations	T/S	Yard	P	[61],[50]	No

Table 3.1: KPI analysed in the literature divided by planning phase (O = Operational, T = Tactical, S = Strategical), KPI Class and type of railway (P = Passenger, F = Freight) in which the KPI has be applied.

To analyze the KPIs that will be needed to assess the goodness of the presented methodologies for strategical and tactical planning, we present in Table 3.1 an overview of the literature through the lens of the KPIs chosen for their results. Two major elements defined how we decided to approach this problem: one is the lack of purely strategic studies within this topic, leading to an absence of KPIs to analyze for this problem and the lack of maintenance integration within the shunting operation processes. For this, through Table 3.1, we identified KPIs that might be of relevance to this problem. Performance indicators that are important for our analysis in assessing the maintenance impact on shunting operation on the strategical planning are:

- Number of maintenance operations: this is a metric that in strategic planning plays an important role, given that maintenance is a source of costs. Integrating mileage-based maintenance, even at the tactical level, becomes crucial as the number of maintenance operations to be performed is strictly related to the potential delays and cancellations that might come from shunting out one wagon for maintenance reasons. This problem is connected to the definition coming from [17], in which a shunting operation can be performed on either one or a block of wagon. When it comes to shunting operations to be performed for maintenance reasons though, it becomes nearly impossible to create a 1-to-1 connection between a specific shunting operation with a maintenance operation, as with one block we can remove one or multiple wagons, some of which might require maintenance.

- Number of shunting operations performed: this metric helps us understand how the model(s) are performing from the clustering point of view. Within the same instance, the higher the number of shunting operations, the more exploitative could be the model of the system, and the better the performance of the system might be. Moreover, this is also a metric that when applied on the single wagons can provide us with a measure of how much the model is under/over using the fleet.

- Delays, cancellation: this metric helps us understand the feasibility of the model. From a strategic point of view, we are not so interested in reaching perfect feasibility as much as for the tactical level, given that we can account for rescheduling by penalizing the model for this. Nonetheless, having low cancellation rates overall, as well as low delays, is important to ensure the quality of the service.

- Fleet Size: this is both a strategic and tactical KPI, depending on the point of view. On the tactical side, the absence of an available fleet might lead to cancellations/delays for retrieval, while the surplus might not be relevant information. When this surplus is then propagated through time, therefore on the strategical side, this becomes of uttermost importance, as is one of the metrics that tell us how good are we at optimizing fleet usage, and therefore reducing the costs for investments in rolling stock in the long term. Furthermore, the maintenance impacts directly on the fleet size required for a specific instance, as it creates unavailability for a specific amount of time.

- Mileage distribution of the fleet: this KPI is more important from the strategic point of view when we can start seeing the rotation of the fleet and different maintenance performed on wagons, rather than on a tactical level where our planning horizon is smaller. This is connected to the concept of exchangeability of the fleet, as stated in the previous RC, where ideally we would like the mileage of the parked wagons and the ones going around to be as aligned as possible.

Regarding the tactical perspective, the KPIs for assessing the integration of maintenance into freight train operations are as follows:

- Operating Cost: it is defined as the total cost associated with the operation of freight train services, including maintenance and shunting. It is a metric for evaluating the efficiency of operational strategies. Lower operating costs can indicate effective maintenance scheduling, as well as shunting operations performed.

- Overall Shunts: this refers to the total number of shunting operations performed within the instance to fulfill the demand of an outbound train considering maintenance. This KPI provides insights into the operational workload of shunting yards and is directly connected to delays and resource efficiency. Efficiently managing the number of shunts can lead to reduced operational times and lower costs.

- Total Wagons Moved: this metric displays the efficiency in allocating the rolling stock for an outbound service. This metric allows us to understand shunting efficiency, as more wagons moved with less shunt means more clustering, thereby improving the overall operations performed.

- Actual Departure: this metric defines the actual train departures, reflecting punctuality and efficiency. Actual departure times are crucial for maintaining service reliability and customer satisfaction. Delays can be costly and disrupt downstream operations, and higher delays can cause cancellations. In tactical planning, having a better performance in the average actual departures signifies better efficiency in performing operations, reducing cancellation rates.

- Risk Term: for this study, this metric measures the risk associated with operational decisions of keeping or removing a wagon for maintenance. The risk term helps in assessing the model performance in terms of potential cost loss.

3.3 RO3: Development of a Hybrid Modeling Approach for Long-Term Assessment through MILP and Simulation

We decided to address RC3 by developing a hybrid modeling framework that combines a Mixed Integer Linear Programming (MILP) model with a simulation environment, as for [57, 86], to evaluate the long-term efficacy and robustness analysis of shunting yard operations, particularly under maintenance constraints.

The proposed MILP models a similar problem to the RSP, in which we need to decide both which wagon to remove from an inbound train and which wagon to add to form the final composition for the next service. In the first part, defined as *shunt-out*, the wagon is chosen based on mileage-based maintenance considerations, clustering possibility, and delays consideration to reduce the number of shunting operations to perform. In the second part, defined as *shunt-in*, wagons in the shunting yard that are more suitable are selected based on the concept of *shunting policies*. These policies are defined as the criteria for assigning to a wagon its time of service, looking at attributes of the wagon of the wagon seen as the mileage covered from different perspectives, and are explained in detail in Section 4.5.

Given the computational complexity brought by the inclusion of a strategical approach to this model, namely considering multiple instances of multiple trains within a larger timespan at the same time, the chosen solution is to break down each incoming train service and solve it singularly (if needed) while having the whole system managed by a higher-level interface to allow a continuous simulation. What it means in practice is that the proposed simulation environment provides information regarding the management of rolling stock status, different delays and cancellations, loading and unloading, and maintenance operations to be performed. It allows us to simulate the TUSP by having the structure of the shunting yard modeled within it. This comes with the strong assumption that we do not consider the potential conflicts coming from the routing of wagons as in the TUSP problem, but just the retrievability of the wagon. We opted for this assumption as previously made in [42, 43]. In their work, they assume that the routing is performed in the following step, modeling only the position of the wagons in the shunting yard and the estimated time for moving rolling stock. In our model, the shunting time for the retrieval from the shunting yard is treated as a pseudo-random variable, while the removal from the inbound train is dependent on the position of each wagon that needs shunting. This is further explained in Section 4.3.2. This assumption holds significant importance for tactical and operational planning; however, its impact is less critical when viewed from a strategic perspective, allowing for a degree of flexibility in its application at that level. The choice of the simulation environment is important as a modeling choice for the MILP problem. A major limitation is that the solution space of the MILP is limited on the single train and shunting yard status. This effect is mitigated by the *shunting policies* embedded in the model, which has a longer-term vision aiming to maximize the cluster that can be created in the long term. This allows the MILP to effectively use the simulation environment to propagate its effect through time and within the system, resulting in the capability of performing long-term analysis. For the classical RSP, this can be disregarded, as usually, the aim is to minimize delays and cancellations, assigning wagons to the outbound train which minimizes the operational time. When incorporating maintenance, and looking long-term in minimizing these operations, then the choice of which wagon to place in and in which position becomes crucial in the light of the scenario as explained in Section 2.5, RC1. The policies aim to improve the likelihood of creating clusters within the inbound train for maintenance reasons and improve the controllability of the system in terms of maintenance.

3.4 RO4: Data-Driven Modeling for Condition-Based Maintenance and Unplanned Disruption, and its Integration in MILP Modeling through a Risk-Management Approach

As stated in Section 2.3, mileage-based or condition-based maintenance are the preferred approaches for maintenance scheduling. While the former has been addressed in RC3, integrating the latter requires a more data-driven modeling approach, as usually these data are computed through sensors on the rolling stocks. In this thesis, this problem is solved by developing a Machine Learning (ML) model, which is then plugged into an extended version of the MILP model compared to the one presented in RC3 through a risk-assessment approach. We developed a Binary Classification ML model whose aim is to predict the likelihood of rolling stock breakdowns based on real data to create a condition-based maintenance model to use in our MILP model. This model identifies key features that can generate disruption, and therefore unplanned maintenance, aiding in preemptive decision-making. To improve the MILP presented in Section 4.2 and adapt it to a more tactical vision, we developed a new version of the MILP, presented in Section 5.3 that addresses the shortcomings of the previous model, which includes a better representation of the retrieval time for each wagon from the shunting yard. As the ML model is our source of information regarding disruption, acting like a ”sensor” for our rolling stock, we want to be able to consider its capability of prediction within our model. This is because the failure of the ML model can mean the failure of the wagon, which can lead to a disruptive event. In this work, we propose a *mild* integration of the ML input through a risk management framework. The proposed solution risk assesses the ML model against its failure within our MILP model, using the classification model's metrics as probabilities of failure/success to compute the different risks associated with the decision of following the ML direction or not. The idea is to use the ML outputs not as direct inputs but as advisory tools, assisting in strategic decisions about maintenance and shunting operations. The model evaluates the risks associated with various operational scenarios, balancing the impact of potential disruptions against the costs and benefits of different maintenance strategies, read as deciding whether to remove one particular wagon or not for preemptive maintenance. This approach is particularly effective in mitigating the effects of unpredictable and severe disruptions, such as rolling stock failures, which can have widespread consequences on network efficiency. Moreover, in a context where preemptive maintenance can be time-consuming and resource-draining, having this approach will help mitigate the risk to an acceptable level.

A visual framework of the proposed RCs and ROs and how these are going to be integrated is provided in Figure 3.1.

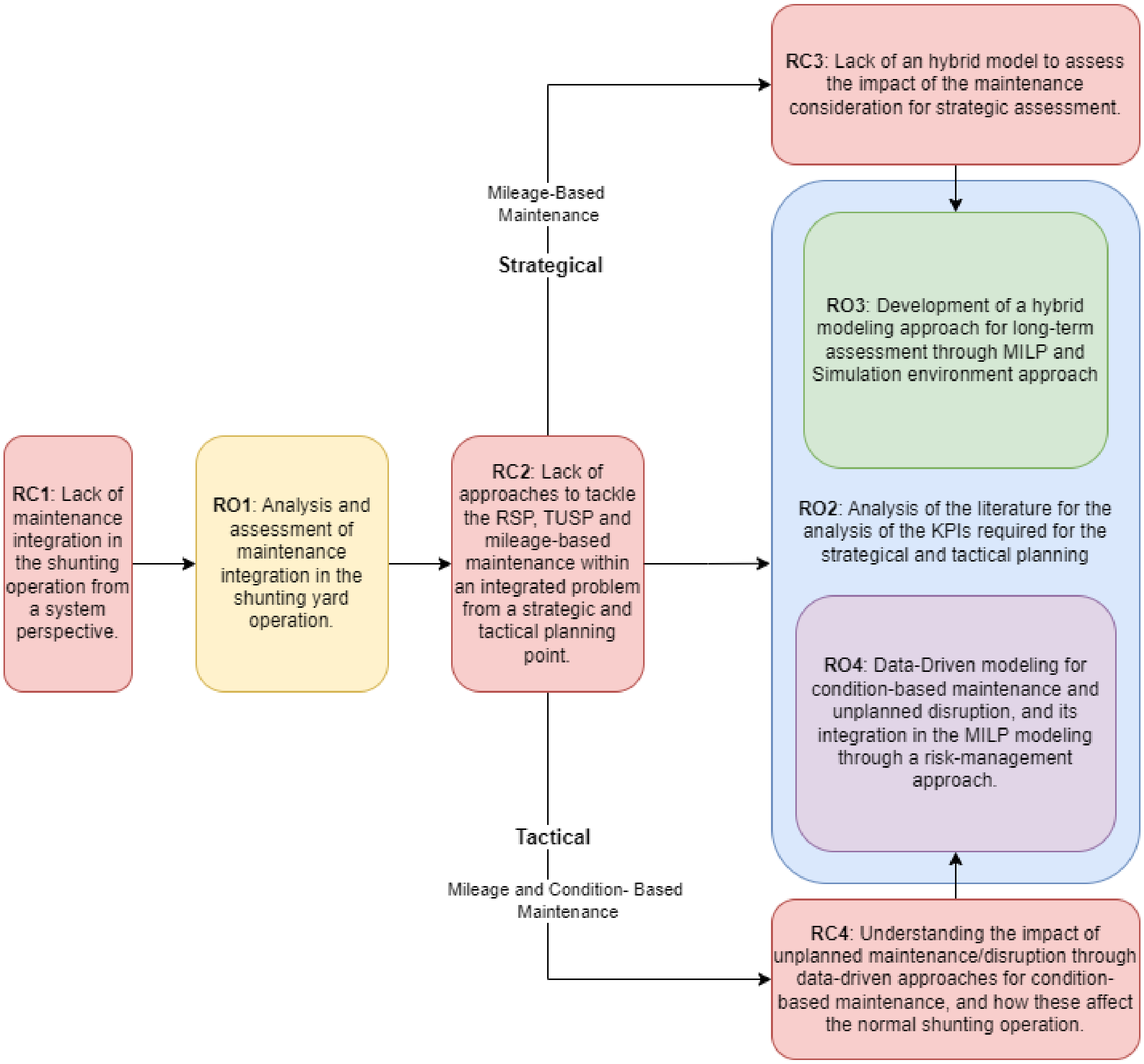


Figure 3.1: Visual framework of the proposed RCs and ROs

Chapter 4

Simulation and MILP Modeling for the Shunt-In/Shunt-Out Problem (SISO)

My friend, do you fly away now?
To a world that abhors you and I?

Loveless Act III

4.1 Introduction

This chapter covers RO3, Section 3.3, "Development of a Hybrid Modeling Approach for Long-Term Assessment through MILP and Simulation", and has been published in [11, 13]. We formalize the Shunt-In Shunt-Out (SISO) problem in Section 4.2, within the context of railway shunting operations, focusing on the impact of maintenance operations. This problem is a formulation of the RSP with TUSP considerations and maintenance constraints. We aim to optimize shunting operations with rolling stock clustering consideration and evaluate the impact of maintenance through a mileage-maintenance threshold approach. We say TUSP consideration as we consider the organization and order of the rolling stocks within the inbound train for shunting consideration, but we omit the computation of routing or conflict. The chapter then focuses on the formalization of the SISO problem, in which we present its mathematical modeling through Mixed Integer Linear Programming (MILP) as in Section 4.4 and Appendix A, and the various strategies and policies for effective wagon management, in Section 4.5. We then present a detailed simulation framework in Section 4.6. This tool is necessary to integrate the mathematical model within a strategic vision, providing a comprehensive approach to understanding and solving the SISO problem.

4.2 The Shunt-In Shunt-Out Problem (SISO)

We want to analyze the impact of mileage-based maintenance on the shunting operations and strategic KPIs through the use not of the classical RSP and TUSP problem, but by creating an integrated problem that is suitable for this kind of assessment. We formalize this new problem as the *Shunt-In Shunt-Out problem* (SISO). We define them as two coupled problems, as they deal with the aspects of removing wagons and recomposing the train for the outbound service. Specifically, we define a model for selecting which wagons have to be removed from the inbound train due to leasing contract, timetable constraints, or for reducing operational time *(Shunt-Out, SO)*, and criteria for replacing the latter with shunting yard's wagons to make up the outbound train *(Shunt-In, SI)*, taking into account parameters such as the time to shunt, the shunting yard's supplies availability and so on.

To do this, we propose an integrated MILP model resolving both the Shunt-In and Shunt-Out problems. The mathematical model considers rolling stock maintenance as mileage-based maintenance and timetable constraints, as well as a multi-objective function that aims to minimize the number of shunts performed. The multi-objective function further considers weighted delay terms of outbound trains, the SI policy applied, and weighted terms associated with shunting binary variables used to avoid a quick shortage of wagons inside the shunting yard. The SO problem can be influenced by various constraints such as maintenance constraints, operational costs, seasonal wagon demand, etc. The cost of the make-ready stage of the shunting locomotive is the biggest part of the SO operation's cost, so creating clusters of shunts by triggering *optional* shunts that aren't caused by maintenance or demand constraints can bring cost benefits. Moreover, as a constraint for this model, the outbound train's composition must be fulfilled and wagons must not be moved for maintenance unless their mileage is within the maintenance threshold range. The SI problem is a complementary problem to the SO, where the goal is to minimize time and economic costs by replacing each shunted-out wagon with a suitable wagon from the shunting yard. A *suitable* wagon is defined as a wagon that has enough mileage for the next trip, filling the correct type requested. The basic problem is the replacement of a shunted wagon with one that takes less time to shunt-in the shunting yard. However, considering just one parameter for a strategic vision might be short-sighted, as we are not considering many parameters unique to that specific wagon. As an example, the shunted-in wagons might already be close to the mileage limit, which can potentially trigger an SO maintenance request when the train returns from its service, and thereby require additional shunting. A multi-component objective function that focuses on economic costs can avoid this additional cost. For the SI problem, different policies are described, each of which is characterized by particular wagon assignment criteria, mileage-based. Due to their different assignment criteria, each SI policy has pros and cons, therefore, should be used considering the target that the user wants to achieve.

4.3 Assumptions

For this specific model for the SISO problem, we apply the following assumptions:

- We consider two types of wagons, $SIMPLE$ (Fig. 4.1a) and $DOUBLE$ (Fig. 4.1b) to simplify the initial problem. These two classes of wagons are defined by the number of containers they can carry.

- We define two types of SO operations: the *mandatory* and the *optional* ones. The *mandatory* operations are performed on rolling stocks that need to be removed due to maintenance rules or demand-matching constraints, as further explained in Section 4.3.1. The *optional* operations are performed on rolling stocks between successive mandatory shunts to create bigger clusters and reduce shunting costs and time.

- A *cluster* of shunts (namely, two or more adjacent wagons requiring SO operations) is associated with a single economic cost and a single temporal cost.

- The wagon matching between inbound and outbound wagons is not positional, as explained in Section 4.3.1.

- The operational time to SO is fixed and defined by the shunt time for a wagon/cluster. The time required to SI, given the shunting yard position and therefore the expected time for routing, changes from wagon to wagon, as further explained in Section 4.3.2.

- The optional shunt can be performed only when the wagon's *virtualmileage* ranges between the minimum and maximum mileage or exceeds the maximum mileage defined by the corresponding leasing contract. The virtual mileage is defined as the kilometers covered by the wagon i-th (m_i) once it has performed the outbound train's next trip (r_T), and described in Eq. 4.1:

$$v_{ms_i} = m_i + r_T \tag{4.1}$$

- The *maintenance* shunt can be performed only when the virtual mileage exceeds the maximum mileage defined for that wagon. Further explanations are provided in the definition of the maintenance constraint in Section 4.3.1.

- If the operational shunting time exceeds the planned departure time of the outbound train, a penalty due to the lowering of the service level is considered. This penalty is given by the departure delay function in the objective function of the MILP.

(a) Example of a 40' freight wagon.

(b) Example of a T3000e wagon.

Figure 4.1: Example of a SIMPLE wagon (a) and a DOUBLE wagon (b).

4.3.1 Mandatory Shunts

As for [38], we have identified two scenarios for shunting: demand matching and mileage-based maintenance constraints. These are shunts that are performed on rolling stocks that need to be replaced for one of the two above-mentioned reasons. For this work, we define them as *Mandatory shunts* and they can be seen as the type of shunts that are inevitable. The demand-matching shunts are triggered by wagons that are necessary to remove due to a change between the inbound and outbound train composition. To achieve this, we determine the difference in wagon types (num_k) between the inbound and outbound trains without considering the positional constraints of rolling stocks. Then, we select wagons to SO from the inbound train based on the number required of each type and the maintenance threshold (Eq. 4.2). For the second type of mandatory shunts, due to maintenance, the state of each wagon $i \in \mathcal{T}$ can be described by its current mileage, m_i, and the maximum mileage, m_{max_i}, before it requires maintenance, which is either determined by its leasing contract or based on practitioner policy. Formally, we define the *maintenance constraint*, which can be expressed formally as Eq. 4.2 using the virtual rate definition from Eq. 4.1:

$$v_{ms_i} \leq m_{max_i}, \tag{4.2}$$

This equation expresses that if the current mileage exceeds the maximum mileage with the next trip, the wagon must be removed from the inbound train and sent for maintenance. If a wagon selected for demand matching has also surpassed its m_{max_i}, it is immediately sent for maintenance. Otherwise, it is parked inside the shunting yard. This process results in an intermediate list of wagons with the new train composition.

4.3.2 Computing Time to Shunt-In

Routing time for the wagon to SI from the shunting yard to the inbound train depends on several factors, such as the shunting yard layout, the number of rail tracks, and is usually the solution of the individual TUSP problem, given the input of the RSP. For this study, we opted not to determine this time explicitly. Given that our study focuses on estimating the impact of maintenance considerations and how neglecting it can lead to the underestimation of strategic decisions on assets and parameters such as fleet requirement, we have decided not to address the optimization problem associated with daily, operational time scales in great detail, following the assumption for routing of each wagon of the shunting movement and the optimization of it as presented in [41].

Instead, we rely on assumptions based on information provided by practitioners, which are the following:

- Based on the wagon selected to be shunted in, we compute the necessary time t_{sin} sampling a normal distribution describing the time that should be computed through the TUSP problem.
- The time t_{sout} required for each SO maneuver is considered fixed and dependent on the number of wagons/clusters to be removed from the inbound train.

As mentioned above, the t_{sin} is computed through distribution sampling as the time required to retrieve a wagon from the shunting yard depends on multiple exogenous and endogenous variables, and is usually the output value of the TUSP problem and the Classification Problem, both NP-hard problems. Additionally, the shunting yard configuration greatly influences the actual time required for shunting, t_{sin}. To avoid overly specific results, we opted for a more general approach, assuming no optimization of shunting movements and no conflict, representing the shunting yard as a vector with dimension $\mathcal{S}$. We calculate a normal distribution, with mean μ and standard deviation σ, denoted as $\mathcal{N}(\mu, \sigma^2)$, referred to as Δ. The first distribution, Δ, represents the average time to move one wagon from a random track inside the shunting yard to the outbound train. Once the policy chooses a wagon w, we sample its associated time $t_{w_{\text{out}}}$ from Δ and use it to generate a pseudo-random variable, with mean $t_{w_{\text{out}}}$ and a fixed standard deviation ζ^2. This standard deviation represents the deviation in time to move one wagon from a random position in the track up to the end of it. Therefore, this second distribution, $\mathcal{N}(t_{w_{\text{out}}}, \zeta^2)$, denoted as $\Delta_{|SH_t|}$ can be seen as the average time needed to move a wagon from a random position in the shunting yard to the outbound train, therefore as the distribution of t_{sin}.

4.4 Mathematical Model

Section 4.4 presents the nomenclature for the SISO problem formulation; Section 4.4 explains in-depth a base version of the MILP model; Section 4.5 presents several SI policies translatable as different versions of the MILP objective function. The following model is the final iteration of the model used in the simulation environment, developed through these years and presented in [13].

Nomenclature

Sets	
Name	**Description**
$\mathcal{T}, i \in \mathcal{T}$	Set of inbound train's wagons.
$\mathcal{S}, j \in \mathcal{S}$	Set of shunting yard's wagons.
$\mathcal{K}, k \in \mathcal{K}$	Set of wagon types.
$\mathcal{TR}, T \in \mathcal{TR}$	Set of train destinations.
Parameters	
Name	**Description**
a_T	Integer value expressing the inbound train's arrival time, $T \in \mathcal{TR}$.
d_T	Integer value expressing the outbound train's planned departure time, $T \in \mathcal{TR}$.
dd_T	Integer value expressing the outbound train latest time before cancellation, $T \in \mathcal{TR}$.
ts	Integer value expressing the time required by a shunting locomotive to perform a single SO operation.
r_T	Integer value expressing the kilometers the train T will perform during the next trip, $T \in \mathcal{TR}$.
m_i	Integer value expressing the current mileage of the wagon, $i \in \mathcal{T}$
ms_j.	Integer value expressing the current mileage of the wagon, $j \in \mathcal{S}$.
m_{max_i}	Integer value expressing the maximum mileage before the maintenance of the wagon, $i \in \mathcal{T}$.

ms_{max_j}	Integer value expressing the maximum mileage before the maintenance of the wagon, $j \in \mathcal{S}$.
m_{min_i}	Integer value expressing the minimum mileage to shunt the wagon, $i \in \mathcal{T}$.
$type_{in_i}$	Integer value equal to 1 or 2 expressing the type of the wagon, $i \in \mathcal{T}$.
$type_{S_j}$	Integer value equal to 1 or 2 expressing the type of the wagon, $j \in \mathcal{S}$.
$code_{in_i}$	Integer value expressing the unique code associated with the wagon in the inbound train, $i \in \mathcal{T}$.
$code_{S_j}$	Integer value expressing the unique code associated with the wagon of the outbound train, $j \in \mathcal{S}$.
$type_k$	Integer value equal to 1 or 2 expressing the wagon type on the outbound train that must increase due to the demand, compared to the inbound train $\mathcal{T}$, $k \in \mathcal{K}$.
num_k	Integer value expressing the surplus of wagons of the type $type_k$ in the outbound train new composition, compared to the inbound train $\mathcal{T}$, $k \in \mathcal{K}$.
n_{ms_j}	Float value expressing the *virtual rate* (Eq. 4.1), $j \in \mathcal{S}$.
c_{u_i}	Float value expressing the shunting convenience cost used as a preemptive tool to avoid infeasibility of the shunting yard $\mathcal{S}$.
$c_{s_{i,j}}$	Float value expressing the temporal cost to replace the wagon i-th on the inbound train $\mathcal{T}$ with the wagon j-th inside the shunting yard $\mathcal{S}$, normalized through the Min-Max normalization.
M	Big-M coefficient.
β	Float value between 0 and 1 expressing the percentage of operational time left before the outbound train's deadline once all the SO operations are performed.
α	Float value equal to 1 - β.
Simulation Output	
Name	**Description**

$code_{\mathrm{out}_i}$	Integer value expressing the unique code associated with the wagon i-th on the outbound train.
Decision Variables	
Name	**Description**
ad_T	Integer value expressing the actual departure time of the outbound train T once all the shunting operations are performed, $i \in \mathcal{T}$.
y_i	Binary variable equals to 1 if on the wagon i-th on the inbound train $\mathcal{T}$ a maintenance or optional shunt is performed, $i \in \mathcal{T}$.
$x_{i,k}$	Binary variable equals to 1 if on the wagon i-th on the inbound train $\mathcal{T}$ a demand shunt is performed and it is replaced by a shunting yard's wagon of type k, $i \in \mathcal{T}, k \in \mathcal{K}$.
$z_{i,j}$	Binary variable equal to 1 if the wagon i-th on the inbound train $\mathcal{T}$ is replaced by the wagon j-th inside the shunting yard $\mathcal{S}$, $i \in \mathcal{T}, j \in \mathcal{S}$.
γ_i	Binary variable equals to 1 if the wagon i-th on the inbound train $\mathcal{T}$ is shunted out, regardless of the shunt type, $i \in \mathcal{T}$.
σ_1	Binary variable equals to 1 if $dd_T \geq ad_T > d_T$, and to 0 if $d_T \geq ad_T$.
σ_2	Binary variable equals to 1 if $ad_T > dd_T$, and to 0 if $ad_T \leq dd_T$.
σ_3	Real variable equals to $\frac{ad_T - d_T}{dd_T - d_T}$ if $\sigma_1 = 1$, and to 0 if $\sigma_1 = 0$.
$adj_{i,i+1}$	Binary variable equals to 1 if both the wagon i-th and its adjacent wagon $i+1$-th on the inbound train $\mathcal{T}$ are shunted out, $i \in \mathcal{T}$.

4.4.1 Shunt-In/Shunt-Out (SISO) Model

The following formulation is presented also without comments in Appendix A.
The objective function for the MILP model is presented in Eq. 4.3. For this problem, we propose a weighted multiobjective function.

$$\min \overbrace{\sum_{i \in \mathcal{T}} \gamma_i - \sum_{i=1}^{|\mathcal{T}|-1} adj_{i,i+1}}^{Eq.(4.4)} + \frac{|\mathcal{T}|}{2} \overbrace{(\sigma_2 + \sigma_3)}^{Eq.(4.5)} + \\ + \frac{|\mathcal{T}|}{4} \overbrace{(\sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{S}} (c_{s_{i,j}} \alpha + n_{ms_j} \beta) z_{i,j} + \sum_{i \in \mathcal{T}} y_i c_{u_i}}^{Eq.(4.6)} \tag{4.3}$$

Eq. 4.4 represents the actual number of shunts performed considering the assumption on clustering (Section 4.3).

$$\sum_{i \in \mathcal{T}} \gamma_i - \sum_{i=1}^{|\mathcal{T}|-1} adj_{i,i+1} \tag{4.4}$$

It considers the overall number of wagons to be removed for SO (γ_i), regardless of them being mandatory or optional shunts, minus the adjacencies selected ($adj_{i,i+1}$). This way, we count the number of clusters of shunts activated rather than single wagons shunted-out, taking into account the assumption that states that two or more adjacent wagons shunted out are considered as a single economic cost. Shunt-out operations are not multiplied by a weight, since we have set the other objective function terms in such a way that they are all comparable in number of wagons.

Eq. 4.5 expresses the penalty related to the possible departure delay $\mu(ad_T)$ of the outbound train.

$$\frac{|\mathcal{T}|}{2} \overbrace{(\sigma_2 + \sigma_3)}^{\mu(ad_T)} \tag{4.5}$$

Due to its non-linear behaviour, $\mu(ad_T)$ has been handled by introducing three different binary variables σ used for the constraints (4.7)-(4.16). If the actual departure time (ad_T) is smaller than its planned departure time (d_T), σ_2 and σ_3 will be both equal to zero and $\mu(ad_T)$ will be equal to zero as well. If ad_T is between d_T and the maximum time before cancellation (dd_T), the model imposes σ_2 to zero and σ_3 and $\mu(ad_T)$ to $\frac{ad_T - d_T}{dd_T - d_T}$, representing the normalized delay for the expected departure. $\mu(ad_T)$ is then multiplied by a weight proportional to the number of wagons on the inbound train. The weight can not be proportional to the number of wagons SO, as it would both produce a non-linear term and enter into contradiction with the clustering assumption.

Eq. 4.6 includes the decision criteria of the selected shunt-in policy and a term that considers the shunting yard availability of wagons.

$$\frac{|\mathcal{T}|}{4}\overbrace{(\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{S}}(c_{s_{i,j}}\alpha + n_{ms_j}\beta)z_{i,j}}^{\text{Add-on SI Policy}} + \sum_{i\in\mathcal{T}} y_i c_{u_i}) \tag{4.6}$$

According to practice priority, Eq. 4.6 is weighted with a halved value compared to the second term. This term takes into account two parts: the shunt-in policy applied and the shunting convenience costs c_{u_i}. The *Add-on SI Policy* is explained in detail in Section 4.5, as is connected with all the shunt-in policies. The second aspect is a preemptive tool to avoid too many optional shunts that could lead to an unfeasible state of the shunting yard capacity over the simulation iterations. The value c_{u_i} is defined as the shunting convenience, and it prevents the model from performing too many optional shunts to avoid the possibility of a shortage of rolling stock in the shunting yard. Therefore, if the current mileage of the wagon i-th in the outbound train is low and the number of wagons in the shunting yard is below a certain threshold, the cost c_{u_i} is computed so that will be high, and the model will opt not to shunt-out the wagon.

Eq. 4.7 - 4.16 represents the constraints related to the delay $\mu(ad_T)$ function. Based on ad_T, they define if the outbound train is on time, late, or if it will be canceled due to the exceeding of the planned deadline. This is done by using three temporal variables σ, as described in Section 4.4, to add a delay penalty in the objective function.

$$ad_T \leq d_T + \sigma_1 M + \sigma_2 M \tag{4.7}$$

$$dd_T + (1 - \sigma_1)M \geq ad_T \tag{4.8}$$

$$ad_T > d_T - (1 - \sigma_1)M \tag{4.9}$$

$$ad_T \leq dd_T + \sigma_2 M \tag{4.10}$$

$$ad_T > dd_T - (1 - \sigma_2)M \tag{4.11}$$

$$\sigma_1 + \sigma_2 \leq 1 \tag{4.12}$$

$$\sigma_3 \leq \sigma_1 M \tag{4.13}$$

$$\sigma_3 \geq \frac{a_T - d_T}{dd_T - d_T}\sigma_1 \tag{4.14}$$

$$\sigma_3 \leq \frac{ad_T - d_T}{dd_T - d_T} + (1 - \sigma_1)M \tag{4.15}$$

$$\sigma_3 \geq \frac{ad_T - d_T}{dd_T - d_T} - (1 - \sigma_1)M \tag{4.16}$$

Eq. 4.7 - 4.9 ensure that the conditions of σ_1 are satisfied: if both σ_1 and σ_2 are equal to 0 then the outbound train must be on time; if σ_1 is equal to 1 then ad_T must range between d_T (not included) and dd_T. Instead, Eq. 4.10 and 4.11 express σ_2's conditions, saying that if σ_2 is equal to 0 then ad_T has not already reached the outbound train's deadline, otherwise, the outbound train will be canceled. While constraint 4.12 links σ_1 to σ_2 by forcing them not to be simultaneously active, constraints 4.13 - 4.16 link σ_1 and σ_3 and represent σ_3's conditions. The latter states that if σ_1 is equal to 0 then σ_3 will be equal to 0 as well, while if σ_1 is equal to 1 then ad_T ranges between d_T and dd_T, and σ_3 will be equal to $\frac{ad_T - d_T}{dd_T - d_T}$.

Eq. 4.17 sets the ad_T equal to the inbound train arrival time (a_T) plus the time required to perform both the shunt-out and shunt-in operations.

$$a_T + \overbrace{(\sum_{i \in \mathcal{T}} \gamma_i - \sum_{i=1}^{|\mathcal{T}|-1} adj_{i,i+1})}^{SO\ Ops.} ts + \overbrace{\sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{S}} c_{s_{i,j}}}^{SI\ Ops.} = ad_T \tag{4.17}$$

Constraints 4.18 and 4.19 set α and β, the percentage of operational time used by the shunt-out (β) and the shunt-in (α) operations.

$$\frac{a_T \ + \ (\sum_{i\in\mathcal{T}} \gamma_i - \sum_{i=1}^{|\mathcal{T}|-1} adj_{i,i+1})ts}{dd_T} \ = \ \alpha \tag{4.18}$$

$$1 - \alpha = \ \beta \tag{4.19}$$

The operational time is defined in this case by the arrival time of the train in the shunting yard and the latest time before its cancellation. Both the values of α and β range between 0 and 1.

Equations 4.20 and 4.21 ensure that shunt-out operations are performed properly.

$$y_i \geq \frac{m_i + r_T}{m_{max_i}} - 1 - (\sum_{k\in\mathcal{K}} x_{i,k})M \qquad \forall i \in \mathcal{T} \tag{4.20}$$

$$y_i \leq (1 - \sum_{k\in\mathcal{K}} x_{i,k})\frac{m_i + r_T}{m_{min_i}} \qquad \forall i \in \mathcal{T} \tag{4.21}$$

The above constraints represent two conditions:

- If the next trip of the train (r_T) plus the actual mileage (m_i) exceeds the maximum mileage (m_{max_i}), then the wagon i-th must be shunted out, therefore we y_i is equal to 1.
- If a demand shunt is already activated for the wagon i-th, it is not possible to perform maintenance or an optional shunt.

Eq. 4.22 is an optional constraint that should be considered if we want to force the model to activate only mandatory shunts once m_{max_i} is exceeded.

$$y_i \leq \frac{m_i + r_T}{m_{max_i}} \qquad \forall i \in \mathcal{T} \tag{4.22}$$

Eq. 4.23 - 4.25 ensure that the new composition on the outbound train will be satisfied.

$$\sum_{i\in\mathcal{T}:type_{in_i}\neq type_k} x_{i,type_k} = num_k \tag{4.23}$$

$$\sum_{k\in\mathcal{K}} x_{i,k} = 0 \qquad \forall i \in \mathcal{T} : type_{in_i} = type_k \tag{4.24}$$

$$\sum_{i\in\mathcal{T}} x_{i,k} = 0 \qquad \forall k \in \mathcal{K} : k \neq type_k \tag{4.25}$$

The sum of the wagon type to shunt-in ($x_{i,k}$) with a k different from the $type_k$ (the wagon type that must increase on the outbound train) must be equal to the surplus of the wagon of $type_k$.

Moreover, the sum of $x_{i,k}$ of $type_k$ type must be equal to 0, because in this way the other part of the composition will not be altered.

$$|type_{out_i} - type_{in_i}| = p_i \qquad \forall i \in \mathcal{T} \quad (4.26)$$

Practice might require an orderly composition of the departure train, meaning that for each wagon position, a specific type is requested there. To express this Eq. 4.23 - 4.25 must be replaced by constraint 4.26, where $type_{out_i}$ represents the type of wagon i required on the outbound train and the binary variable p_i is equal to 1 when the inbound train wagon i is shunted out due to demand matching. However, constraint 4.26 works when there are exactly two wagon types. This is connected with the SIMPLE/DOUBLE assumption as for Section 4.3.
Eq. 4.27 - 4.30 models the shunt-in operations.

$$\sum_{j \in \mathcal{S}: type_{S_j} = type_{in_i}} z_{i,j} = y_i \qquad \forall i \in \mathcal{T} \quad (4.27)$$

$$\sum_{j \in \mathcal{S}: type_{S_j} \neq type_{in_i}} z_{i,j} = x_{i,type_k} \qquad \forall i \in \mathcal{T} \quad (4.28)$$

$$z_{i,j} \leq 2 - \frac{ms_j + r_T}{ms_{max_j}} \qquad \forall i \in \mathcal{T}, \forall j \in \mathcal{S} \quad (4.29)$$

$$\sum_{i \in \mathcal{T}} z_{i,j} \leq 1 \qquad \forall j \in \mathcal{S} \quad (4.30)$$

Constraints 4.27 and 4.28 force the model to activate the shunt-in ($z_{i,j}$) with the proper type from the suitable wagons. Eq. 4.27 forces the wagons shunted out with maintenance or optional shunt to be replaced by shunting yard's wagons of the same type. For Eq. 4.28, the ones shunted out with demand shunt will be replaced by wagons of the opposite type. This is because we consider only two types for this formulation, therefore if a demand shunt is requested, the opposite type is needed. Constraints 4.29 and 4.30 ensure that we cannot select wagons from the shunting yard unable to perform the next trip r_T, and that the same shunting yard wagon will not replace more than one inbound train wagon. These constraints define the feasible region where the SI policy applied will select shunting yard wagons.

The following Eq. 4.31 and 4.32 allow the model to comply with the clustering assumption.

$$\sum_{k \in \mathcal{K}} x_{i,k} + y_i = \gamma_i \qquad \forall i \in \mathcal{T} \quad (4.31)$$

$$2adj_{i,i+1} \leq \gamma_i + \gamma_{i+1} \qquad \forall i = 1, ..., |\mathcal{T}| - 1 \quad (4.32)$$

Constraints 4.31 and 4.32 assure that if two or more wagons on the inbound train will be SO, then the respective adjacency variables will be activated and counted in the objective function. This is done by summarizing in a single variable γ_i both the demand, the optional and the maintenance shunts performed, and by forcing the activation of $adj_{i,i+1}$ only if both γ_i and γ_{i+1} are equal to 1.

$$\sum_{j \in \mathcal{S}} z_{i,j} code_{S_j} + (1-\gamma_i) code_{in_i} = code_{\text{out}_i} \qquad \forall i \in \mathcal{T} \quad (4.33)$$

Eq. 4.33 is an optional one, not strictly necessary to optimize the problem, but quite useful to keep track of wagons' codes that will be on the outbound train once all the SO and SI operations are performed. If the wagon i-th has been replaced by activating $z_{i,j}$, this constraint associates the wagon j-th's code to the position i-th, alternatively, the wagon i-th's code remains unchanged. This constraint becomes important when giving back the solution to the simulation environment, as it will speed up the process of computing the different states of each wagon.

4.5 Shunt-In Policies

In this section, we propose different shunt-in policies and discuss the composition and add-ons to the multi-component objective function (4.4)-(4.6). The shunt-in model's basic version only considers the time cost required to move a wagon i-th into the position j-th, defined as $c_{s_{i,j}}$. This cost is the matrix that is composed of the routing times defined as the assumption from Section 4.3.2. While this approach could be limiting from a strategic point of view, the shunt-in policies allow the model to exploit different features impacting the number of future shunting operations. These features are directly linked to assumptions (Section 4.3), such as those on clustering, and have proved to be impactful also on the wagon fleet size and average mileage performed by each wagon. The general structure of a shunt-in policy, defined as the *Add-on SI Policy*, is described in Equation 4.34:

$$\min ... \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{R}} (\alpha * W_1 + \beta * W_2) z_{i,j} ... \qquad (4.34)$$

Based on the policy criteria, weights W_1 and W_2 can assume different meanings, to steer the model for an objective function that covers more strategic than tactical/operational goals. Specifically for this study, we will refer to W_1 as the cost $c_{s_{i,j}}$, while W_2 will change depending on the policy applied, and are both normalized for measurement unit reason. α and β are complementary weights describing the temporal state of the system between the two stages, the shunt-out and the shunt-in respectively. α represents the percentage of operational time used by SO, while β represents the remaining percentage of the operational time that SI can use. Depending on the value assumed by α and, consequently, by β, the solver will decide whether to weigh more the cost $c_{s_{i,j}}$, therefore opting for a more greedy approach, or the policy criteria W_2.

As a note from the author, it can be pointed out that the following policies are designed to be both simple and pragmatic, as they rely on metrics that can be easily derived from the data that has been collected. This was a specific design choice, to allow for easy and no-cost implementation from the practitioners' side fighting the classical resistance that comes with overly complex models.

4.5.1 MIN

This policy aims to shunt in wagons from the shunting yard with the minimum virtual rate n_{ms_j}. Therefore, the *Add-on SI Policy* is modeled as follows:

$$\min \ldots \sum_{i\in\mathcal{A}} \sum_{j\in\mathcal{R}} (\alpha * c_{s_{i,j}} + \beta * n_{ms_j}) z_{i,j} \ldots \tag{4.35}$$

Based on (4.35), if the percentage of operational time required by the shunt-out operations is predominant, the solver will opt to shunt in wagons with a lower cost $c_{s_{i,j}}$. Otherwise, the solver will be directed toward the policy criteria, choosing wagons with the lowest n_{ms_j}. The goal of this policy is to keep the average current mileage of the departing train $\mathcal{T}$ low by assigning the wagon with the lowest current mileage ms_j.

4.5.2 AVG Long - Short (AVG L-S)

The AVG L-S policy aims to make the most of the wagon's mileage capacity, based on its services assignment record. The goal is to balance the number of long trips ($long_j$) and short trips ($short_j$) assigned to each wagon by utilizing terms: the *degree of unbalance* δ_j as described in Eq. 4.36, and the *threshold tresh*. The threshold *tresh* is used to determine whether a trip r_T is considered *long* or a *short* trip and strongly depends on the instance. The goal is to choose a suitable $j \in \mathcal{S}$ with the maximum degree of unbalance, defined as:

$$\delta_j = n_{long_j} - n_{short_j} \tag{4.36}$$

The objective function will therefore become:

$$\min \dots \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{S}} (\alpha c_{s_{i,j}} - \beta((-1)^{long} \theta_j)) z_{i,j} \dots \tag{4.37}$$

Where:

- *long* is a metric that is 1 if the trip is a *long* trip, 0 otherwise.
- θ_j is the normalized vector of unbalances of the wagons in the shunting yard.

If a wagon has performed a large number of long trips, it's likely that it has used much of its max mileage ms_{max_j} and therefore will be assigned to a short trip to maximize its remaining usage.

This is a policy that has been developed based on the current policy applied by CFL to keep the wagons rotating to leverage the mileage performed by each wagon.

4.5.3 NCLD

The NCLD policy, which stands for *miNimum distanCe simiLar Deadline*, aims to optimize shunting operations by minimizing the difference between the *index of use* of individual wagons and the average index of use of wagons in the inbound train. The *index of use* is defined as the ratio of distance traveled by a wagon to its $m_{max_w}, w \in \mathcal{T} \cup \mathcal{S}$ before maintenance. The index of use is computed as Eq. 4.38

$$use_w^{ind} = \frac{a_w + r_T}{m_{max_w}}, \forall w \in \mathcal{T} \cup \mathcal{S} \tag{4.38}$$

This is computed for both the wagons on the inbound train (α_w) and those in the suitable set $\mathcal{S}$ associated with the train T and the type of wagon in position j (β_w).

The goal is to maximize both wagon utilization and create as many clusters of wagons with a similar index of use as possible, such that we can increase the likelihood of minimizing the shunting to be performed for maintenance. This is achieved by aligning the β_w of selected wagons with the average α of non-exchangeable wagons within the train. The final element necessary for this policy to work is the definition of the C, set of n wagons on the train not to be shunted.

If wagons with a homogeneous index of use are shunted-in, then it is likely that, when one of the wagons requires maintenance for mileage reasons, a cluster will be created as more wagons will need maintenance. This means that multiple wagons will be shunted out altogether.

The add-on in the multi-component objective function will be:

$$\min ... \sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{S}}(\alpha c_{s_{i,j}} + \beta|A_{SO} - A_{SI}|)z_{i,j}... \tag{4.39}$$

A_{SO} is defined as the average virtual rate on the inbound train once all the shunt-out operations have been performed; A_{SI} is defined as the average virtual rate of the wagons shunted in, as expressed by (4.40) and (4.41).

$$A_{SO} = \frac{\sum\limits_{i\in\mathcal{T}}(1-\gamma_i)v_i}{\sum\limits_{i\in\mathcal{T}}(1-\gamma_i)} \tag{4.40}$$

$$A_{SI} = \frac{\sum\limits_{i\in\mathcal{T}}\sum\limits_{j\in\mathcal{S}} z_{i,j}}{\sum\limits_{i\in\mathcal{T}}\sum\limits_{j\in\mathcal{S}} z_{i,j}} \tag{4.41}$$

If β is higher than α, (4.39) will minimize the distance between the average virtual rate of the left wagons on the outbound train and one of the wagons shunted in.

This method results in a leaner solution, reducing the number of wagons needed in the long term to fulfill services, and promoting the reusability of the wagons. However, a notable side effect is increased shunting operations due to overuse. As the policy tends to favor wagons with non-zero mileage over those with zero mileage, the policy tends to use fewer wagons on average, preferring those that have $\beta_w > 0$. This preference leads to an overstressing of all wagons that have been used previously, therefore requiring more shunting operations due to maintenance reasons on a fewer number of wagons.

4.5.4 Reserving

The Reserving policy leverages the concepts of β_w from Section 4.5.3, *tresh*, and *long* from 4.5.2. It evaluates two scenarios based on the trip length r_T of the train T. If r_T exceeds *tresh*, indicating a long trip, the policy prioritizes using the wagon with the lowest mileage in the depot, as indicated by β_w. On the other hand, if the trip is short, the policy chooses the most used wagon with enough residual mileage to complete r_T. This policy promotes a balanced rotation of the fleet and maximizes the number of trips a wagon can complete before maintenance while ensuring a homogeneous usage of the rolling stock.

The add-on in the multi-component objective function will be in this case:

$$\min ... \sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{S}}(\alpha c_{s_{i,j}} - \beta((-1)^{long}\zeta_j))z_{i,j}... \tag{4.42}$$

With ζ_j being the normalized vector of β_j for the shunting yard.

4.5.5 Random Policy

To compare the different policies with the typical operations within a shunting yard, we implemented what we refer to as the random policy. This criterion selects wagons from the shunting yard without taking mileage, nor trips performed, into account, but only looking at the closest to the inbound train. It's a simple yet effective way to represent the actual, non-structured behavior often observed in shunting yards, and it serves as a baseline against which the other, more targeted policies can be assessed.

Finally, the add-on in the multi-component objective function for this policy will be:

$$\min \ldots \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{S}} c_{s_{i,j}} \ldots \tag{4.43}$$

4.6 Simulation

In this section, we present the simulation framework. The simulation environment has been developed by the University of Luxembourg in partnership with CFL Multimodal, which has shared data for the case study as well. The event-based simulation framework, developed in Python, aims to reproduce the mechanisms, events, activities, and objects of a shunting yard. The entire simulation process is depicted in the flowchart presented in Figures B.1 and B.2 of Appendix B. This visual representation provides a clear and concise illustration of all the steps involved in the simulation.

We opted for a simulation approach mainly due to our research objectives and the challenges associated with using a MILP model for long-term scenario analysis, which could lead to scalability issues. Therefore, we propose a simulation approach paired with a MILP model to solve each instance optimally, as explained in Section 4.6.1, with all the assumptions and clear limitations, which allowed us to test and analyze different scenarios over time, offering us the flexibility to adjust and refine the simulations as necessary.

4.6.1 Interaction between the Mathematical Model and the Simulation Environment

The mathematical model used in this simulator is the one presented in Section 4.4. The MILP model defines which wagons have to be shunted-in and out based on both the optimal solution. Once this has been computed, the simulator will subsequently handle the operations to meet the outbound train service requirements as outlined in the model.

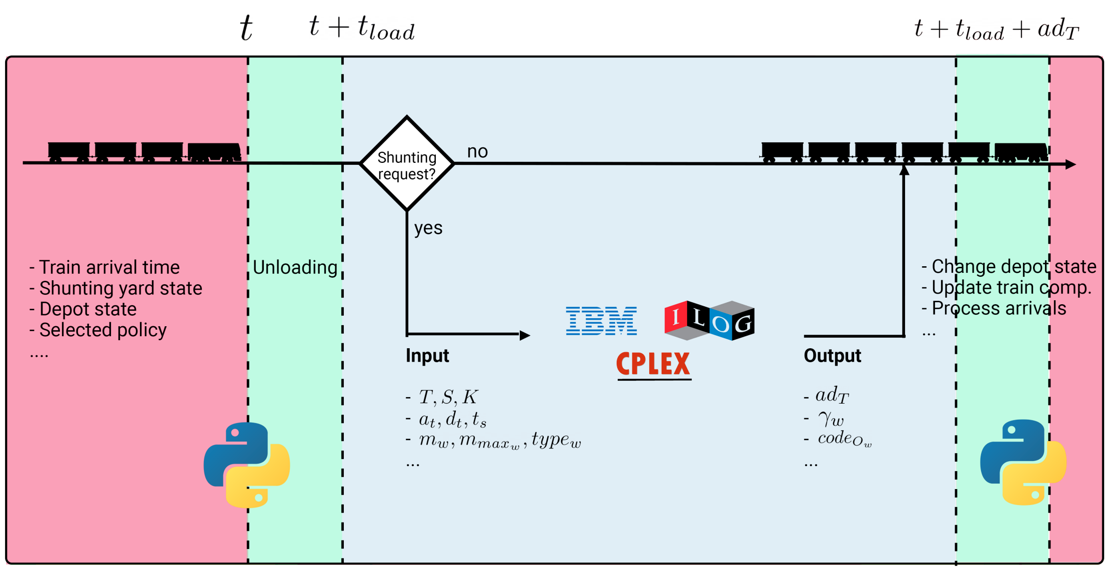


Figure 4.2: Integration between the event-based simulator and the MILP model.

The event-based simulation manages all the high-level information and events that trigger within the shunting yard, including but not limited to:

- Arrival and departure of trains.
- Loading and unloading time, adding them to the current time as well as managing all the queues for the different shunting yard areas.
- Shunting yard and workshop states.
- Data tracking of the policies KPIs.

The simulation environment aims to reproduce the behavior of the shunting yard. The train arrives and gets dispatched following the flow as explained in Figure 4.3. Trains that do not require shunting are still processed and added to the queue if in need of loading and unloading. The simulation starts a shunting operation when is required for either demand or maintenance purposes. The necessary train data is collected and formatted into arrays, sets, and parameters that can be processed by the MILP model. This pre-processing step is followed by the use of the IBM CPLEX solver accessed through the CPLEX Python API, which receives all the information required by the optimization model. Upon receiving a solution from the MILP, the simulation updates the system state by performing the shunts on the trains, computing new arrivals, updating performance indicators (KPIs), and preparing the shunting yard for the next trains. If the optimization model proves infeasible, it indicates that the operations cannot be executed as planned. In such circumstances, there are two potential outcomes:

- If the unfeasibility is due to a lack of rolling stock in the shunting yard, the train will be rescheduled for shunting as soon as the necessary wagons become available. Once the operations have been completed, the train will be rescheduled again.
- If the unfeasibility is caused by a cancellation request, meaning that the model is unable to complete the operations within the given operational time frame available before cancellation, the simulation will proceed to cancel the planned outbound train, while still conducting the shunting operations. The canceled train will be rescheduled for a future rotation.

In both cases, as for the current state of the simulator, no replanner has been introduced for the timetabling.

4.6.2 Input Requirements

The simulation environment requires comprehensive input data and distributions to be able to generate scenarios and assess them. This includes information tracked by practitioners on train movements, such as the number of trains in transit, frequency of operations, and weight and distance of transported freight. In addition, data on the maintenance threshold and resources available must be included to model the impact of maintenance constraints. The input data are listed below:

- A timetable with all services τ, providing specifics on the freight demand and the layout of the wagons for each train.

- A comprehensive list of wagons $W = \mathcal{T} \cup \mathcal{S}$.
- A set of standard time parameters for specific operations.
- Information on the maintenance threshold.
- Data on the freight load for loading into outbound trains.

The simulation operates on a per-train basis, processing each train based on its arrival time. The flowchart detailing the sequence of events for each train entry is presented in Figure 4.3.

4.6.3 Shunting Yard Areas

Each area in the shunting yard is represented through an M/M/1 queue with FIFO serving and an infinite buffer between the areas. Given a train T arriving at time $t_{arr} \in \tau$, as soon as it enters the system, it is moved into the queue of the *Arrival/Departure yard* (from now on referred to as ADY) where some operations as for the indication of CFL are performed. After these, T is moved into the *Train Loading/Unloading area* (from now on referred to as TLA), if is free. Each loading/unloading operation requires on average t_{load} time to be fulfilled. We do not consider loading and unloading operations in the overall shunting operation count, since we are addressing the impact of the maintenance on the rail system and the two problems are not mutually dependent. The TLA, like the other zones, is designed to have an unlimited buffer to accommodate the shunting yard's availability for operations. This buffer serves as a waiting area for post-unloading and pre-loading operations, as well as scheduled activities. After unloading the goods, the train is directed to the queue in the Shunting Yard for shunting operations. This queue also follows an M/M/1 queuing model for demand matching or maintenance purposes. For each SISO operation, respectively, a time t_{sout} and t_{sin} is required. Once all the shunting operations are performed, the train is moved again to the queue of TLA for the loading operations, after which, it is ready for departure. The allowed time window for the departure ranges between the departure time scheduled and a max delay of dd_T minutes. For our specific case study, the threshold used to decide to cancel a trip is 180 minutes after the latest possible operational time. This is specific to our case study and can be generalized if needed. The train is canceled in case it exceeds this deadline and assumed as rescheduled before it arrives at the next station.

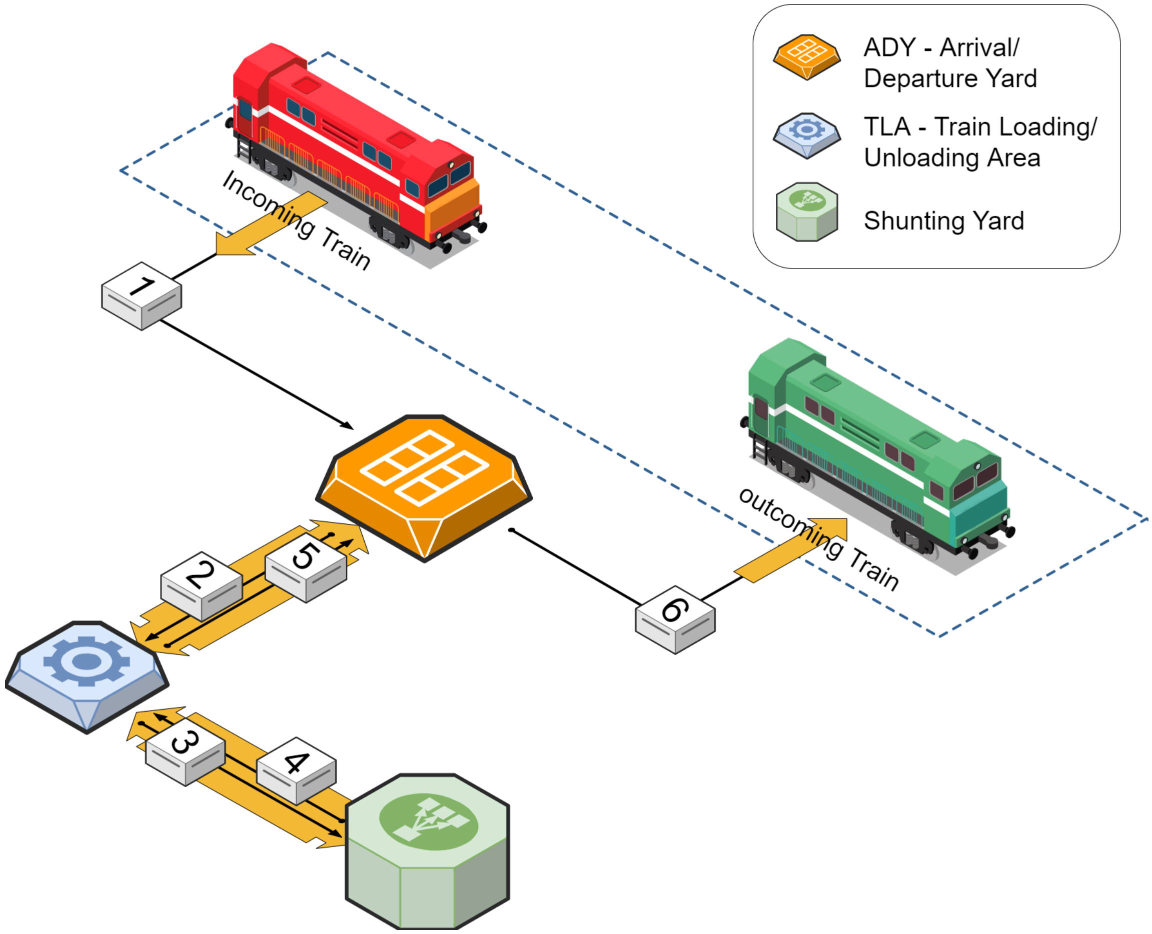


Figure 4.3: Flowchart of the operations performed in the simulator for each train.

4.7 Case Study

Our case study considers the 2020 timetable for freight trains at the Bettemburg Eurohub Sud Terminal, a shunting yard managed by the freight forwarding operator CFL Multimodal, which plays an important role in Europe by connecting various EU countries (Figure 4.4) due to its central location. Bettembourg Eurohub Sud Terminal, one of the 446 intermodal terminals across Europe [3], emerged as a significant hub for freight rail traffic in 2021, handling 2.6 billion tonne-kilometers of intermodal freight traffic, representing roughly 1% of the total freight rail traffic in Europe [5][4]. In the same year, the yard saw the passage of 1384 trains, with approximately 600 recorded shunting operations performed.

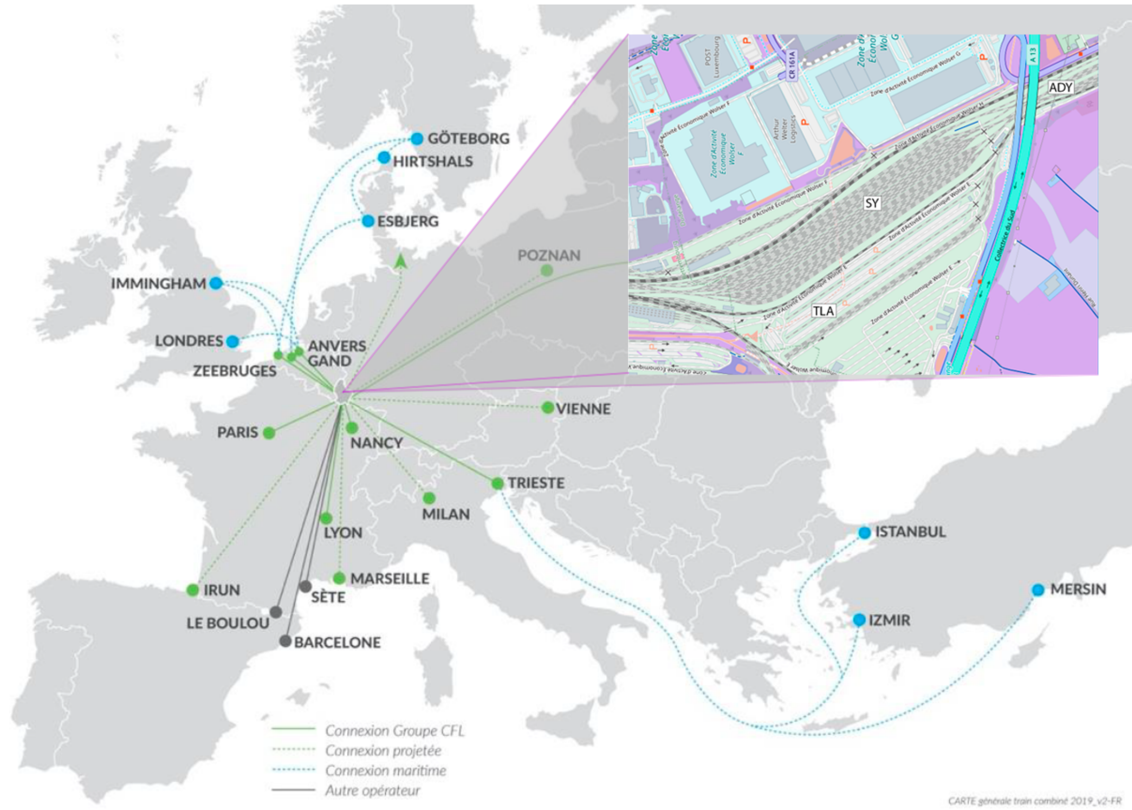


Figure 4.4: The Bettemburg Eurohub Sud Terminal key location. In this case, the shunting yard is connected to an Arrival/Departure yard (ADY) and a Train Loading/Unloading Area (TLA).

Bettembourg shunting yard handles both SWL and IWL traffic, with a specific focus on the latter. It is equipped with a classification yard and a hump specifically designed for classification. This hump is primarily employed for the SWL traffic. Intermodal units, on the other hand, are predominantly managed using shunting locomotives to ensure safe and efficient operations. While the classification yard serves both types of traffic, for IWL specifically, shunting operations are performed in case of changes in the demand composition, or parked when maintenance operations are required on the train within the yard.

4.7.1 Case Study Parameters

For this study, the same 2020 year timetable was used to simulate inbound and outbound trains, including information on train rotation, wagon demand, and destinations of outbound trains. To assess the long-term impact of the maintenance, and how to address them through the concept of shunt-in policies, this timetable has been extended by 20 years. The simulator also implements a predictive model, previously developed by [69], to estimate the trip delays of incoming trains based on factors such as the weight of the wagons and the distance of the train trip. In this study, we classify wagons into two categories, SIMPLE and DOUBLE, based on the wagon models and tare, as for the assumptions in Section 4.3. To fit the initial distribution of types provided by practitioners (in terms of SIMPLE wagons and DOUBLE wagons, 19% and 81% respectively), a group of 1400 wagons was chosen to be available inside the simulation, setting their initial $m_i = 0$. This was done to allow a warm start and understand the complete behavior of the policies, rather than just the impact of these. Moreover, the maintenance mileage limit for this study has been set to $m_{max_w} = 150000$ km, $\forall w \in \mathcal{T} \cup \mathcal{S}$, and maintenance time has been set to t_{maint}=3 days. It is important to emphasize that any available wagon left unused in the simulation will not be considered in the following postprocessing. Conversely, if a wagon is employed at least once, it will be accounted for in the postprocessing and thereby be deemed as 'owned'. For each of the shunt-in policies presented, we run a simulation for both the no-maintenance and maintenance scenarios, with the above-mentioned specification.

4.7.2 No - Maintenance Scenario

To evaluate the impact of maintenance on different KPIs through the shunting policies, we conduct a benchmark analysis in which we simulate, for each policy, the different No-Maintenance scenarios, setting them as a baseline. We aim to compare the Maintenance scenario, viewed as a better approximation of the real-world scenario, with practitioners' previsions which may be underestimated, leading to additional costs to be sustained in practice. In addition, we aim to demonstrate that the exclusion of the maintenance constraint in shunting operations may impede the detection of underlying problems, leading to reactive rather than proactive measures to address issues.

4.8 Results and KPIs

Table 4.2 presents the results of simulations for the 2020-2040 period to evaluate the impact of integrating maintenance with rolling stock management on the wagon fleet size, shunting operations, and annual mileage per wagon for each policy. For this study, we have chosen to analyze all the KPIs that have been identified for the strategic problem in RO2, Section 3.2.

	Min	**NCLD**	**AVG L-S**	**Reserving**	**Random**	
Shunting Ops.	16434	16434	16434	16434	16434	No maintenance scenario
Maintenance Ops.	0	0	0	0	0	
Wagons Used	1199	276	269	1198	1226	
Annual operations on each wagon						
Mean	20.29	88.68	87.46	20.29	17.09	
Median	13	71	30	13	15	
Variance	21.55	75.78	143.59	21.55	19.43	
Annual Miles Performed per wagon (km)						
Mean	5092.12	27717.93	28180.52	6895.58	6740.51	
Median	1902.1	21782.3	8318.5	1320.9	6055.5	
Variance	8941.44	17272.87	41142.57	12108.2	6505.1	
Shunting Ops.	16850	18427	18514	17050	17523	Maintenance scenario
Maintenance Ops.	1326	2288	2615	1429	1660	
Wagons Used	1335	1385	791	1338	1409	
Annual operations on each wagon						
Mean	19.23	19.36	34.06	19.25	16.64	
Median	12	10	10	13	12	
Variance	21.74	38.73	86.24	21.55	17.61	
Annual Miles Performed per wagon (km)						
Mean	17370.46	17458	33986.09	15941.29	17013.11	
Median	14234.42	12490.8	13951.28	12807.19	14566.16	
Variance	19387.4	33683.83	62047.3	18194.7	13706.26	
Shunting Ops.	2%	11%	11%	4%	6%	Comparison
Wagons Used	10%	80%	66%	10%	13%	
Annual Miles Performed	71%	37%	17%	57%	60%	

Table 4.2: Performance comparison between the policies between the No-Maintenance and Maintenance scenario.

4.8.1 No Maintenance Scenario

Shunting Ops

Since no maintenance operations have to be performed, the same number of shunting operations is observed in each policy, given that only the demand constraint has to be met. Nevertheless, this does not apply to the number of wagons used, since each policy can be greedier regarding wagons to shunt-in (Min, Reserving, Random) compared to more exploitative policies (NCLD, AVG L-S).

Fleet Usage

The fleet size varies based on the criteria used to determine which rolling stock is shunt-in by each policy. For instance, NCLD and AVG L-S prioritize a limited set of wagons with high residual mileage, while the other policies tend to utilize a wider range of rolling stock. Since both the composition and the length of the outbound trains can change, MIN and Reserving policies tend to bring in wagons with lower mileages, while wagons removed due to demand matching are likely to be parked in the shunting yard and infrequently used. These policies typically avoid wagons with higher mileage until new wagons become available, leading to a lower average mileage per wagon as observed in the simulation results. This type of trend is also observed in the statistically low performance in terms of *Mileage Performed per wagon.* The *No Maintenance* scenario suggests that NCLD and AVG L-S might have the best wagon fleet usage, based on the high mean, median, and variance for shunting operations.

Mileage Performed

NCLD presents a higher mean compared to Min and Reserving, which is paired with a close median and a similar variance, translating into an almost symmetric distribution of mileage across the fleet, meaning that its fleet covers more km efficiently, using fewer wagons. AVG L-S performs similarly in terms of mean, but with a distant median and high variance, suggesting that fewer wagons are used more frequently, confirmed by the slight decrease in the number of wagons used and higher statistics on operations on wagons compared to NCLD.

4.8.2 Maintenance Constraint Scenario

By including in the shunting operations the maintenance constraint, therefore constraining the maximum mileage for each wagon, significant changes occur.

Shunting Ops

The number of shunting operations performed increases from 2% to 11%, and this disparity can be directly attributed to maintenance constraints. However, it is challenging to determine the exact number of shunting operations performed for maintenance purposes, as the definition of shunting operation does not allow for a clear distinction: if multiple rolling stocks are moved together and, within this group, there is a wagon that must be removed for maintenance reasons, it is incorrect to say that this cluster has been created for maintenance. Instead, we can assess the impact of maintenance constraints on the overall number of shunting operations. In this case, a higher number of shunting operations due to maintenance constraints and higher annual mileage per wagon may suggest better fleet management in terms of wagon requirements. However, uneven utilization of the wagon fleet, as seen in AVG L-S, is not desirable.

Maintenance Operations

The number of maintenance operations performed by each policy varies based on how exploitative it is. With Random setting the benchmark, we can see that Min and Reserving perform fewer maintenance operations on the wagon, tending to spread the effort among different wagons. On the other hand, NCLD and AVG L-S prefer used wagons to new ones, performing, therefore, more operations and using less rolling stock on average compared to the other policies.

Fleet Usage

The number of used wagons required to perform the simulations increases, compared to the benchmark, due to maintenance activities causing fleet unavailability. Despite AVG L-S showing a higher number of shunting operations, it also utilized nearly 40% fewer rolling stocks, resulting in increased operational expenses but significantly reduced overheads (leasing costs per wagon) and variable costs (e.g. storage costs).

Mileage Performed

In terms of mileage performed, AVG L-S presents almost twice the mean and variance compared to the other policies. This could be both an advantage and a disadvantage: the distant mean and median, coupled with the high variance, indicate an uneven utilization of the fleet, where a smaller number of wagons are used more frequently, requiring more maintenance and shunting operations to meet the policy's demands. This is attractive for companies that have a limited owned fleet and rely heavily on leased wagons, but it could also lead to overuse of the rolling stock, causing excessive wear and additional maintenance. Overuse of wagons may also increase the risk of breakdowns.

4.9 Discussion

In this section, we will discuss the presented results focusing on how these can be read in terms of assessment of the impact of the maintenance constraint, so a "vertical" analysis between the No-Maintenance and the Maintenance scenario, the addressment of it, through a breakdown on the policy performance in terms of the analyzed KPIs.

4.9.1 Assessing the Impact of the Maintenance Constraint

Looking only at the No-Maintenance scenario, therefore neglecting the application of the maintenance constraint, the conclusion that could be drawn is that AVG L-S and NCLD outperform Min, Reserving and Random, especially in terms of fleet size. This leads to more operations performed on average on each wagon and more miles performed on average, highlighting the optimization in the fleet use of these two policies.

When we examine the variation in all the KPIs after implementing the maintenance constraint, the extent of the previous undervaluation due to the neglection of the constraint becomes evident. This underestimation's impact is not just a numerical adjustment, but it reflects notably on the practical aspects of our study, such as the distribution of annual mileage performed by each wagon. A concrete illustration of this underestimation effect can be seen in the distribution of annual mileage performed by each wagon, as presented in Figure 4.5.

For Min and Reserving in the No-Maintenance scenario, Figure 4.5 shows high peaks corresponding to the Mean Annual Mileage Performed per Wagon in Table 4.2, which represents the

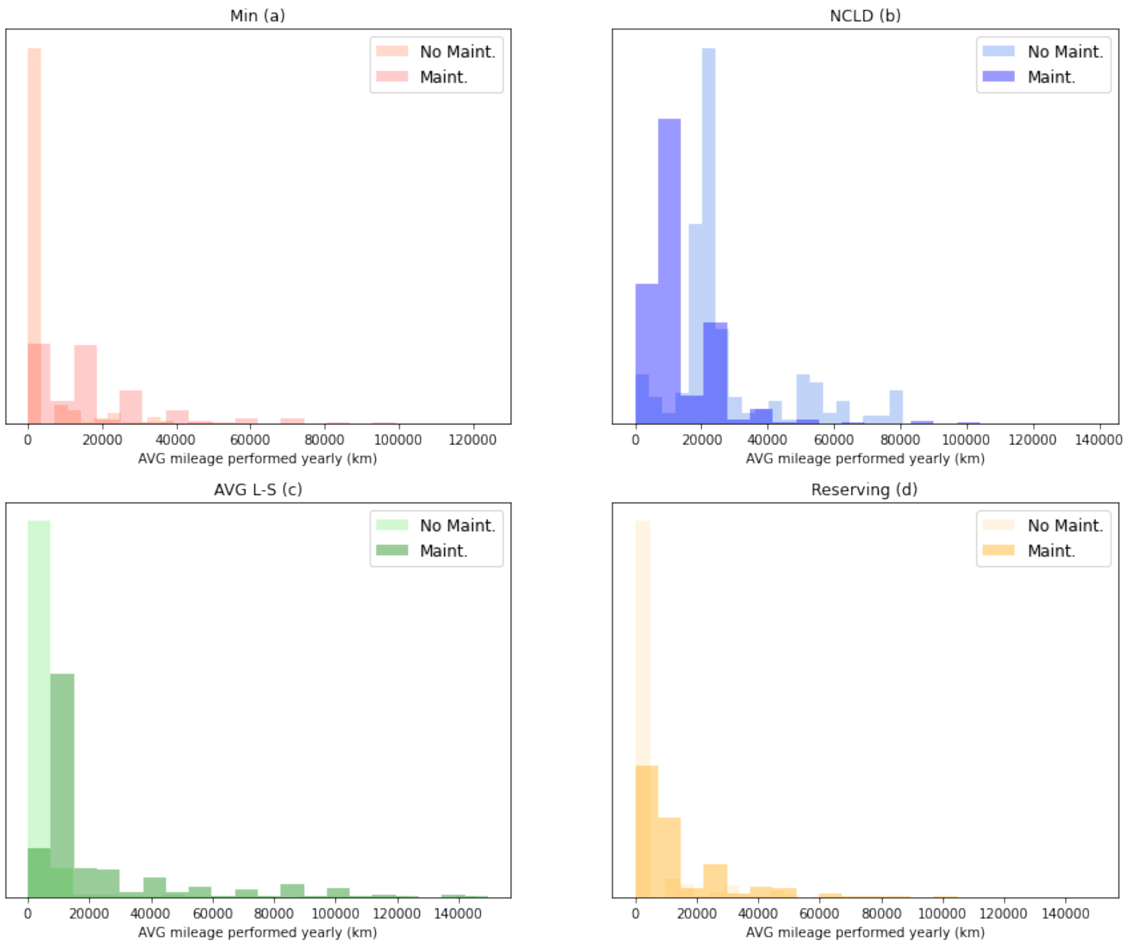


Figure 4.5: Annual Mileage Distribution of each wagon between Benchmark and Maintenance Scenario

fact that a limited number of wagons are utilized only to meet occasional demand fluctuations, leading to many wagons covering minimal distances, and therefore getting counted during the post-processing. In contrast, the majority of the fleet operates with a consistent annual mileage. This behavior changes completely when looking at the Maintenance Scenario, where a more balanced fleet usage is observed by these two policies. Interestingly though, the Operations performed on each wagon do not increase so much between the No-Maintenance and the Maintenance scenarios (2% and 4% respectively), given the high demand shunts performed in the No-Maintenance scenario compared to the additional effort required in the Maintenance scenario. Overall, the two policies seem to perform in a very similar way. The increase in the spread of all the distributions passing from the No-Maintenance to the Maintenance scenario is strongly connected to the increase in the number of wagons required for performing the simulation, which remains stable for MIN and Reserving as the turnover of wagons for maintenance reduces the increase of wagons by working on the average mileage per wagon while increasing for NCLD and AVG L-S. The maintenance limitations lead policies to opt for wagons that may not meet their decision criteria, resulting in up to an 80% increase in the fleet size for the NCLD policy and a 66% increase for AVG L-S, spreading the distribution of annual mileage performed by increasing the "owned" fleet usage (Section 4.7.1) and requiring more effort from an increasing number of wagons, concluding in more maintenance efforts which were unplanned, and therefore a more widespread annual mileage distribution by all the policies. Nonetheless, both these two policies in the Maintenance scenario show a quite compacted use of the fleet, showing a more controlled behavior even under maintenance constraints. Looking at the Min and Reserving policies in the No-Maintenance scenario, which can be seen as the initial practitioner's perspective, what can be observed is that the fleet's wagon requirements look overestimated compared to NCLD and AVG L-S, potentially leading to the other two policies as a choice. Yet, this initial overestimation does not find its validation when we introduce the maintenance constraint. These two policies present a smaller differential in terms of Wagon Used among the policies. Nonetheless, this initial corrected view on the No-Maintenance scenario for the fleet requirements for Min and Reserving significantly underestimates the fleet's annual mileage performed distribution, as seen from the Annual Miles Performed differential for these two policies in Table 4.2.

Looking at the comparison in Table 4.2, what can be seen is that, even though the shunting difference assesses itself between 2 and 11%, which can be directly connected to the maintenance operations, the highest difference in terms of impact lies within the core business of the freight companies, which are the mileage to be performed each year by the wagons and the number of wagons required to fulfill all the services.

4.9.2 Addressing the Impact of the Maintenance Constraint

To address the maintenance constraint, we want to analyze if there is a policy that outperforms the others in terms of different KPIs. This may completely depend on the points of view, and needs of the company, with an example provided in Figure 4.6, which represents the average mileage of the wagons available inside the shunting yard at any point in the simulation per policy for the Maintenance scenario.

This parameter can be read as the *resilience* of the overall wagon fleet, where a more linear trend corresponds to a control on the fleet distribution mileage through time, which results in multiple benefits such as a better answer to eventual disruption and less wagon fleet requirement to fulfill all services. In grey, for all the plots as presented in Figure 4.6, is represented the Random behavior for comparison reasons. In this case, what is observed is that for SIMPLE wagons Min, NCLD and Res (Fig. 4.6 (a),(c),(g)) suffer all from a bullwhip effect, indicating an underestimation of the SIMPLE wagon fleet requirement and a constant overuse of this portion of the fleet. This trend has not been observed in AVG L-S, which provides a more stable behavior. In this context, Random policy seems to perform slightly better than Min and Res, but still with an initial bullwhip effect observed. For the DOUBLE wagons instead, while the size of the DOUBLE fleet allows for a steady growth in the actual mileage available in the depot for Min and Res (Fig. 4.6 (b),(h)), which will eventually result in the same bullwhip effect as observed in the case of the SIMPLE wagons, two interesting behaviors arises. First, the NCLD trend (Fig. 4.6 (d)) shows complete control of the fleet, given that the criteria will always prefer a wagon with $m_i > 0$ compared to a fresh wagon, only reusing the new wagons when there is no availability. For AVG L-S (Fig. 4.6 (f)), a similar trend is observed, where a steady state is reached, together with a more controllable situation in terms of wagons. In both cases, a reduction of the wagon fleet ownership could be applied, still resulting in a stable scenario. In this context, the Random policy appears to perform slightly better than Min and Res, providing a somewhat stable behavior in the fleet average mileage. However, it's important to note that this stable trend observed may be more attributable to the specific setup of the instance (as explained in Section 4.7.1) rather than a reflection of the policy's controllability

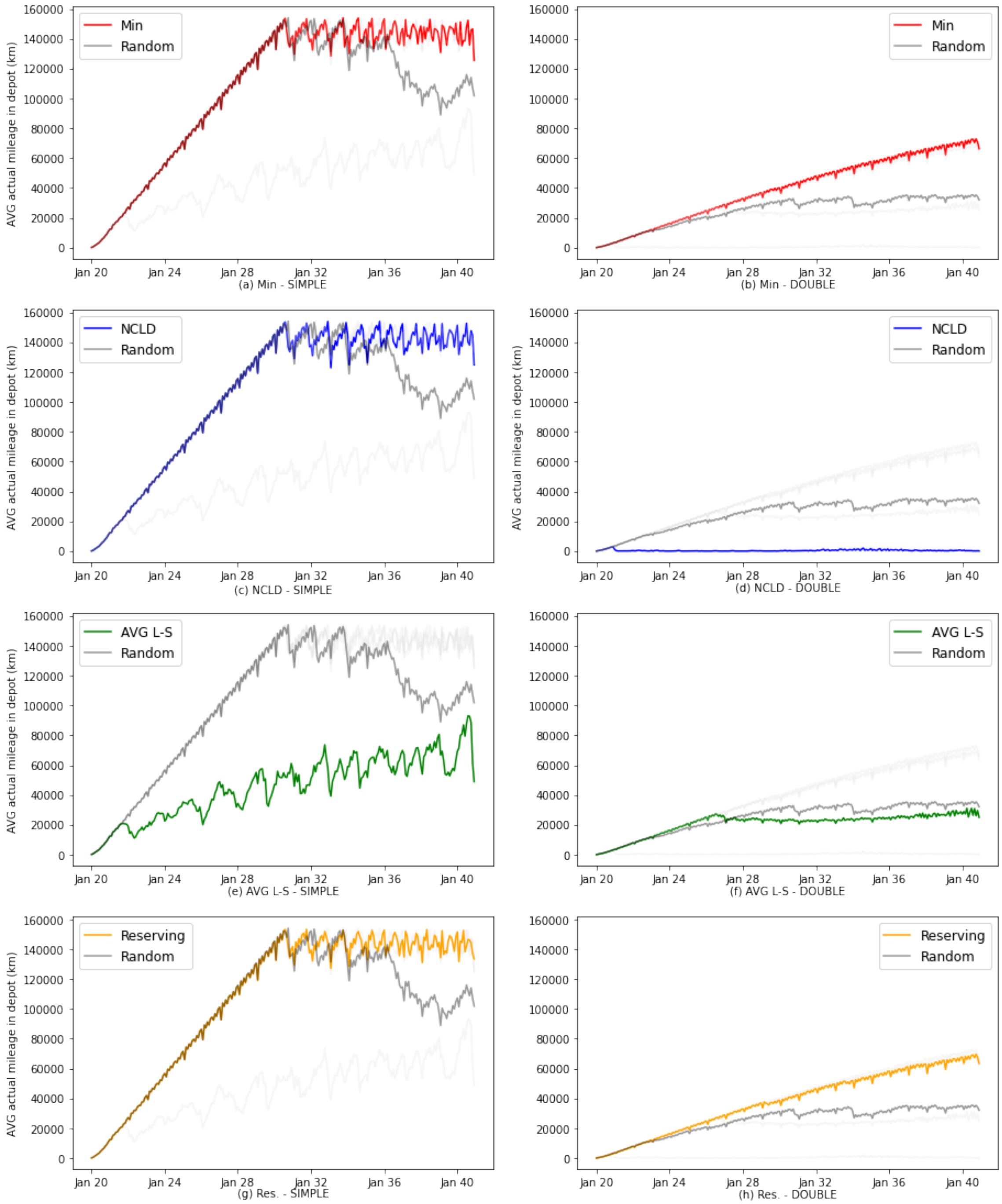


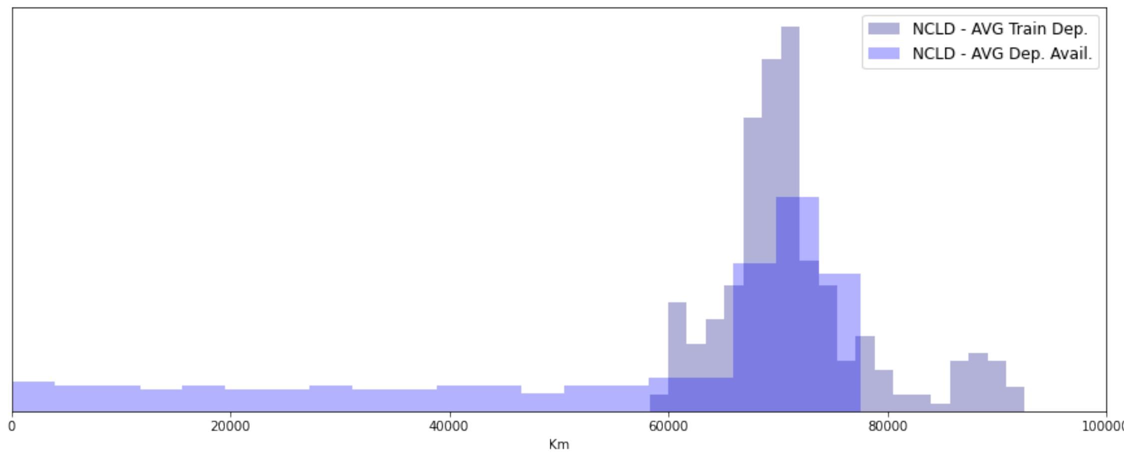
Figure 4.6: AVG mileage available in the depot per year per policy.

over fleet management. Given the Random shunt-in criterion, its performance advantage in this scenario may not necessarily translate to a consistent or intentional control strategy. To proceed further, we investigated the overall distribution of the average yearly mileage of the fleet at the end of the simulation from the NCLD and AVG L-S policies, divided into the "active fleet" (wagons in service) and the "passive" fleet (wagons parked in the shunting yard). This has not been done for Min and Res. as they have shown to perform worse in terms of fleet management and all the presented KPIs. As for this case study, Min and Reserving exhibited a pronounced bullwhip effect for both SINGLE and DOUBLE wagons, a phenomenon observed with NCLD in the SINGLE wagon case, but with much better mileage management for the DOUBLE wagons. Moreover, while their KPIs are relatively aligned, they are distinctly outperformed by both NCLD and AVG L-S in terms of operations within the fleet and annual miles performed, where NCLD and AVG L-S demonstrate a better efficiency, further highlighting the optimization and effectiveness of these two policies in comparison to Min and Reserving. To assess the acceptability of the fleet management behavior of NCLD and AVG L-S, we compared the average actual mileage within the depot (passive fleet) to the average mileage of the departed trains (active fleet), which provides an overview of the fleet's overall condition, together with the behavior of the Random policy, as depicted in Figure 4.7.

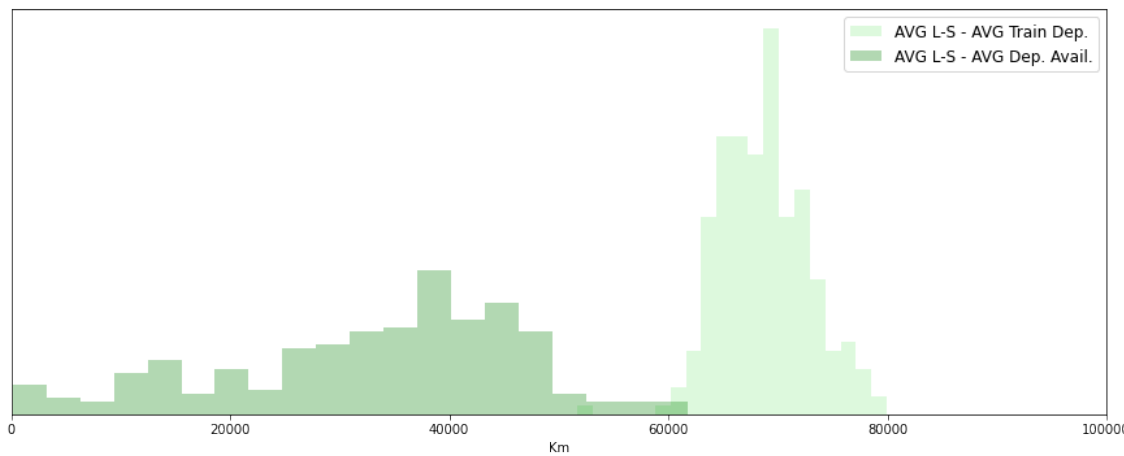
What is desirable is that these two distributions are close to each other, describing overall good management of the fleet in terms of wagons available in the depot and wagons active for services; moreover, in case of misalignment, we would still prefer to have a passive fleet mileage distribution that has:

- Either an average mileage uniformly distributed, to more likely fit better the policies criteria which
- is centered more toward the lower mileage, rather than toward the maximum available mileage.

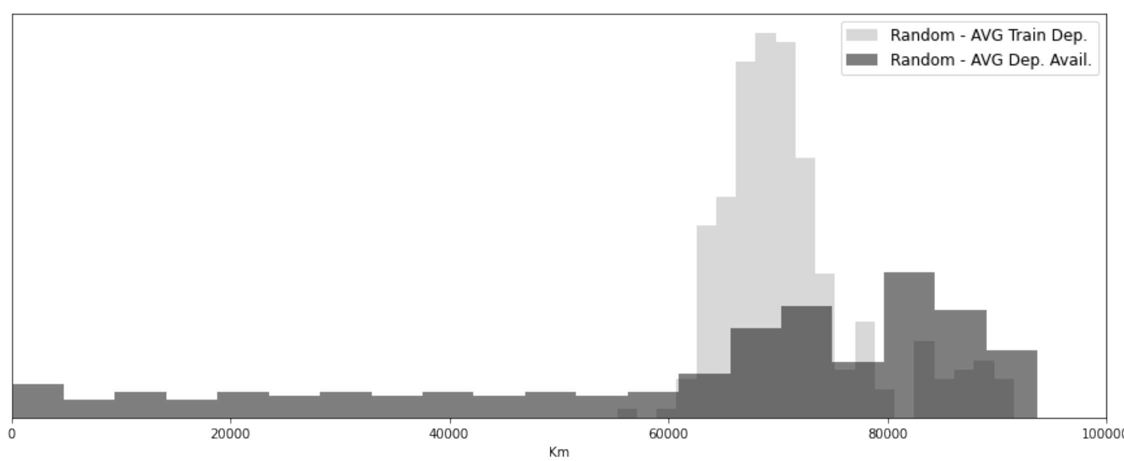
This can be seen as the interchangeability of the active and passive fleet. As can be seen in Figure 4.7 (a), NCLD belongs to the first case, providing a pseudo uniform distribution. This can be attributed to the discrepancy between types (SIMPLE, DOUBLE) in the mileage of the available wagons as shown in Figure 4.6 (c),(d); moreover, the distribution of average mileage of the active fleet shows a good alignment, showing that the wagons that are in service are not overused, which suggests good fleet management. In Figure 4.7 (b), we can see that AVG L-S shows less alignment



(a) NCLD



(B) AVG L-S



(C) Random

Figure 4.7: Distrib. of AVG mileage available in the depot vs Train Dep. AVG mileage

between the two distributions, with an active fleet mileage distribution aligned with the other two policies, and slightly better management in terms of the passive fleet mileage distribution. Looking at Figure 4.7 (c), we can see that instead the Random policy provides a similar active fleet mileage distribution as NCLD and AVG L – S. Nonetheless, the mileage of the passive fleet is much higher and flattened towards the high 80000km, suggesting that this mass will likely move onward towards the maximum mileage. Overall, it can be stated that based on our simulations both AVG L-S and NCLD provide reliable fleet management resilient to disruptions compared to the Random policy, with AVG L-S providing better performance due to the more controlled management of the SIMPLE wagons available in the depot and better control over the active fleet. Moreover, for both these policies, this might also suggest that the wagons provided for the simulation and accounted as "owned", as stated in Section 4.1.1, might have been instead only leased to reduce the economic cost to be sustained by the company. Among the policies proposed, some final consideration has to be made:

- The MIN and Reserving policies exhibit a more balanced utilization of the fleet for the Maintenance scenario, showing less variance in the annual miles performed per wagon, as for Table 4.2, which results in a more compact distribution of the annual mileage distribution, as for Figure 4.5 (a), (d). Nonetheless, their lack of fleet optimization due to their simple shunt-in criterion is reflected in the bullwhip effect for both SIMPLE and DOUBLE wagons for these two policies as for Figure 4.6.

- AVG L-S tends to use a small portion of the fleet, calling back the other wagons only to address exceptional situations where the preferred wagons are unavailable due to maintenance, providing good reliability in terms of disruption management and having better control of the wagons on service.

- NCLD policy provides statistics closer to the more conservative fleet management seen in MIN and Reserving while still compactly utilizing fewer wagons due to higher variance, similar to AVG L-S. The management of the SIMPLE fleet could be improved by increasing the size of the SIMPLE wagons "owned", while decreasing dramatically the DOUBLE ones and still having a stable and controllable fleet.

4.10 Summary

This chapter presents a detailed study on the Shunt-In/Shunt-Out (SISO) problem, focusing on optimizing railway shunting operations with an emphasis on the impact of maintenance activities. We introduce a MILP model to formalize the problem, integrating maintenance considerations into the rolling stock management to optimize shunting operations. We then present a simulation framework that evaluates the efficacy of various Shunt-In policies (Min, NCLD, AVG L-S, Reserving, Random) over 20 years, highlighting the consequences of maintenance constraints on operational efficiency and rolling stock utilization.

Key findings from the simulation under both No Maintenance and Maintenance scenarios reveal that the underestimation of shunting operations between these two scenarios can vary from 2% to 11%, with almost 10% to 80% difference in the underestimation of the fleet requirement. Moreover, under maintenance constraints, AVG L-S and NCLD policies outperform others in terms of strategic fleet utilization and operational efficiency. The inclusion of maintenance constraints significantly varies for each policy, with AVG L-S demonstrating better fleet management, achieving balanced utilization and high mileage efficiency.

Chapter 5

ML model and MILP Modelling with Condition-Based Maintenance Integration

My friend, the fates are cruel
There are no dreams, no honor remains.

Loveless Act IV

5.1 Introduction

This chapter covers RO4, Section 3.4, ”Data-Driven Modeling for Condition-Based Maintenance and Unplanned Disruption, and its Integration in MILP Modeling through a Risk-Management Approach”, and a portion of it is in review for the Optimization and Decision Science (ODS) 2024 conference. In this chapter, we describe in depth the generation of the ML model in Section 5.2, from the data selection up to the validation of the finalized model. We then describe the deterministic MILP model informed by uncertainty in Section 5.3, whose aim is to overcome the shortcomings of the previous model described in Section 4.4 and implement the ML model as an input through a risk-assessment approach for the evaluation of the shunting operations.

The idea is to incorporate inside our MILP model information about rolling stock's condition information, which might trigger potential unplanned disruption. To do this, we developed a ML model to predict these unplanned events based on real data. In this context, our data-driven model is the source of our information on potential disruptions, as its output describes the state of the rolling stock. For this study, rather than implementing the input of the ML model *as is* in our MILP, we want to perform a risk assessment on the trustworthiness of its predictions. The ML model becomes the *asset* that we are interested in assessing against its failure, as this could lead to the actual disruption of our asset of interest, the rolling stock, which can set off a cascade effect of other disruptive events.

ML models have proven effective when it comes to big data, understanding patterns within the latter, particularly when it comes to predicting future outcomes based on past events. However, some of these models also have the disadvantage of being hard to interpret, potentially resulting in unreliable outcomes. These models tend to be insensible to specific unseen events, which comes from the limitation of not always being able to implement all the necessary data that would allow perfect predictions. This limitation is worsened by data-cleaning processes, where resampling techniques like oversampling and undersampling can help balance imbalanced datasets and improve model performance, while risking introducing bias or overfitting, further reducing the reliability of these models.

Therefore, matching these models with a higher-level evaluator that accounts for the risk of a specific decision on a single event and also looks at the performance of the models in terms of predictions and different system KPIs can be beneficial for better evaluation. The proposed ML model aims to predict, based on real data, rolling stock status in terms of potential disruption, which may lead to both the interruption of normal shunting yard functioning and unplanned maintenance.

Given how diverse, severe, and unpredictable the impacts of disruption can be, being able to predict and mitigate these events requires the development and implementation of sophisticated maintenance strategies, as well as preemptive maintenance models. While in the best-case scenario, these disruptions lead to additional maintenance operations once the wagon reaches the maintenance site, based on the severity of the problem, the worst-case scenario can be illustrated by a freight train derailment that happened in Florence on 20/04/2023. Media claimed that this was caused by an axle breakdown, and the impact of the derailment ended up not only disrupting the wagon itself, but also damaging the infrastructure in the section through which it passed. As a result, the whole high-speed Italian network was blocked, with delays propagating throughout Italy and disruption registered from Milan to Rome [76]. Moreover, unplanned maintenance operations can also disrupt the maintenance schedule of companies, making it challenging to plan maintenance operations. This can result in a reactive maintenance approach, which can be more costly and time-consuming than proactive maintenance [11]. Train disruption analysis and recovery have long been recognized as crucial components of the transportation industry, with a particular focus on the area of real-time railway rescheduling [16, 93]. As this field continues to evolve, machine learning has become an increasingly popular tool for enhancing disruption analysis and recovery [46, 73]. There is still a significant gap in our understanding of the root causes of disruptions that are triggered by rolling stock breakdowns. Existing research has generally focused on mechanical issues, more comprehensive studies are needed to identify the key features that can trigger these breakdowns, taking into account both network and rolling stock attributes. By adopting a more preemptive approach that proactively identifies potentially disrupted rolling stocks, transportation providers may be able to reduce the occurrence of disruptions and minimize the associated costs, leading to greater efficiency and a more sustainable transportation system overall.

5.2 ML Model

The goal of our proposed data-driven model is to predict the condition of our rolling stock, and therefore its potential disruption, based on both rolling stock and network characteristics. We propose a binary classification supervised ML model that can be used for the prediction of eventual future hazards. We use incident data, GPS information of the rolling stock, train schedules, and other relevant data sources to train and validate the model. We conducted extensive preprocessing of data from CFL Multimodal, the freight railway company in Luxembourg, to remove outliers, standardize numerical features, and treat categorical features to ensure data consistency and quality. We first present the Data collection, preprocessing and feature engineering in Section 5.2.1, the selection of the model in Section 5.2.2, and finally the performance assessment in Section 5.2.3.

5.2.1 Data Collection, Data Pre-Processing and Feature Engineering

Incident Selection

By initially filtering the data related to the accidents, we realized from the database that the disruptions that lead to unplanned maintenance accounted for around 34% of the total recorded disruptions for the year of analysis.

These accidents are divided by *gravity*, which is defined as the impact that the disruption has on the specific train. The description of the *gravity* and the percentage of occurrence among our unplanned maintenance cases are shown in Table 5.2:

Gravity	Description	% of Occurrence
1	*Événement*: This deviation did not lead to a train delay (> 60 min) or cancellation, and had a low probability of impacting the train schedule.	71%
2	*Presqu'incident*: This deviation did not lead to a train delay (> 60 min) or cancellation, and had a medium to high probability of impacting the train schedule.	21%
3	*Incident*: This deviation led to a train delay (> 60 min).	0%
4	*Incident grave*: This deviation led to a train cancellation.	8%

Table 5.2: Different gravity with related distribution of unexpected maintenance accidents.

Data Joining and Feature Engineering

For our model, we collected data on freight rail operations for the years 2022 and 2023 from CFL Multimodal. The data includes, but is not limited to, information on rolling stock attributes, such as incident records, and train schedules and operations. The complete framework of how the data was joined and filtered is presented in Appendix C, Fig. C.1 and C.2. We merged a total of 8 different databases, each contributing to specific features of the final dataset. The most critical among these was the *incident_list_unplanned_maint*, a subset of the full incident list recorded in the year of analysis. This list contains a wide spectrum of disruptions ranging from more severe breakdowns - such as wagon failure due to unlisted reasons, wheel flats, derailments, and brake problems - to more straightforward issues such as electrical failures on the wagons or cuts in the wagon cloth. The final dataset consisted of various attributes derived from the available data, aligning with recommendations from the literature. We further refined this selection by choosing data with the lowest correlation that provided the best performance of the model. These are presented in Figure 5.1 and Table 5.4.

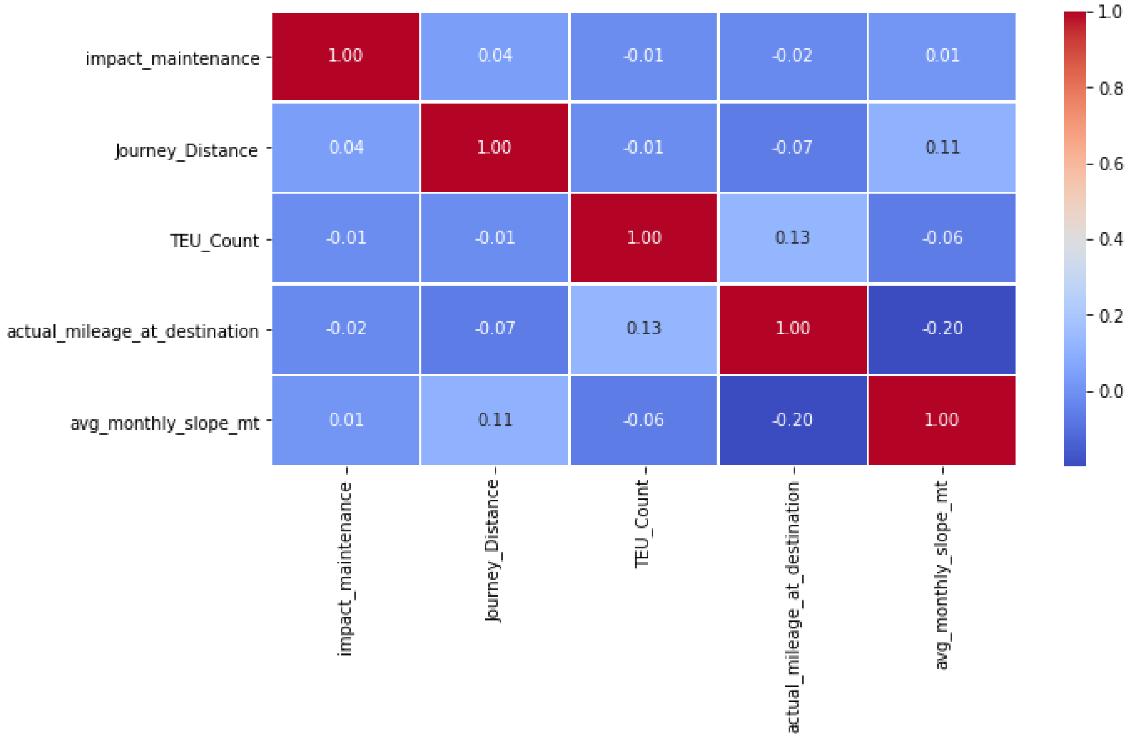


Figure 5.1: Correlation matrix of the selected attributes for training the model.

Feature	**Type**	**Explanation**	**Note**
impact_maintenance	Binary	If the rolling stock suffered of disruption	Target Feature
Journey_Distance	Numeric	Distance towards destination	-
wagon_model	Categorical	Wagon model	-
TEU_Count	Numeric	Expected number of TEU count transported by rolling for a specific destination in the month of the trip	-
actual_mileage_at_destination	Numeric	Expected actual mileage at data point	-
code_lat_long	Categorical	Data point recorded for the position of the wagon	Precision of 11.1 km
avg_monthly_slope_mt	Numeric	Estimated total elevation change for the specific rolling stock between all its recorded data points for that month	-

Table 5.4: Explanatory table of selected features.

The flag *impact_maintenance* is the binary feature that we want to predict based on the initial data, representing that a wagon has encountered a disruption at a specific point in time and space. False means that the rolling stock is in a condition to proceed with its scheduled services, and True signifies that a disruption might have happened in its road that we are not aware of. The *Journey_Distance* represents the distance that the rolling stock covers when assigned to a specific destination (r_T). The *wagon_model* represents a categorical variable detailing the specific model of the wagon. These three are the features that we inherited from the initial merged datasets. The attribute *TEU_Count* represents the mean weight carried by a rolling stock considering as a unit the TEU (Twenty-foot Equivalent Unit) based on the assigned destination throughout the month of the performed trip. This attribute has been computed for each wagon based on the mean monthly

TEU transported by a rolling stock for a specific destination $\mathcal{TR}$.

The attribute *actual_mileage_at_destination* (m_i) represents the actual mileage as recorded on a specific data point. We computed it using:

- Initial estimated mileage at the first entry in the system, computed using:
  - The initial service time of the wagon (t_{start}).
  - Mileage at initial service time ($m_i^{t_{start}}$) from the CFL dataset, inputting the missing value using the average for the t_{start},
  - average mileage performed by a specific model given a specific year $m_i^{t_{start}}$.
- GPS position of the different trajectories performed by the wagon since the initial entry in the system, computing the distance between the point using the Haversine formula.

The attribute *code_lat_long* represents the province code of the wagon's position, which has been computed by clearing the GPS position of the rolling stock using a z-score approach for detecting outliers, using a 1 decimal range (around 11.1 km precision). The data cleaning, especially due to the z-score outlier detection, sometimes removed wagon positions due to significant noise, causing sudden location changes not aligned with previous recordings, often occurring with wagons reassigned to services different from their usual assignments. The beeline distance between such disjointed entries was calculated to compensate for the data loss and added to the m_i to enhance precision. Finally, the attribute *avg_monthly_slope_mt* represents the average monthly absolute elevation performed, considering the GPS position. We computed this attribute for each rolling stock by computing the absolute vertical distance traversed performed throughout a month for each of its data points. This attribute has been specifically added to account for the heavy breaking and effort to go up and downhill.

Training and Dataset Preparation

We transformed categorical features into dummy variables using categorical encoding and scaled using the numerical features through *MinMax* normalization method. Lastly, we carried out a correlation analysis to choose the features for model training, aiming to reduce redundant predictors and avoid multicollinearity, which is presented in Figure 5.1.

The target feature is to predict whether a rolling stock, throughout the analyzed time span, experienced disruption as the one computed in Section 5.2.1. To achieve this, the most suitable data-driven class of models is binary classification, which uses supervised machine-learning models to predict binary classes. Following data pre-processing and feature engineering, we obtained a total of 50,754 records. Due to the high data imbalance, we opted for an undersampling resampling [60]. The undersampling strategy focused on keeping all instances of the minority class and a randomized selection of the remaining instances, thereby reducing the dataset size from 50,754 to 17.484 records, with a distribution of 33% data points belonging to the positive class. This was chosen over the oversampling due to the better performance of the final model. Finally, for the training phase, data partitioning was executed using a k-stratified 10-fold cross-validation strategy with an 80/20 split for training and testing datasets respectively.

5.2.2 Selection of the Model

We tested and evaluated a range of binary classification models, prioritizing Recall, the fraction of accurately identified positive instances from all actual positive instances, and Precision, the proportion of correctly predicted positive instances among all predicted positive cases. This choice is linked to our specific operational context: the cost and operational implications of a false positive - scheduling maintenance when it is not required - are far less severe than those associated with a false negative - overlooking a critical maintenance event, as also further explained in Section 5.3.1. The outcomes of this evaluation are presented in Table 5.5.

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
Decision Tree Classifier	0.8975	0.8925	0.8775	0.8262	0.8510	0.7729	0.7739	0.0680
Extreme Gradient Boosting	0.8549	0.9296	0.7353	0.8118	0.7715	0.6657	0.6675	0.6400
Light Gradient Boosting Machine	0.8296	0.9129	0.6354	0.8131	0.7126	0.5942	0.6038	1.6020
Random Forest Classifier	0.7941	0.8589	0.6218	0.7220	0.6680	0.5200	0.5232	0.4320
Extra Trees Classifier	0.7676	0.8224	0.5875	0.6740	0.6276	0.4599	0.4623	0.4410
Gradient Boosting Classifier	0.7397	0.8138	0.2975	0.7917	0.4320	0.3059	0.3677	1.7080
K Neighbors Classifier	0.7092	0.7405	0.5159	0.5707	0.5417	0.3296	0.3307	0.4640
Ada Boost Classifier	0.7036	0.7209	0.2353	0.6582	0.3456	0.2059	0.2521	0.5230
SVM - Linear Kernel	0.6701	0.0000	0.0206	0.6468	0.0398	0.0202	0.0690	0.1060
Ridge Classifier	0.6691	0.0000	0.0489	0.5337	0.0895	0.0363	0.0761	0.0230
Linear Discriminant Analysis	0.6677	0.6416	0.0624	0.5109	0.1111	0.0419	0.0778	0.1250
Logistic Regression	0.6662	0.6399	0.0491	0.4930	0.0893	0.0307	0.0626	0.1670
Naive Bayes	0.6635	0.6109	0.0395	0.4461	0.0724	0.0194	0.0417	0.0230
Quadratic Discriminant Analysis	0.4360	0.4936	0.6664	0.3276	0.4322	-0.0109	-0.0144	0.0600

Table 5.5: ML model tested.

The Decision Tree Classifier (DT) model was chosen as the go-for ML model for this dataset given that it provided the best performance out of the prioritized metrics.

DT Model

Decision Trees are a class of machine learning models used for both regression and classification models. It operates by partitioning the data into subsets based on the values of input features, effectively building a tree-like model of decisions. At each node of the tree, a feature is selected and the data is split according to a criterion such as Gini impurity. This process continues recursively, resulting in a set of terminal nodes or leaves, each representing a class label or a continuous output value.

Our DT model training was performed by incorporating an early stopping criterion to mitigate the risk of overfitting. Overfitting, where the model performs exceedingly well on the training data but poorly on unseen data, can lead to over-optimistic initial results but poor real-world applicability. By capping the tree's depth, we aimed to create a model that generalized better to new data, thereby ensuring more robust and reliable predictions. For our model, a maximum depth of 15 was set for the tree. The performance metrics of the trained model are presented in Table 5.6.

	Accuracy	**AUC**	**Recall**	**Prec.**	**F1**	**Kappa**	**MCC**
0	0.8284	0.9006	0.7961	0.7190	0.7556	0.6240	0.6259
1	0.8192	0.8862	0.7618	0.7143	0.7373	0.5996	0.6003
2	0.8406	0.8992	0.7554	0.7636	0.7594	0.6403	0.6403
3	0.8220	0.8996	0.7940	0.7075	0.7482	0.6113	0.6137
4	0.8463	0.9125	0.8069	0.7505	0.7777	0.6605	0.6615
5	0.8320	0.8935	0.7666	0.7397	0.7529	0.6257	0.6259
6	0.8406	0.9140	0.8223	0.7328	0.7750	0.6522	0.6548
7	0.8197	0.8958	0.7318	0.7286	0.7302	0.5949	0.5949
8	0.8255	0.8943	0.8047	0.7102	0.7545	0.6199	0.6228
9	0.8526	0.9219	0.7790	0.7790	0.7790	0.6685	0.6685
Mean	0.8327	0.9018	0.7818	0.7345	0.7570	0.6297	0.6309
SD	0.0112	0.0104	0.0264	0.0226	0.0156	0.0238	0.0238

Table 5.6: Initial DT Results

Despite these preventive measures, an overfitting from the model was observed, as observed by the learning curve presented in Figure 5.2. This overfitting can be observed in the slightly descending training score to the cross-validation score.

Figure 5.2: Learning curve of the DT model: in this case, the stability of the training score compared to the growing trend of the cross-validation score suggests the overfitting of the model.

To address the issues related to overfitting, we performed hyperparameter tuning. This step aims to fine-tune the model by adjusting various hyperparameters to find the optimal configuration that minimizes overfitting while improving the model's predictive accuracy.

Tuned DT Model

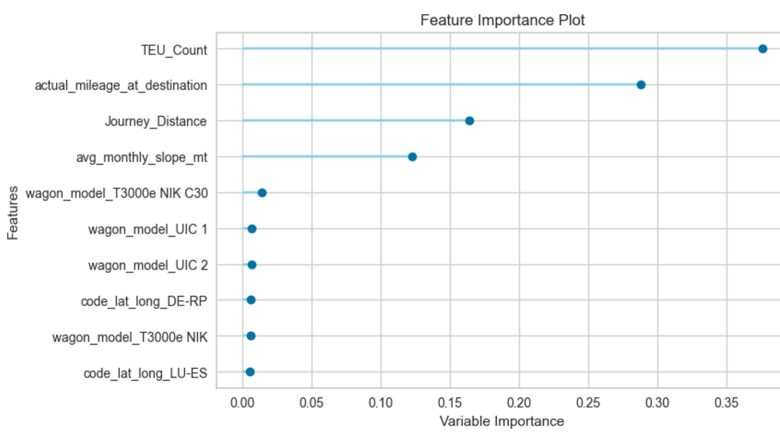
To tune the DT model and obtain Tuned Decision Tree (TDT), we opted for the optimization of the *F1* score, an evaluation metric that measures a model's accuracy that combines the Precision and Recall scores of a model. This was done to ensure that when maintenance was predicted, it was indeed necessary, thereby minimizing the potential for unnecessary maintenance activities. The metrics are presented in 5.7.

	Accuracy	**AUC**	**Recall**	**Prec.**	**F1**	**Kappa**	**MCC**
0	0.8106	0.8919	0.7639	0.6967	0.7288	0.5837	0.5852
1	0.8034	0.8901	0.7318	0.6945	0.7126	0.5634	0.5639
2	0.8299	0.9045	0.7060	0.7651	0.7344	0.6095	0.6106
3	0.7956	0.8876	0.7446	0.6751	0.7082	0.5514	0.5530
4	0.8256	0.9026	0.7661	0.7256	0.7453	0.6128	0.6134
5	0.8127	0.8905	0.7216	0.7186	0.7201	0.5794	0.5794
6	0.8106	0.8952	0.7516	0.7020	0.7260	0.5815	0.5823
7	0.8033	0.8972	0.6888	0.7118	0.7001	0.5538	0.5540
8	0.7961	0.8786	0.7532	0.6737	0.7112	0.5545	0.5565
9	0.8290	0.9134	0.7339	0.7484	0.7411	0.6135	0.6135
Mean	0.8117	0.8952	0.7362	0.7111	0.7228	0.5804	0.5812
SD	0.0121	0.0093	0.0238	0.0280	0.0141	0.0235	0.0233

Table 5.7: Tuned DT Results

The decrease in the TDT model's performance is paired with less overfitting, as shown in Figure 5.3a. Therefore, we chose the TDT model to further investigate the role of input features on the predicted outcomes. This allowed us to study how different attributes interact with each other and how they contribute to the possible disruptions of rolling stocks.

(a) Learning curve of the TDT model: in this case, the stability of the training score compared to the steeper growing trend of the cross-validation score suggests less overfitting of the model compared to the DT model.



(b) Feature importance plot on incidents causing unplanned maintenance on rolling stock.

Figure 5.3: TDT learning curve and Feature Importance Plot.

Figure 5.3b shows the feature importance plot for the prediction based on the TDT. For this dataset, the TEU count (monthly) emerged as one of the most influential features, followed by the

Actual Mileage, Journey_Destination, and AVG Slope (monthly). The significance of these features is logical, given that both Slope and Journey Distance are highly dependent on the destination and the route taken, and when paired with a high AVG TEU count these might trigger one of the mentioned accidents.

5.2.3 Performance Assessment and Probability Threshold in the TDT model

To assess the trustworthiness of the predictions generated by our TDT model, we use the concepts of True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN), which compose the Confusion Matrix, and Precision, Recall (True Positive Rate, TPR), Specificity (True Negative Rate, TNR), False Positive Rate (FPR), and False Negative Rate (FNR). The explanation and formulas of them are presented in Table 5.8.

Name	**Explanation**	**Formula**
True Positive (TP)	Correctly predicted positive	-
False Positive (FP)	Incorrectly predicted positive	-
True Negative (TN)	Correctly predicted negative	-
False Negative (FN)	Incorrectly predicted negative	-
Precision	Proportion of true positives in predicted positives	$\frac{TP}{TP+FP}$
Recall (TPR)	Proportion of true positives out of all actual positives	$\frac{TP}{TP+FN}$
Specificity (TNR)	Proportion of true negatives out of all actual negatives	$\frac{TN}{TN+FP}$
False Positive Rate (FPR)	Proportion of false positives out of all actual negatives	$\frac{FP}{TN+FP}$
False Negative Rate (FNR)	Proportion of false negatives out of all actual positives	$\frac{FN}{TP+FN}$

Table 5.8: Confusion Matrix Metrics and Formulas.

These values are all dependent on a key factor, which is also related to risk assessment, namely the probability threshold. The probability threshold in classification decisions is key to understanding the model behavior and providing good predictions. It is a predetermined value, usually between 0 and 1, that determines the cutoff at which a data point is classified into one of the two categories based on the predicted probability. In risk assessment, it can be seen as the level of risk exposure that the model is willing to accept when making predictions. A lower threshold implies a more conservative approach, where the model identifies more instances for preventive maintenance, although this comes at the expense of increased false positives, leading to potentially unnecessary maintenance actions, and more shunting operations, increasing the probability of delays and cancellation. Conversely, a higher threshold expresses a more tolerant stance, likely resulting in fewer false positives but at the risk of overlooking some essential preventive maintenance, as captured by the false negative rate, and therefore increasing the exposure to the risk event. We provide an example for further clarification. Let's consider a scenario where our TDT model is used for rolling stock disruption prevention. The model predicts whether a particular rolling stock is likely to fail soon.

- **Scenario**: The model analyzes data from sensors on the rolling stock and outputs a probability of failure shortly.
- **Probability Threshold Setting**: Let's say we set the probability threshold at 0.7. This means that if the model predicts a failure probability of 70% or higher, it flags this as a potential disruption.
- **Outcome Interpretation**:
  - If the model predicts a failure probability of 75% for a disruption, this exceeds our threshold of 70%. Consequently, maintenance actions are triggered to prevent potential disruption.
  - Conversely, if the model predicts a 60% probability of failure, this does not exceed the threshold, and no immediate action is taken, though monitoring may continue.

The key here is the balance between safety and efficiency:

- A High Threshold (e.g., 90%) might result in fewer false alarms but could miss some potential failures, risking disruptions.
- A Low Threshold (e.g., 30%) increases sensitivity to potential problems, possibly leading to more frequent maintenance actions but could also result in unnecessary inspections and repairs, reducing operational efficiency.

The choice of threshold becomes a trade-off between different evaluation metrics and ultimately depends on the specific operational context and the associated costs of false positives (useless maintenance) and false negatives (missed preventive maintenance). For instance, in a scenario where the cost of a false negative is significantly high—indicating a severe consequence for missing a necessary maintenance action—a lower threshold may be more acceptable despite the higher false positive rate.

To empirically determine the optimal threshold for our problem, we examined a range of threshold values, incrementing by 0.05, and observed the corresponding impact on the key performance metrics. Results are presented in Table 5.9.

TP	**FP**	**TN**	**FN**	**Precision**	**Recall**	**Specificity**	**FPR**	**FNR**	**Threshold**
5801	25356	19763	27	0.19	1	0.44	0.56	0	0.05
5795	21447	23672	33	0.21	0.99	0.52	0.48	0.01	0.1
5783	18606	26513	45	0.24	0.99	0.59	0.41	0.01	0.15
5759	16123	28996	69	0.26	0.99	0.64	0.36	0.01	0.2
5732	13631	31488	96	0.3	0.98	0.7	0.3	0.02	0.25
5639	11011	34108	189	0.34	0.97	0.76	0.24	0.03	0.3
5499	9221	35898	329	0.37	0.94	0.8	0.2	0.06	0.35
5303	7644	37475	525	0.41	0.91	0.83	0.17	0.09	0.4
5113	6404	38715	715	0.44	0.88	0.86	0.14	0.12	0.45
4572	3759	41360	1256	0.55	0.78	0.92	0.08	0.22	0.5
3928	2181	42938	1900	0.64	0.67	0.95	0.05	0.33	0.55
3496	1441	43678	2332	0.71	0.6	0.97	0.03	0.4	0.6
3082	880	44239	2746	0.78	0.53	0.98	0.02	0.47	0.65
2686	575	44544	3142	0.82	0.46	0.99	0.01	0.54	0.7
2222	340	44779	3606	0.87	0.38	0.99	0.01	0.62	0.75
1885	198	44921	3943	0.9	0.32	1	0	0.68	0.8
1447	109	45010	4381	0.93	0.25	1	0	0.75	0.85
904	42	45077	4924	0.96	0.16	1	0	0.84	0.9

Table 5.9: Performance metrics at varying probability thresholds.

This analysis, presented in Table 5.9, illustrates the trade-offs in selecting a particular threshold and provides a data-driven basis for choosing a threshold value that aligns with our operational priorities and risk tolerance.

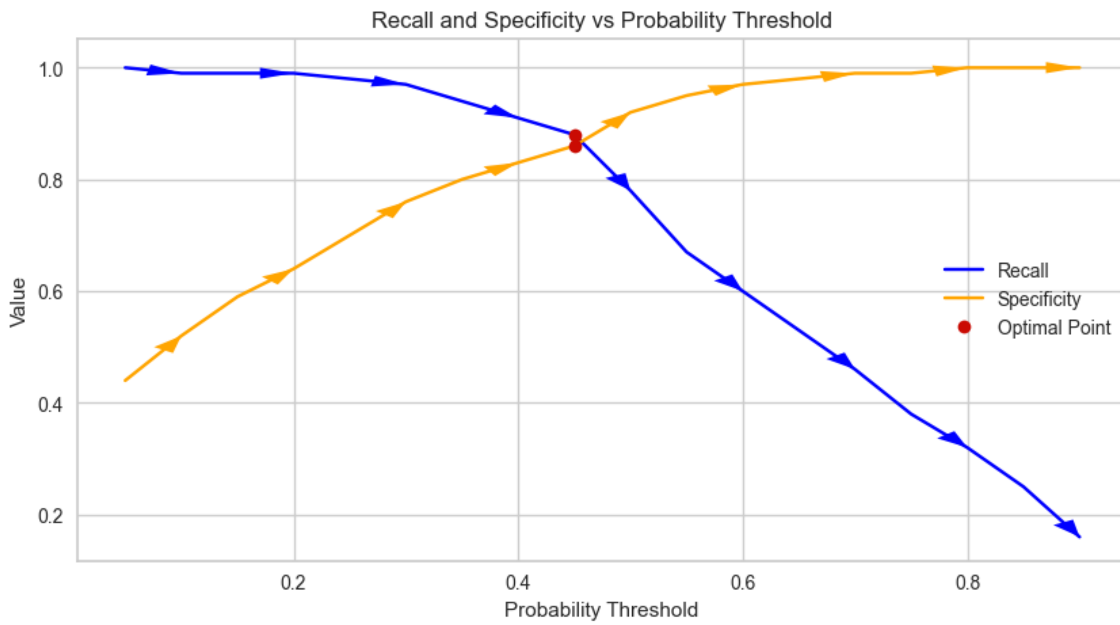


Figure 5.4: Recall (TPR) vs Specificity (TNR) through probability thresholds.

Figure 5.4 shows the trend of Recall and Specificity and TNR through probability thresholds. To balance these two, an optimal point has been identified around the 45% mark.

5.3 MILP Modelling with Condition-Based Maintenance Integration

In this section, we present an improved deterministic optimization model informed by a probabilistic forecast. The new model addresses the limitations identified in earlier MILP models as presented in Section 4.4 by including a more sophisticated representation of the operational dynamics that happen in the shunting yard, together with the risk assessment on the ML's prediction. The key improvements can be summarized as follows:

- **Spatial Granularity:** we introduced the concepts of tracks and positions within the shunting yard, offering a more granular representation of the spatial layout. Specifically, the set $\mathcal{S}$ has been replaced with two new sets TRK and $PTRK$, representing the shunting yard track number and the position relative to the track number respectively.
- **Distribution Revision:** the assumption of a normal distribution that defined the shunt-in time (Section 4.3.2) has been replaced with a fixed shunting time calculation based on the different tracks and positions of the wagons.
- **Wagon Type Expansion:** the model now can deal with all the possible wagon types, aligning with the actual heterogeneity of the fleet observed in practice.
- **Operational Constraints:** the model introduces more specific delay and cancellation constraints, informing the model on the risk of over-shunting and the costs associated with it.
- **Objective Transition:** the objective function shifted from minimizing shunting and delays to a maximization of the expected revenue of the train.
- **Machine Learning Integration:** we integrated ML analytics to risk assess it against its prediction failure, using the MILP model as the final decision maker regarding the preemptive maintenance.

We introduce the conceptualized risk analysis framework and explain the metrics developed to quantify the ML model prediction's vulnerability in Section 5.3.1, where propose a methodology to compute the risk profile for each wagon within the fleet. Following this, we propose an integration of ML metrics and the prediction, and how the latter is mildly inputted within this MILP model using the ML metrics as vulnerability/model trustworthiness in Section 5.3.1. Lastly, we present the complete MILP model in Section 5.3.2.

5.3.1 Risk Assessment

Risk assessment is defined as the systematic process of identifying hazards and evaluating any associated risks within a specific frame, then implementing reasonable control measures to remove or reduce them. The Risk associated with an event is classically defined by the formula:

$$Risk = probability\,of\,occurrence * impact\,of\,event \tag{5.1}$$

Where:

- The *probability of occurrence* is defined as the likelihood of the analyzed event materializing.
- The *impact of event* is defined by the impact (economical, social, health-related) that the event materialization has.

In the FAIR Risk Taxonomy ([32]), the risk analysis includes the identification of the different assets at risk for the targeted event and the analysis of the likelihood of the event to assess the frequency of potential threats. This topic is related to our research as our goal is to implement, within our MILP model, information regarding the condition, and therefore potential disruption, of rolling stocks using the predictions that come from our ML model. The *asset* that we want to preserve from a disruptive event, for our study, is the rolling stock, with the *event* that we're trying to prevent being an unexpected disruption. To assess a potential disruption within our digital environment, we use our TDT model as presented in Section 5.2.2. As this model is the source of our information regarding the potential disruption, this is the *asset* at risk that we are interested in assessing against its failure.

Therefore, the likelihood of our event is defined as the probability that the ML model could provide a wrong prediction, potentially leading to an event that affects both rolling stocks and shunting operations. Specifically, to each prediction, we can always associate its relative model performance, False Positive Ratio, True Positive Ratio, True Negative Ratio, and False Negative Ratio. These are rates that define how trustworthy is the model on its prediction versus the actual event happening and can be computed based on the validation data.

In the FAIR assessment, the impact of the event is defined as Loss Magnitude (LM). The LM is formalized as the loss related to a specific event happening. The losses analyzed in this study are relative to the potential events that are triggered by the failure of the ML model:

- Loss in productivity (PL) is defined as the loss of availability due to disruptions occurring. For this study, the PL has been defined as the potential loss in availability of the fleet due to unplanned disruption, or delays and cancellations based on the gravity of an event.

- Replacement loss (RL) is defined as the cost of replacing the disrupted rolling stock. For this study, is assumed as the repair cost of a specific wagon, which depends on the gravity of the disruption. We computed the repair costs from the PGV Agreement, an agreement on the use of freight wagons in international traffic from the OSJD (Organisation for Cooperation between Railways) Committee. We used this data as they provided a good heterogeneity in terms of repair. To achieve this, we split the distribution of prices according to 4 quantiles, performing a Monte-Carlo simulation to obtain the expected costs for each gravity.

- Monetary loss (MnL) is defined as fees associated with the disruptive event. For our study, this cost was provided to us by CFL multimodal.

Risk of Pointless Disruption

To model the costs associated with the False Positive prediction, a preemptive maintenance operation that is not required, we define the Risk of Pointless disruption R_{point} for each wagon w. This risk expresses the PL of a wagon being out for maintenance which was not needed, and the RL as the cost of removing it from the train due to the additional operation to be performed. This risk is independent of the gravity of the accident.

$$R^{\text{point}} = \overbrace{\bar{t}_{out}^{\text{prev}} * f^{\text{dest,yr}} * R^{\text{dest}} * Rev_{rail}}^{\text{PL}} + \overbrace{C_{\text{single}}^{\text{shunt}}}^{\text{RL}} \tag{5.2}$$

where:

- $\bar{t}_{out}^{prev}$ is the expected time that a wagon will be out to perform preemptive maintenance.

- $f^{dest,yr}$ is the frequency yearly of a wagon for a specific destination.

- R^{dest} is the distance to be covered for a specific destination (r_T in the model presented in Section 4.4).
- Rev_{rail} is the kilometric revenue towards a specific destination.
- $C^{\text{shunt}}_{\text{single}}$ is the cost of a single shunt.

Risk of Disruption

The Risk of Disruption, R^{disr}_{grav}, defines instead the risk associated with encountering a disruptive event and is associated with the False Negative prediction. We formalize this risk as:

$$R^{\text{disr}}_{grav} = \overbrace{(m_{max_i} - m_i) * rev^{\text{dest}}}^{\text{PL}} + \overbrace{C^{\text{corr,maint}}_{grav} + C^{\text{shunt}}_{\text{single}}}^{\text{RL}} + \overbrace{P(grav == 4) * C^{\text{canc}}}^{\text{MnL}} \tag{5.3}$$

where:

- $C^{\text{shunt}}_{\text{single}}$ is the cost of a single shunt.
- $C^{\text{corr,maint}}_{grav}$ is the cost for corrective maintenance based on the gravity level.
- $m_{max_i} - m_i$ are the residual km to be performed before the scheduled maintenance.
- rev^{dest} is the kilometric revenue towards a specific destination.
- $P(grav == 4) * C^{\text{canc}}$ express that if the gravity is of level 4, we also include the cancellation costs.

In our formulation, we assume different types of disruptive events, represented by *grav* and explained in Table 5.2, the gravity of the event. The *PL* is defined as the economic loss related to the wagon losing its potential mileage before the scheduled maintenance. The *RL* is represented by the cost of performing corrective maintenance, which depends on *grav* as the cost of repair increases with the intensity of the disruption, and the cost of removing the wagon from the system. Finally, the *MnL* is specifically for our study, as the highest *grav* of disruptive event forces us into cancellation.

Risk Assessment Conceptual Framework

A conceptual framework is presented for clarity in Figure 5.5.

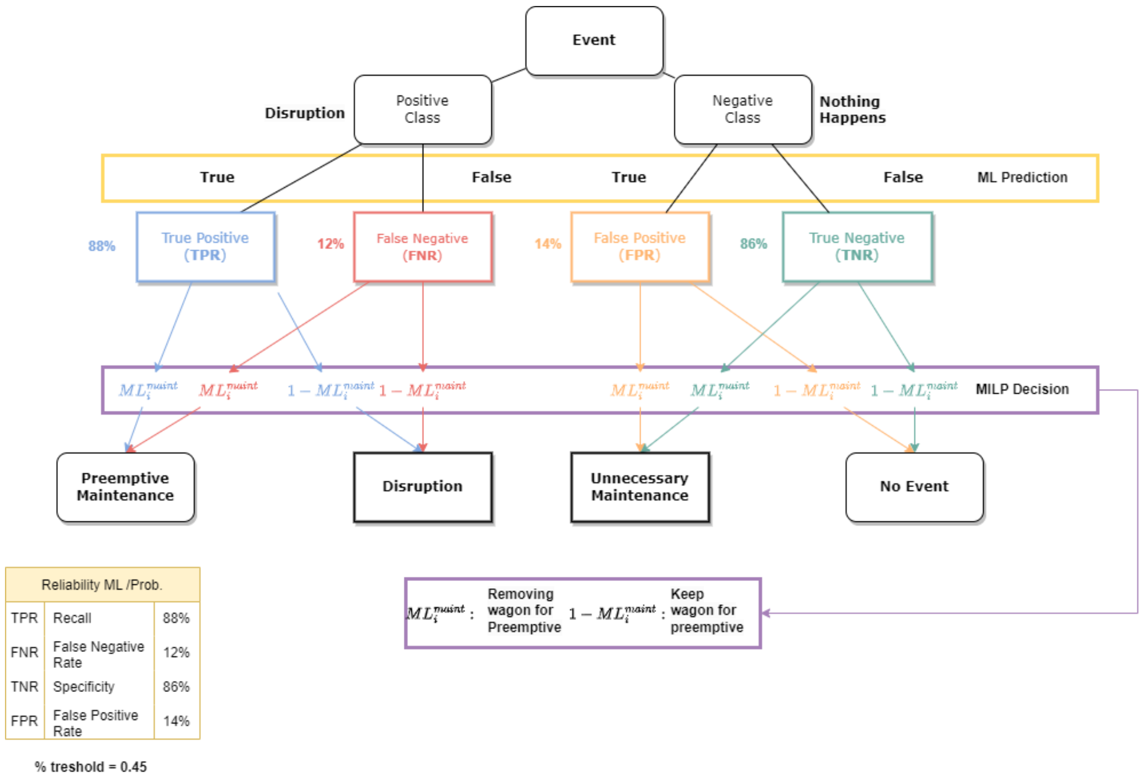


Figure 5.5: Conceptual framework of the risk assessment.

For the True Positive and True Negative, we associate the cost of preemptive maintenance $C^{\text{prev,maint}}$ and No Cost respectively. Figure 5.5 shows the impact of the assessment on the prediction of the shunting model, as well as the proposed mild integration of the ML inside the MILP. The rolling stock arrives with an unknown status, defined in our case as the *event*. The ML model is then run, providing its prediction regarding the status of the wagon. This prediction is then not executed right away, but instead placed inside the MILP, together with the performance metric of the ML model, to make the former decide whether to agree with the ML model based on the associated risks, operations to perform, delays, and cancellations.

The strength of this framework lies in the capability of the MILP model to make more informed decisions based on how trustworthy is the ML model, as well as being able to direct the MILP model by choosing a probability threshold to the impact of each risk. Analyzing the example provided in Figure 5.5, we can see that the proposed model presents good metrics of Recall and Specificity. Within the framework, as for these metrics, the MILP model will tend to trust more the ML prediction. Let's suppose a rolling stock arrives in the station, with our ML model providing us with a *False* prediction. This means that the wagon's prediction can either be a False Negative or a True Negative prediction. Based on the performance of the model, as well as the operations that have to be conducted on the other rolling stocks, the MILP will decide whether to shunt-out the rolling stock or not for preemptive maintenance weighting in this case the Risk of Disruption only by 12%, rather than the expected probability of disruption of the wagon.

5.3.2 Operational Research Model

In this section, we will explain the improved MILP model developed for the Shunt-In/Shunt-Out problem for the condition-based maintenance approach. The following formulation is presented also without comments in Appendix D.

Nomenclature

Sets	
Name	**Description**
$gravity$	Set of possible disruption level $\{g \mid g \in \mathbb{N}_0, 1 \leq g \leq gr\}$
$\mathcal{T}$	Set of position in the inbound train $\{i \mid i \in \mathbb{N}_0, 0 \leq i \leq rail\}$
TRK	Set of track within the shunting yard $\{j \mid j \in \mathbb{N}_0, 0 \leq j \leq tr\}$
$PTRK$	Set of position in the shunting yard track $\{k \mid k \in \mathbb{N}_0, 0 \leq x \leq pos\}$
TYP	Set of types of wagons $\{l \mid l \in \mathbb{N}_0, 1 \leq l \leq typ\}$
Parameters	
Name	**Description**
$rail$	Number of wagons in the train, $rail \in \mathbb{N}_0$
tr	Number of tracks in the SY $tr \in \mathbb{N}_0$
pos	Number of positions for tracks in the SY $pos \in \mathbb{N}_0$

typ	Number of types available in the fleet, $typ \in \mathbb{N}_0$
gr	Possible number of gravity of disruption, $gr \in \mathbb{N}_0$
$rev_{j,k}$	Potential revenue of each wagon in the SY, $rev_{j,k} \in \mathbb{R}^{tr*pos}$
$R^{\mathrm{disr}}_{j,k,g}$	Risk of disruption for each wagon in the SY by gravity, $R^{\mathrm{disr}}_{j,k,g} \in \mathbb{R}^{tr*pos*gr}$
$adj_{j,k}$	Adjacent wagons in the SY, $adj_{j,k} \in \mathbb{N}_0^{tr*pos}$
$suit_{j,k}$	Suitable wagons in the SY, $suit_{j,k} \in \mathbb{N}_0^{tr*pos}$.
$typ_{j,k,l}$	Wagon in the shunting yard per type, $typ_{j,k,l}\mathbb{N}_0^{tr*pos*typ}$
$pred^{\mathrm{ML}}_{j,k}$	prediction of the ML model for the wagon in the SY. $prob^{\mathrm{ML}}_{j,k} \in 0, 1^{tr*pos}$
Train parameters	
Name	**Description**
$difference_{i,l}$	Difference in the wagons of the train by type, $difference_{i,l} \in \mathbb{N}_0^{rail*typ}$
$WI_{i,l}$	Type composition for the inbound train, $WI_{i,l} \in \mathbb{N}_0^{rail*typ}$
$WO_{i,l}$	Type composition for the outbound train, $WO_{i,l} \in \mathbb{N}_0^{rail*typ}$
t_{op}	Expected operational time of the train, defined as the time before cancellation happens, $t_{\mathrm{op}} \in \mathbb{R}$
$R^{\mathrm{disr}}_{i,g}$	Risk of disruption for each wagon in the train, $R^{\mathrm{disr}}_{i,g} \in \mathbb{R}^{rail*gr}$
R^{point}	Risk of pointless disruption, $R^{\mathrm{point}} \in \mathbb{R}$
Rev_{rail}	Potential revenue of the train as a sum of the single wagon's revenue, $Rev_{rail} \in \mathbb{R}^{rail}$
$maint_i$	Wagons to be removed due to maintenance constraint $maint_i \in 0, 1^{rail}$
$pred^{\mathrm{ML}}_i$	prediction of the ML model for the wagon in the inbound train. $pred^{\mathrm{ML}}_i \in 0, 1^{rail}$
ML & Risk Parameters	
Name	**Description**
C_{shunt}	cost of a single shunt, $C_{\mathrm{shunt}} \in \mathbb{R}$
$C^{\mathrm{corr,maint}}_{gr}$	cost of corrective maintenance per gravity, $C^{\mathrm{corr,maint}}_{gr} \in \mathbb{R}^{gr}$

$C^{\text{prev,maint}}$	average cost of preemptive maintenance, $C^{\text{prev,maint}} \in \mathbb{R}$
t_{shunt}	time to perform one shunting operation, $t_{\text{shunt}} \in \mathbb{R}$
C^{canc}	cost of cancelling one train, $C^{\text{canc}} \in \mathbb{R}$
TPR	True Positive Ratio from the ML model, $TPR \in \mathbb{R}$
TNR	True Negative Ratio from the ML model, $TNR \in \mathbb{R}$
FPR	False Positive Ratio from the ML model, $FPR \in \mathbb{R}$
FNR	False Negative Ratio from the ML model, $TPR \in \mathbb{R}$
P^{gr}	probability of gravity disruption, $P^{gr} \in \mathbb{R}^{gr}$
Decision Variables	
Name	**Description**
t_{dep}	expected departure time, $t_{\text{dep}} \in \mathbb{N}_0$
$Canc^{\text{trig}}$	cancellation variable for costs, $Canc^{\text{trigg}} \in [0, 1]$
$Risk$	risk term overall between SI & SO, $Risk \in \mathbb{R}$
C_{maint}	overall maintenance cost, $C_{\text{maint}} \in \mathbb{R}$
Dem_i	wagons to be SO for demand reason, $Dem_i \in [0, 1]^{rail}$
w_i^{SO}	wagons to be shunted out from the train, $w_i^{\text{SO}} \in [0, 1]^{rail}$
$w_i^{\text{IN,staying}}$	wagons that are staying after the SO, $w_i^{\text{IN,staying}} \in [0, 1]^{rail}$
$shunt^{\text{SO}}$	number of shunting operations performed for SO, $shunt^{SO} \in \mathbb{N}_0$
adj^{SO}	number of adjacent wagons moved for the SO, $adj^{SO} \in \mathbb{N}_0$
$adj_{i,i}$	check if the wagons to SO are adjacent, $adj_{i,i} \in [0, 1]^{rail}$
R_i^{SO}	risk of the train after the SO, $R_i^{\text{SO}} \in \mathbb{R}^{rail}$
ML_i^{maint}	wagons to be SO due to the preemptive maint, $ML_i^{\text{maint}} \in [0, 1]^{rail}$
$w_{i,j,k}^{\text{SI}}$	wagons to be shunted in rail from the SY at (j,k), $w_{i,j,k}^{\text{SI}} \in [0, 1]^{rail*tr*pos}$
adj^{SI}	number of adjacent wagons moved for the SI, $adj^{\text{SI}} \in \mathbb{N}_0$
$adj_{j,k}^{\text{SI}}$	flag for adjacent wagon in the SY, $adj_{j,k}^{\text{SI}} \in [0, 1]^{tr*pos}$
$shunt^{\text{SI}}$	number of shunting operations performed for SI, $shunt^{\text{SI}} \in \mathbb{N}_0$
$shunts$	overall number of shunting operations performed, $shunts \in \mathbb{N}_0$
w^{SI}	number of wagons to SI, $w^{\text{SI}} \in \mathbb{N}_0$

w^{moved}	overall number of wagons moved, $w^{\text{moved}} \in \mathbb{N}_0$
$W^{\text{out}}_{i,l}$	train overall outbound type composition, $W^{\text{out}} \in [0,1]^{rail*typ}$
$diff^{\text{SISO}}$	difference between SI and SO, $diff^{\text{SISO}} \in \mathbb{N}_0$
$R^{\text{SI}}_{j,k}$	risk term of the wagon to SI, $R^{\text{SI}}_{j,k} \in \mathbb{R}^{tr*pos}$
R^{SI}_{i}	risk term of the wagon to SI (connect with the train), $R^{\text{SI}}_{i} \in \mathbb{R}^{tr*pos}$
R^{staying}_{i}	overall risk term of the wagon staying in the train between SI and SO, $R^{\text{staying}}_{i} \in \mathbb{R}^{rail}$
Rev_{rail}	overall train revenue, $Rev_{rail} \in \mathbb{R}^{rail}$

Table 5.10: Nomenclature of the Deterministic Risk model.

Two notes on nomenclature:

- $C^{\text{prev,maint}}$ is computed by taking the average cost of preemptive maintenance and adding to it the loss due to unavailability for that period, namely $L^{prev,maint} = t^{prev}_{out} * f_{daily} * R * Rev_{rail}$, where t^{prev}_{out} is the expected time out for preemptive maintenance in days, f_{daily} is the daily frequency for a destination. Therefore, $C^{\text{prev,maint}} = \mathbb{E}[C^{\text{prev,maint}}] + L^{prev,maint}$.
- $suit_{j,k}$ is defined by precomputing if the wagons in the SY can perform the next trip without surpassing the maximum mileage threshold.

Objective Function and Constraints

Eq. 5.4 is the objective function. We want to maximize the revenue of the inbound train, defined as the sum of the potential revenue of each wagon throughout its services until the maintenance deadline. The objective function considers the risk associated with the status of each wagon, cancellation, shunting to perform, and maintenance operations associated with the train.

$$\max \quad Rev_{rail} - (C^{\text{maint}} + shunts \cdot C_{\text{shunt}} + Risk + Canc^{trig} \cdot C^{\text{canc}}) \tag{5.4}$$

The set of constraints 5.5 models the behavior of the model in terms of risk and revenue.

$$\sum_{i \in \mathcal{T}} (1 - w_i^{\text{SO}}) \cdot rev_i + \sum_{j \in TRK} \sum_{k \in PTRK} w_{i,j,k}^{\text{SI}} * rev_{j,k} = Rev_{rail} \tag{5.5a}$$

$$\sum_{i \in \mathcal{T}} R_i^{\text{SI}} + R_i^{\text{SO}} = Risk \tag{5.5b}$$

$$\begin{aligned}
& pred_i^{\text{ML}} \cdot \{TPR \cdot [\\
& (\sum_{g \in \text{gravity}} P_g \cdot R_{i,g}^{\text{disr}}) \cdot (1 - ML_i^{\text{maint}}) + (C_{\text{single}}^{\text{shunt}} + C^{\text{prev,maint}}) * ML_i^{\text{maint}} \\
&] + FPR * [(C_{\text{single}}^{\text{shunt}} + C^{\text{prev,maint}} + R^{\text{point}}) * ML_i^{\text{maint}}]\} \\
& +(1 - pred_i^{\text{ML}}) \cdot \\
& \{TNR \cdot (C_{\text{single}}^{\text{shunt}} + R^{\text{point}} + C^{\text{prev,maint}}) \cdot ML_i^{\text{maint}} \\
& +FNR \cdot [(\sum_{g \in \text{gravity}} P_g * R_{i,g}^{\text{disr}}) \cdot (1 - ML_i^{\text{maint}}) \\
& +(C_{\text{single}}^{\text{shunt}} + C^{\text{prev,maint}}) \cdot ML_i^{\text{maint}}]\} \\
& = R_i^{\text{SO}} \quad \forall i \in \mathcal{T}
\end{aligned} \tag{5.5c}$$

$$\begin{aligned}
& \sum_{j \in TRK} \sum_{k \in PTRK} \{[pred_{j,k}^{\text{ML}} \cdot \\
& TPR \cdot (\sum_{g \in \text{gravity}} P_g \cdot R_{i,j,k}^{\text{disr}}) \cdot w_{i,j,k}^{\text{SI}}] + \\
& +(1 - pred_{j,k}^{\text{ML}}) \cdot FNR \cdot [(\sum_{g \in \text{gravity}} P_g \cdot R_{i,j,k}^{\text{disr}}) \cdot w_{i,j,k}^{\text{SI}}]\} \\
& = R_i^{\text{SI}} \qquad \forall i \in \mathcal{T}
\end{aligned} \tag{5.5d}$$

Equation 5.5a models the revenue of the train defined by the revenue of the wagons staying in the train after the operations and the one chosen for the shunting. Equation 5.5b defines the overall risk of the train and is composed of two factors: R_i^{SO}, risk related to the shunt-out operations, and R_i^{SI}, risk related to the shunt-in operations.

R_i^{SO} is defined by Eq. 5.5c and models the decision of which wagons should be removed or not from the inbound train using the ML model as an advisor. The first part of the equation reads:

- If $pred_{j,k}^{\mathrm{ML}} = 1$, ML model predicting disruption, the cost of not removing the wagon $(1 - ML_i^{\mathrm{maint}})$ from the inbound train is $R_{i,g}^{\mathrm{disr}}$, the risk of disruption. Trusting the model, therefore removing the rolling stock (ML_i^{maint}), has an associated cost of $(C_{\mathrm{single}}^{\mathrm{shunt}} + C^{\mathrm{prev,maint}})$, the shunting cost and the cost of preemptive maintenance. These costs are then weighted based on the capability of the model to recognize positive class, as defined by the TPR. To this, we add the risk of performing unnecessary maintenance if this prediction is instead a False Positive. This is represented by $(C_{\mathrm{single}}^{\mathrm{shunt}} + C^{\mathrm{prev,maint}} + R^{\mathrm{point}})$, weighted on the capability of the model to find False Positives, defined as the FPR.

- If $pred_{j,k}^{\mathrm{ML}} = 0$, ML model predicting no disruption, we investigate the risk of looking at the negative class. The risk of removing a wagon for unnecessary maintenance is represented by $(C_{\mathrm{single}}^{\mathrm{shunt}} + R^{\mathrm{point}} + C^{\mathrm{prev,maint}})$, weighted on the capability of the model to find True Negative, TNR. To this, we add the risk of being a False Negative, with the respective decision and risks as for the TPR case.

In a similar way, we define R_i^{SI} in 5.5d as the risk of the wagon that needs to be shunted-in from the shunting yard inside the outbound train. Given that there is always an associated risk in placing a wagon for the destination, we can in this case only decide which wagon should be shunted in, and not which one has to be removed. This reduces the problem of choosing the wagon at minimum risk of future disruption.

The set of constraints 5.6 defines the time, cost, and constraints related to the number of shunts.

$$\sum_{i \in \mathcal{T}} \sum_{g \in \mathrm{gravity}} (P_g \cdot C_g^{\mathrm{corr,maint}}) \cdot maint_i + C^{\mathrm{prev,maint}} \cdot ML_i^{\mathrm{maint}} = C^{\mathrm{maint}} \tag{5.6a}$$

$$t_{\mathrm{dep}} - t_{\mathrm{op}} \leq M \cdot Canc^{trig} \tag{5.6b}$$

$$shunts \cdot t_{\mathrm{shunt}} = t_{\mathrm{dep}} \tag{5.6c}$$

$$shunts^{\mathrm{SI}} + shunts^{\mathrm{SO}} = shunts \tag{5.6d}$$

Eq. 5.6a defines the overall cost of maintenance C^{maint} to be sustained between the preemptive and corrective maintenance. Eq. 5.6b and 5.6c define the timing constraints for the cancellation and the expected departure time after operations. Eq. 5.6d defines the overall number of shunts to be performed on the train.

The set of equations 5.7 models the shunt-out operations.

$$\sum_{i \in \mathcal{T}} w_i^{\text{SO}} - \sum_{i=1}^{rail-1} adj_{i,i+1} = shunts^{\text{SO}} \tag{5.7a}$$

$$2 \cdot adj_{i,i+1}^{\text{SO}} \leq w_i^{\text{SO}} + w_{i+1}^{\text{SO}} \qquad \forall i \in [\mathcal{T}, \mathcal{T}+1] \tag{5.7b}$$

$$maint_i + ML_i^{\text{maint}} + dem_i \leq 1 \qquad \forall i \in \mathcal{T} \tag{5.7c}$$

$$dem_i \leq (1.05 - prob_i^{\text{ML}}) \qquad \forall i \in \mathcal{T} \tag{5.7d}$$

$$\sum_{i \in \mathcal{T}} dem_i \leq \sum_{l \in TYP} |diff_l| \tag{5.7e}$$

$$maint_i + ML_i^{\text{maint}} + dem_i = w_i^{\text{SO}} \qquad \forall i \in \mathcal{T} \tag{5.7f}$$

Eq. 5.7a defines the overall number of shunts that are accounted for the shunt-out operations. Eq. 5.7b defines the number of wagons that have to be removed from the train by adjacency. Eq. 5.7c defines that we can't perform demand shunts and preemptive shunts on the same wagon. Eq. 5.7d and 5.7e defines the shunt-out conditions for the demand shunts: we cannot remove a wagon for demand reason if the ML model predicts a probability of disruption for the specific wagon less than 5%. For this type of shunt, we bound the number of wagons that can be removed by the sum of the absolute value of the difference of demand types between inbound and outbound composition. Eq. 5.7f defines that a wagon to shunt out is defined by the corrective, preemptive, and demand removal.

The set of equation 5.8 defines the connection between the shunt in and shunt out, as well as the type and demand management.

$$\sum_{l\in TYP} diff_l - \sum_{i\in\mathcal{T}}(w_i^{\mathrm{SO}} - \sum_{j\in TRK}\sum_{k\in PTRK} w_{i,j,k}^{\mathrm{SI}}) = 0 \tag{5.8a}$$

$$\sum_{j\in TRK}\sum_{k\in PTRK} w_i^{\mathrm{SO}} \cdot typ_{i,j,k} + (1-w_i^{\mathrm{SO}})\cdot WI_{i,l} \geq w_{i,l}^{OUT} \qquad \forall i\in\mathcal{T}, \forall l \in TYP \tag{5.8b}$$

$$\sum_{i\in\mathcal{T}} w_{i,l}^{OUT} - WO_{i,l} = 0 \qquad \forall l \in TYP \tag{5.8c}$$

$$\sum_{l\in TYP} w_{i,l}^{OUT} = 1 \qquad \forall i\in\mathcal{T} \tag{5.8d}$$

Eq. 5.8a ensures that if a wagon is requested to be shunted-out, then it will be replaced by a shunt-in wagon considering the difference between the types of input and output. This not only allows for management of the change in the size of the inbound and outbound train but connects the $w_{i,j,k}^{\mathrm{SI}}$ and w_i^{SO} variables. Eq. 5.8b allows the model to fit the demands type between the wagons that stay inside the train after the shunt-out operations and the wagons that will be placed in from the shunt-in operations. Eq. 5.8c and 5.8d model that the demand of the outbound train has to be to filled without replacement.

The set of equations 5.9 defines the shunt-in constraints.

$$\sum_{\hat{i}=1}^{\mathcal{T}-1} w_{\hat{i},j,k}^{\mathrm{SI}} \cdot \sum_{\tilde{i}=\hat{i}+1}^{\mathcal{T}} w_{\tilde{i},j,k}^{\mathrm{SI}} \cdot adj_{j,k} \geq adj_{j,k}^{\mathrm{SI}} \qquad \forall j \in TRK, \forall k \in PTRK \tag{5.9a}$$

$$\sum_{i\in\mathcal{T}}\sum_{j\in TRK}\sum_{k\in PTRK} w_{i,j,k}^{\mathrm{SI}} - \sum_{k=1}^{PTRK-1} adj_{j,k}^{\mathrm{SI}} = shunts^{\mathrm{SI}} \tag{5.9b}$$

$$\sum_{i\in\mathcal{T}} w_{i,j,k}^{\mathrm{SI}} \leq 1 \qquad \forall j \in TRK, \forall k \in PTRK \tag{5.9c}$$

$$\sum_{j\in TRK}\sum_{k\in PTRK} w_{i,j,k}^{\mathrm{SI}} \leq 1 \qquad \forall i\in\mathcal{T} \tag{5.9d}$$

$$w_{i,j,k}^{\mathrm{SI}} \leq suit_{j,k} \qquad \forall i\in\mathcal{T}, \forall j \in TRK, \forall k \in PTRK \tag{5.9e}$$

Eq. 5.9a defines which wagons that we are shunting-in are adjacent in the shunting yard. Eq. 5.9b defines the number of overall shunts to be performed for the shunt-in operations. Eq. 5.9c and 5.9d constraint the shunt-in variable only to take 1 wagon from the SY to place inside the train. Eq. 5.9e force the model to choose only the suitable wagon. This last constraint is both optional and critical: while it might look unnecessary, it reduces the search for the wagon to replace within the shunting yard only among the ones that can perform the trip, saving for big instances some computational time. More importantly, it can allow for the implementation of the different shunt-in policy, by creating a ranking of the rolling stock or not showing to the MILP model inadequate wagons.

5.3.3 Comparison with the Classical Approach

To compare our approach to the traditional approach to ML implementation, we use the same model by treating ML_i^{maint} as an input rather than a decision variable. This is modeled as Eq. 5.10:

$$ML_i^{\text{maint}} = pred_i^{\text{ML,in}} \qquad \forall i \in \mathcal{T} \quad (5.10a)$$

This allows for a direct comparison between the two models in terms of both risk and key performance indicators (KPIs). The modifications include the introduction of a new input, $pred_i^{\text{ML,in}}$, set equal to $pred_i^{\text{ML}}$ if $maint_i = 0$, and 0 otherwise; the removal of Equation 5.7d, to provide more flexibility in demand management due to the forced acceptance of the input.

This modification allows the use of Equations 5.5c and 5.5d as the risk model, enabling the computation of risk associated with ML-driven decisions.

5.4 Case Study and Result

5.4.1 Case Study

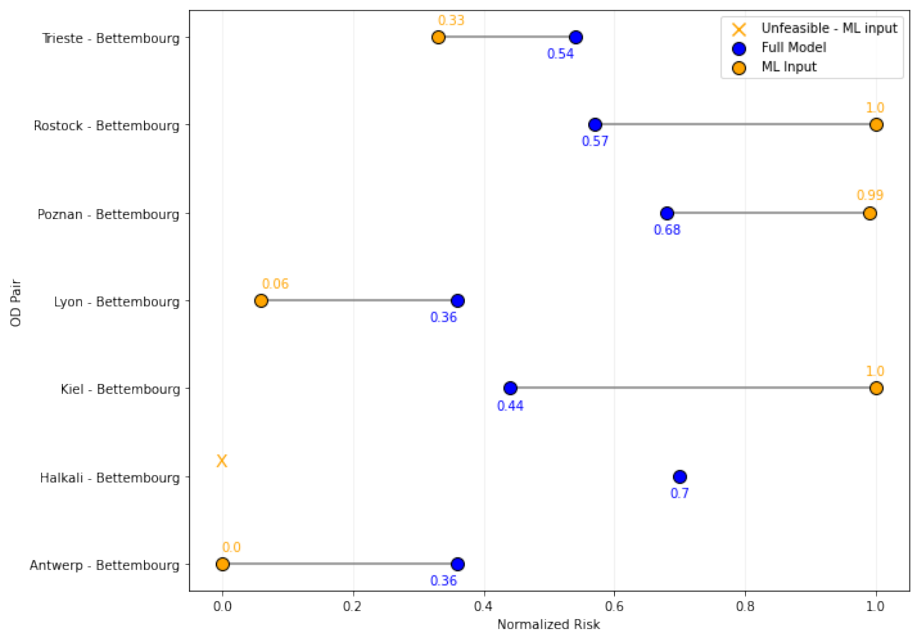
For the case study, we used the TDT model explained in Section 5.2.2. The instance for the study was created using real statistics from CFL Multimodal for the year 2023, covering 7 different destinations. This dataset includes information such as the average number of wagons required per destination, associated revenue, frequency of cancellations, distance required for each journey, and expected travel times. We associated each data point of the year 2023 with the missing data required to perform the ML model prediction. The data regarding the preemptive and corrective maintenance costs and time, expected delay, cancellation fees, revenue per km, as well as shunting costs, and the data for the risk assessment were provided by CFL Multimodal. The inbound train's wagon's actual mileage m_i is defined by sampling daily frequencies from a normal distribution based on the mean and standard deviation of the destination for each year. If the standard deviation or mean frequency for a year is zero or missing, it defaults to a general average value. The total kilometers are computed by multiplying daily frequency samples by the distance covered for that service for each day, then summing these products across all days and years considered. The maximum mileage is instead set at 180000 km, as for CFL information. The shunting yard is populated with a fixed numbers wagon that performs random services across the 7 destinations. For the case study, we run the models for both the full Risk Function and the ML as an input implementing a 1-year-old and 3-year-old fleet, with probability thresholds ranging between 0.05 and 0.95. We made this choice to test the model under both very unlikely and likely scheduled maintenance, as the year of the fleet is reflected on the m_i of the rolling stocks arriving on the inbound train based on the destinations. Cancellation is triggered by this model when the operational time surpasses 180 minutes, and consider 15 minutes per shunting operation.

5.4.2 Result

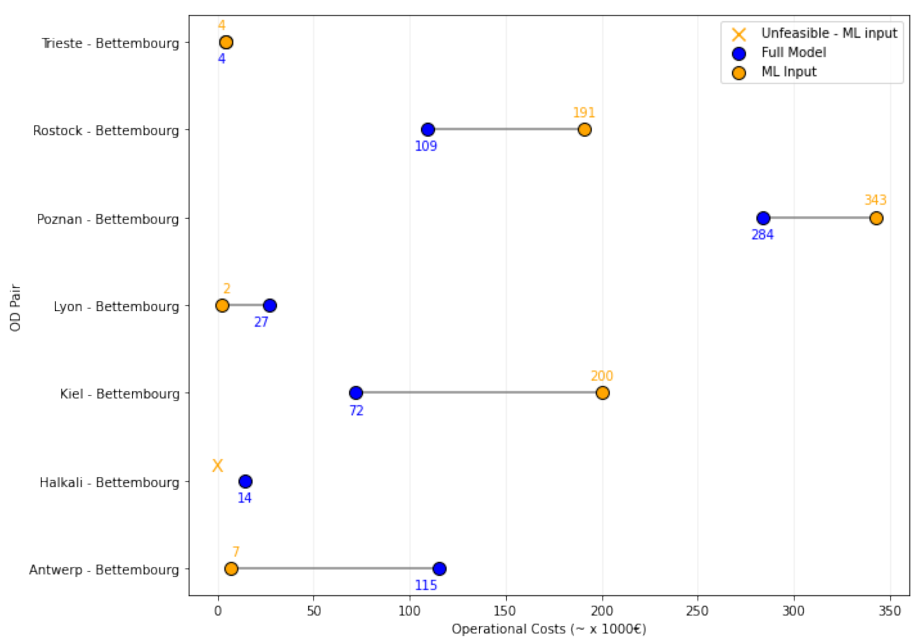
	1 Year fleet			3 Years fleet		
	Full Risk	**ML Input**	**% Diff Risk-ML**	**Full Risk**	**ML Input**	**% Diff Risk-ML**
Max Wagon-Demand In-Out	7.14		-	5.33		-
Predicted Disrupted Wagons	7.6		-	4.55		-
Operating Cost	89815.13	84394.96	-6.03	189439.24	187936.37	-0.79
Overall Shunts	13.34	13.65	2.3	18.33	18.19	-0.8
Total Wagons Moved	20.51	20.61	0.5	30.48	30.79	1
Actual Departure	200.12	204.72	2.3	275	272.8	-0.8
Risk Term	787139.25	887916.07	12.8	795515.9	809940.47	1.81

Table 5.11: Result Model for 1 and 3 years.

Table 5.11 presents the result for the 2 instances, based on the KPIs that relate to the operative/tactical context, as explained in Section 3.2. These values have been computed as the average for each destination and each probability threshold. For the 1-year fleet, the Full Risk model slightly surpasses the ML as an input, as the latter is forced to implement the prediction of the ML model. The operational cost for this analysis is defined as: $C^{\text{maint}} + shunts \cdot C_{\text{shunt}}$, so the overall cost of wagons to send to maintenance for either scheduled or preemptive maintenance and the cost of shunting operation. We can see that the Full Risk has higher operational costs, due to more maintenance and therefore operations in all instances, which also leads to an improvement of 12.8% in the Risk Term compared to the ML Input.



(a)



(b)

Figure 5.6: Normalized Risk (a) and Operational Cost (b) per destination for the 1-year fleet.

Figure 5.6a and 5.6b show an in-depth regarding each single destination's performance for the risk and the operational cost. The risk in Figure 5.6a has been Min-Max normalized for visualization reasons. What can be observed here is that for this instance the Full Risk model presents a more stable trend compared to the ML one for both the KPIs. Moreover, what can be observed is that for this instance ML input returned unfeasible for the Halkali - Bettembourg destination pair, while the Full Risk model managed to solve it at the optimum. The unfeasibility is due to the high amount of wagons to remove from the ML input.

Figure 5.7: Normalized Risk per destination and probability threshold for the Risk Model.

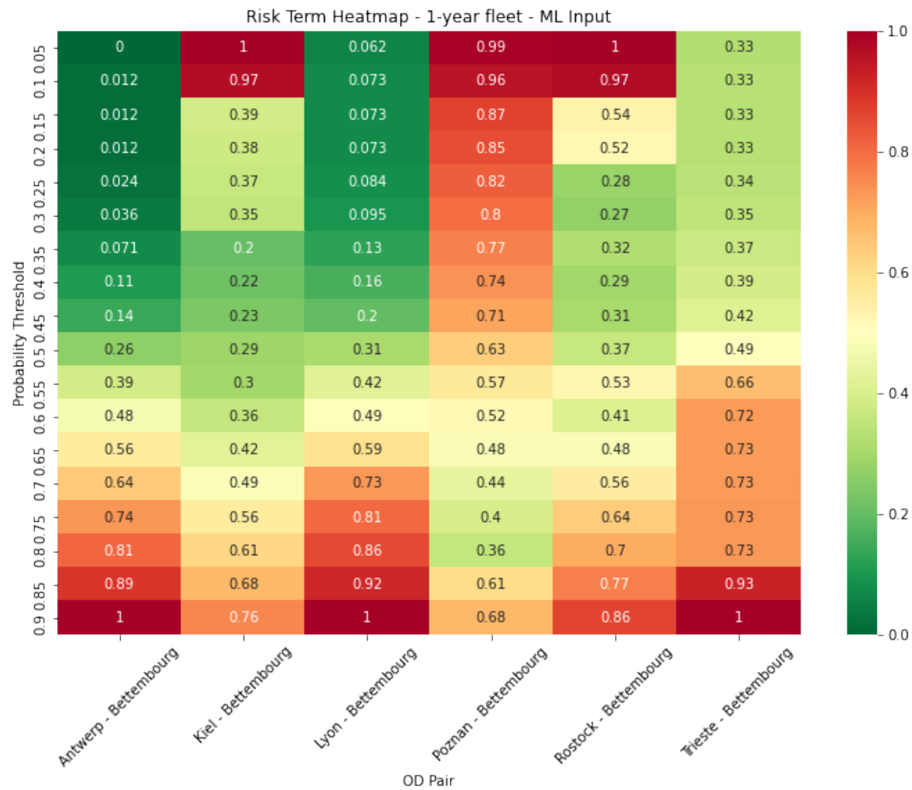


Figure 5.8: Normalized Risk per destination and probability threshold for the ML Input.

Figure 5.7 and 5.8 show the trend of the normalized risk per destination for the probability threshold. Here as well we can see that for almost every destination the risk taken by the ML Input model is higher compared to the Risk Model, due to its inflexibility in evaluating the ML model performance and acting accordingly. What is interesting to note is that there are two types of risk trends observed among the multiple destinations:

- Increasing Risk: this is the expected one, where the risk increases according to the increase in the probability threshold. As the threshold increases, the model tends to be less prone to preemptive maintenance, therefore exposing itself to the Risk of Disruption. This is the case of Antwerp, Lyon, and Trieste.

- Concave Risk: this is an unexpected one, where the risk initially decreases to then increases again steeply as we get closer to the 0.9 probability. This might be due to different wagons within that service having different probabilities of disruption and is referred to the services of Poznan, Kiel, and Rostock.

The average number of Overall Shunts is similar between the two instances, showing the overall capacity of the model in clustering. The Actual Departures behave similarly, as the shunts influence directly the departure time.

As operational time varies from service to service and depends on the rotation of the fleet trains, we cannot state anything regarding the cancellation rate. Considering the threshold for our instance (180 minutes of operational time), results of the cancellation rate are displayed in Figure 5.9.

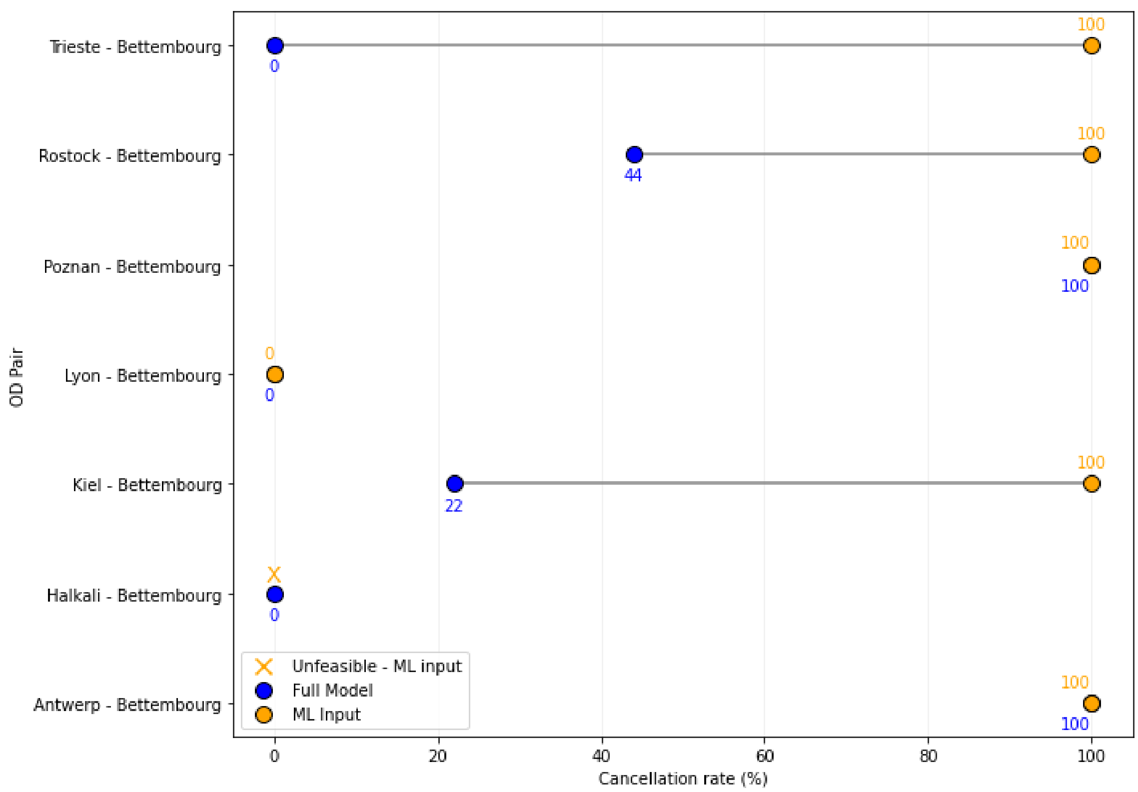


Figure 5.9: Cancellation Rate for the 1-year fleet.

We can see that, for some destinations such as Kiel, Rostock, and Trieste, the Risk Model manages to decrease up to 80% of the cancellations for these destinations.

Looking at the 3-year fleet, what we observe is that the two models are performing similarly.

This is due to the high number of scheduled maintenance operations which are triggered by the age of the fleet. This is suggested by the higher number of Operating Costs, as well as the Total Wagons Moved, Overall Shunts, and Actual Departure.

Regarding the latter, looking at Table 5.11, applying the 180 minutes threshold cancellation would result in a 100% cancellation rate for both models, as the average Actual Departure is much higher than it.

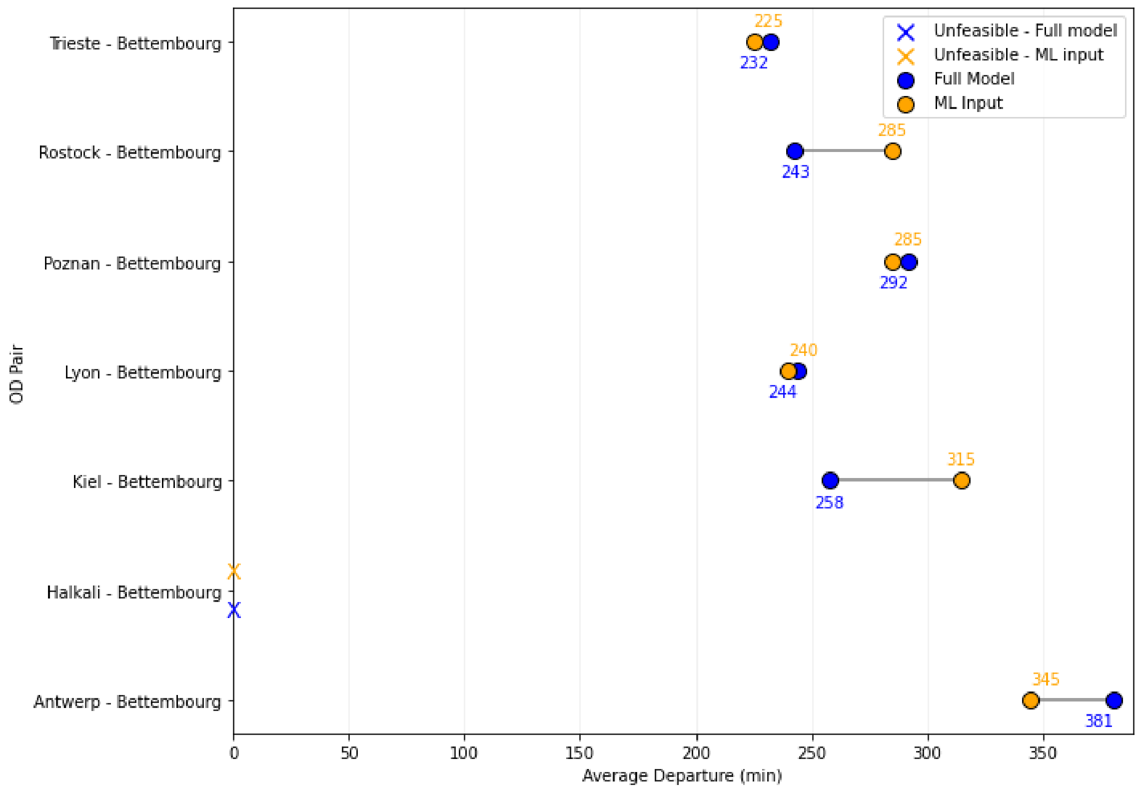


Figure 5.10: Average Departure time for the 3-year fleet.

Results for the average Actual Departure are presented in Figure 5.10. We can still confirm that for this instance, these two models behave almost equivalently, with some exceptions regarding Rostock and Kiel, where the Risk Model performed better than the ML, and Antwerp, where instead the ML model outperformed the Risk Model.

To summarize, what can conclude by looking at the result is that there is an advantage in using a risk-assessment approach compared to the ML Input model, especially when younger wagon fleets are considered and the ML model is not fed with enough data to represent a real-world scenario. The highest benefit of the application of this model is on the 1-year fleet, where there is less Risk Associated with the overall operations, together with the ability to conclude all the services, which instead the ML Input model failed at.

5.5 Summary

This chapter explores the integration of data-driven models into Mixed-Integer Linear MILP for condition-based maintenance and unplanned disruption management in freight rail operations. We propose a risk management strategy to assess the ML model against its failure, as it is our way to incorporate information regarding unplanned disruption. To do this, we developed a binary classification supervised learning for predicting unplanned maintenance, which is then tuned focusing on Recall and Precision. The final selected model, a Tuned Decision Tree (TDT) has been selected as the go-for model given its superior performance. The analysis of input features highlights the importance of monthly TEU count and actual mileage in predicting disruptions, aligning its performance and expectations with the literature. We then proposed the extension of the model proposed in Section 4.4 surpassing the shortcomings of the previous MILP, implementing probabilistic consideration, and the comparison with traditional methods. Finally, we propose a case study based on CFL Multimodal destination, showcasing the difference between the two models. Results suggest that the risk model surpasses the ML as an Input model in the 1-year-old fleet instance, behaving similarly for older fleets.

Chapter 6

Conclusion

Even if the morrow is barren of promises,
nothing shall forestall my return.

Loveless, Act V

The purpose of this dissertation was to identify and address the neglect of maintenance considerations with an integrated approach for improving shunting yard operations and management for freight trains from a strategic and tactical perspective.

6.1 Answer to the Research Questions

The main research question we asked ourselves at the beginning of this manuscript is:

What is the impact of maintenance operations in shunting operations? How can we assess the integration of maintenance and shunting operations in freight rail management within a strategic and tactical vision?

To do this, we propose different methods whose goal is to assess the impact of mileage and condition-based maintenance constraints on the operations performed in freight shunting yards in both tactical and strategic terms. We focus on the analysis of the impact of maintenance considerations in shunting operations and how these then affect multiple KPIs such as fleet requirements, mileage distribution of the fleet, delays, and cancellations. We propose two approaches: one MILP model with simulation support for maintenance-based long-term analysis of KPIs related to strategic goals; and an extension of the above-mentioned MILP which introduces condition-based maintenance consideration using a ML model. This research question has been then split into multiple research challenges, which are listed below and have been addressed throughout the manuscript.

RC1: Lack of maintenance integration in the shunting operation from a system perspective.
The first research challenge has been formulated into the RO1: *Analysis and Assessment of Maintenance Integration in Shunting Yard Operations.* We highlight this problem in Chapter 2, proposing two methodologies for solving it in the third and fourth chapters of the thesis. The review of the literature highlighted how maintenance is usually treated as a separate problem compared to the RSP and TUSP, even if it impacts the latter and inputs the former. Maintenance operations to be performed on trains trigger shunting operations, as the rolling stock has to be removed from the inbound train. From the RSP perspective, this means that a wagon has to be assigned a time-of-service in the inbound train that had the wagon removed from maintenance operations; for the TUSP aspect, this means 2 additional wagons to route throughout the shunting yard. What we also highlight is that in the literature few models perform integration of the RSP and TUSP together, but none of them tries to look at more strategic aspects of the problem. This is due to the nature of the problems but is a gap that can be addressed. The latter affirmation generated the RC2: *Lack of approaches to tackle the RSP, TUSP, and mileage-based maintenance within an integrated problem from a strategic and tactical planning point.*

RC2: Lack of approaches to tackle the RSP, TUSP and mileage-based maintenance within an integrated problem from a strategic and tactical planning point.
The second research challenge aims to bridge the gap between the integration of RSP, TUSP, and mileage-based maintenance considerations from strategic and tactical planning. This research objective has been discussed in the second chapter of the thesis and has been addressed in the RO2: *Identification of the Most Relevant KPIs Required for Strategic and Tactical Planning.* Our goal is to provide a deeper understanding of which are the strategic and tactical KPIs that can be effectively used to assess the impact of maintenance in shunting operations. This results in a comprehensive review and analysis of existing studies to identify metrics that are not only relevant but also critical

for assessing the performance and efficacy of shunting policies and maintenance strategies in railway operations. The result of this analysis is presented in Table 4.2. For the strategic KPIs, we selected the number of maintenance operations, the number of shunting operations performed, delays, cancellations, fleet size, and mileage distribution of the fleet, while for tactical KPIs we selected operating cost, overall shunts, total wagons moved, actual departure and risk term.

RC3: Lack of an hybrid model to assess the impact of the maintenance consideration modelling shunting operation for strategic assessment in freight rail.

This research challenge has been formulated in the RO3: *Development of a Hybrid Modeling Approach for Long-Term Assessment through MILP and Simulation*, and is discussed in detail in Chapter 4 of the thesis. We developed a MILP model for addressing the RSP with TUSP considerations and mileage-based constraints, creating the shunt-in/shunt-out problem (SISO). The objective of this model is to find out which wagons we need to remove from the inbound train and place in from shunting yard to minimize the shunting, delays, and cancellations when the train is subjected to mandatory and optional shunting constraints. The model has been developed together with 4 shunting policies, which are criteria to select which wagons are more suitable to reduce the strategic KPIs long term. To prove the efficacy of the proposed methodology in assessing and addressing the problem of the integration of the maintenance constraint, we developed a simulation environment that can reproduce the operations that are performed in a shunting yard. Through our case study, we show that there is a significant impact in neglecting the maintenance constraint. Moreover, among the 4 policies, AVG L-S shows the best trade-off in terms of strategic KPIs for long-term management.

RC4: Lack of complete understanding of the impact of unplanned maintenance/disruption for condition-based maintenance, and how these affect the normal shunting operation.

This research challenge has been formulated in the RO4: *Data-Driven Modeling for Condition-Based Maintenance and Unplanned Disruption, and its Integration in MILP Modeling through a Risk-Management Approach*, and is discussed in detail in Chapter 5 of the thesis.

We decided to extend the MILP model presented in Section 4.4 to overcome the shortcomings of the previous model and allow it to implement the prediction from a developed Binary Classification ML model. As the ML model provides information on the condition of the rolling stocks, we decided to implement a risk-assessment framework to assess it against its failure, to provide a better service and not be forced to accept the prediction of the model without considering contour situations, such as delays, potential cancellations etc. The case study reveals that for the 1-year instance, the risk model outperforms the classical integration of the ML as an input, while the 3-year instance provides no valuable results, as the two models are both constrained by the high number of mileage-based maintenance operations to perform.

6.1.1 Main Findings

The main findings and practical contributions of the thesis can be summarized as follows and categorized between the *strategic* impact of maintenance integration and *tactical* impact. For the strategic impact:

- In our case study for the 2020-2040 simulation, an increase of the shunting operation ranging from 2-11% has been observed based on the policy between the no maintenance scenario and the maintenance scenario. This suggests that there is a moderate impact in this matter regarding the implementation of the maintenance constraint within an integrated model.
- This impact becomes critical when looking at the fleet size, whose difference changes between 10% and 80% based on the policy between the two scenarios. This means that the under-estimation of the fleet size can bring additional costs due to leasing and cancellation due to unavailability.
- Another critical point is the increase in the annual mileage performed, which ranges from 17% to 71% based on the policy chosen between scenarios. This means that not integrating maintenance considerations can lead to under/overestimation of the mileage each wagon has to perform, leading to misaligned maintenance that can increase the risk of unplanned disruptions.
- Among the presented policies, NCLD and AVG L-S present the best overall management.

For the tactical impact:

- Key factors for predicting possible rolling stock disruption are the weight carried by each wagon, the actual mileage, the destination, and the average slope performed monthly.
- The capability of the model to decide to reject or accept ML prediction led to a more flexible MILP model, which allowed in a case study a decrease of 12% of the accepted risk for the 1-year scenario.

- For the same scenario, what is notable is that the risk model managed to serve all the destinations, while the ML model input did not. Moreover, what is also notable is that for some destinations the cancellation rate difference between the risk model and the ML as input decreased as much as 80%.
- There is no notable difference in terms of the presented KPIs when instead it comes to older fleets, as represented in the 3-year scenario. Here, the two models performed equivalently.

6.2 Future Research

The different chapters presented in this thesis have all possible value for future research, specifically when looking at the inevitable road-to-rail transition for freight. Specifically, new policies should be developed and studied, possibly tailored to the situation needed. What we believe in is that there is no one good policy among all, but the timetable and the demand that has to be served play an important part in defining which is the suitable wagon for which suitable destination. Moreover, an inside replanner should be developed within the simulator to provide more accurate results, as for now, this is not available due to the complexity of the problem. Finally, regarding the Risk-assessment condition-based model, the implementation within the simulation environment, together with the policy direction, would be desirable. This is because the MILP model developed in Chapter 5 still suits the idea of the strategical approach, the application of the policies within this model could only benefit the system long-term.

Chapter 7

Bibliography

[1] Marshalling Yards — Indian Railway Employee. URL https://indianrailwayemployee.com/content/marshalling-yards.

[2] Mileage-based freight wagon maintenance introduced in 2022 — RailTech.com. URL https://www.railtech.com/digitalisation/2021/06/03/mileage-based-freight-wagon-maintenance-introduced-in-2022/?gdpr=deny.

[3] Overview Terminals - intermodal-terminals.eu. URL https://www.intermodal-terminals.eu/database.

[4] Railway freight transport statistics - Statistics Explained. URL https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Railway_freight_transport_statistics.

[5] Rapport annuel 2022 du groupe CFL. Technical report. URL https://groupe.cfl.lu/getmedia/327259e3-6572-4a6f-b8b7-2a21c24390f8/CFL-rapport-annuel-2022-web.pdf.aspx.

[6] Train Maintenance — The Railway Technical Website — PRC Rail Consulting Ltd, . URL http://www.railway-technical.com/trains/train-maintenance/.

[7] Train Maintenance Schedule Best Practices - MPC, . URL https://www.mpofcinci.com/blog/train-maintenance-schedule-best-practices/.

[8] European Environment Agency. Greenhouse gas emissions from transport in Europe, 2021. URL https://www.eea.europa.eu/data-and-maps/indicators/transport-emissions-of-greenhouse-gases-7/assessment.

[9] Arianna Alfieri, Rutger Groot, Leo Kroon, and Alexander Schrijver. Efficient Circulation of Railway Rolling Stock. *https://doi.org/10.1287/trsc.1060.0155*, 40(3):378–391, 8 2006. ISSN 15265447. doi: 10.1287/TRSC.1060.0155. URL https://pubsonline.informs.org/doi/abs/10.1287/trsc.1060.0155.

[10] S. Beretta and D. Regazzi. Probabilistic fatigue assessment for railway axles and derivation of a simple format for damage calculations. *International Journal of Fatigue*, 86:13–23, 5 2016. ISSN 01421123. doi: 10.1016/J.IJFATIGUE.2015.08.010.

[11] Federico Bigi, Tommaso Bosi, Juan Pineda-Jaramillo, Francesco Viti, and Andrea D'Ariano. Long-term fleet management for freight trains: Assessing the impact of wagon maintenance through simulation of shunting policies. *Journal of Rail Transport Planning & Management*, 29, 3 2024. ISSN 2210-9706. doi: 10.1016/J.JRTPM.2023.100430.

[12] Technology Bigi, Federico; University of Luxembourg ¿ Faculty of Science, Medicine (FSTM) ¿ Department of Engineering (DoE), Tommaso Bosi, Technology Pineda Jaramillo, Juan Diego; University of Luxembourg ¿ Faculty of Science, Medicine (FSTM) ¿ Department of Engineering (DoE), Andrea D'Ariano, Technology Viti, Francesco; University of Luxembourg ¿ Faculty of Science, and Medicine (FSTM) ¿ Department of Engineering (DoE). ADDRESSING THE IMPACT OF MAINTENANCE IN SHUNTING OPERATIONS THROUGH SHUNT-IN POLICIES FOR FREIGHT TRAINS OPERATIONS. *Transportation Research Board (TRB) 102nd Annual Meeting*, 2023. URL https://orbilu.uni.lu/handle/10993/54161.

[13] Tommaso Bosi, Federico Bigi, Andrea D'Ariano, Francesco Viti, and Juan Pineda-Jaramillo. Optimal Management of Full Train Load Services in the Shunting Yard: A Comprehensive Study on Shunt-In Shunt-Out Policies. *Computers & Industrial Engineering*, 2 2023. ISSN 03608352. doi: 10.1016/J.CIE.2023.109865.

[14] Nils Boysen, Malte Fliedner, Florian Jaehn, and Erwin Pesch. Shunting yard operations:

Theoretical aspects and applications. *European Journal of Operational Research*, 220(1):1–14, 7 2012. ISSN 03772217. doi: 10.1016/J.EJOR.2012.01.043.

[15] Nils Boysen, Armin Scholl, and Nico Wopperer. Resequencing of mixed-model assembly lines: Survey and research agenda. *European Journal of Operational Research*, 216(3):594–604, 2012. ISSN 03772217. doi: 10.1016/j.ejor.2011.08.009. URL https://linkinghub.elsevier.com/retrieve/pii/S0377221711007284.

[16] Valentina Cacchiani, Alberto Caprara, and Paolo Toth. Solving a real-world train-unit assignment problem. *Mathematical Programming*, 124(1-2):207–231, 2010. ISSN 00255610. doi: 10.1007/s10107-010-0361-y.

[17] Luis Cadarso and Ángel Marín. Robust rolling stock in rapid transit networks. *Computers & Operations Research*, 38(8):1131–1142, 2011. ISSN 0305-0548. doi: https://doi.org/10.1016/j.cor.2010.10.029. URL https://www.sciencedirect.com/science/article/pii/S0305054810002601.

[18] Luis Cadarso and Ángel Marín. Integration of timetable planning and rolling stock in rapid transit networks. *Annals of Operations Research*, 199(1):113–135, 2012. ISSN 02545330. doi: 10.1007/s10479-011-0978-0.

[19] Chris Carr and Josef Doppelbauer. Making the railway system work better for society. FOSTERING THE RAILWAY SECTOR THROUGH THE EUROPEAN GREEN DEAL Making the railway system work better for society. Technical Report 0, European Union - Agency for Railways, 2020.

[20] Guo Chuijiang. Optimization model and algorithm for the placing-in and taking-out of wagons in railway stations with branch-shaped goods operating sites. *https://doi.org/10.1080/23248378.2020.1823254*, 9(6):579–594, 2020. ISSN 23248386. doi: 10.1080/23248378.2020.1823254. URL https://www.tandfonline.com/doi/abs/10.1080/23248378.2020.1823254.

[21] European Commission. Freight transport statistics, 2023. URL https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Freight_transport_statistics_-_modal_split&oldid=563291.

[22] Elias Dahlhaus, Peter Horak, Mirka Miller, and Joseph F. Ryan. The train marshalling problem. *Discrete Applied Mathematics*, 103(1-3):41–54, 2000. ISSN 0166218X. doi: 10.1016/S0166-218X(99)00219-X.

[23] Andrea D'Ariano, Dario Pacciarelli, and Marco Pranzo. Assessment of flexible timetables in real-time traffic management of a railway bottleneck. *Transportation Research Part C: Emerging Technologies*, 16(2):232–245, 2008. ISSN 0968090X. doi: 10.1016/j.trc.2007.07.006.

[24] Samuel Deleplanque, Pierre Hosteins, Paola Pellegrini, and Joaquin Rodriguez. Train management in freight shunting yards: Formalisation and literature review. doi: 10.1049/itr2.12216{\"{i}}. URL https://hal.science/hal-03736251.

[25] Roger Deuce, Anders Ekberg, and Elena Kabo. Mechanical deterioration of wheels and rails under winter conditions – mechanisms and consequences. *https://doi.org/10.1177/0954409718802437*, 233(6):640–648, 10 2018. ISSN 20413017. doi: 10.1177/0954409718802437. URL https://journals.sagepub.com/doi/abs/10.1177/0954409718802437?journalCode=pifa.

[26] C. Tyler Dick. Influence of mainline schedule flexibility and volume variability on railway classification yard performance. *Journal of Rail Transport Planning & Management*, 20: 100269, 12 2021. ISSN 2210-9706. doi: 10.1016/J.JRTPM.2021.100269.

[27] C. Tyler Dick. Quantifying the Relative Influence of Railway Hump Classification Yard Performance Factors. *Journal of Transportation Engineering, Part A: Systems*, 147(8): 04021037, 5 2021. ISSN 2473-2907. doi: 10.1061/JTEPBS.0000529. URL https://ascelibrary.org/doi/abs/10.1061/JTEPBS.0000529https://ascelibrary.org/doi/10.1061/JTEPBS.0000529.

[28] Anders Ekberg and Elena Kabo. Fatigue of railway wheels and rails under rolling contact and thermal loading-an overview. *Wear*, 258(7-8):1288–1300, 3 2005. ISSN 00431648. doi: 10.1016/J.WEAR.2004.03.039.

[29] Errac. Towards a Competitive, Resource Efficient and Intelligent Rail Transport System for 2050. Technical report. URL https://www.ptferroviaria.es/docs/Documentos/Railroute2050_ERRAC.pdf.

[30] European Commission. A Roadmap for moving to a competitive low carbon economy in 2050. Technical report.

[31] European Commission. Transport 2050: Commission outlines ambitious plan to increase mobility and reduce emissions, 2011. URL https://ec.europa.eu/commission/presscorner/detail/en/IP_11_372.

[32] FAIR Institute. FAIR Institute, 2023. URL https://www.fairinstitute.org/.

[33] Constanze Fetting. "The European Green Deal", ESDN Report. *European Commission*, 53 (9):24, 2020. ISSN 1098-6596. URL https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52019DC0640&from=EN.

[34] Teodor Gabriel Crainic and Tolga Bektas. Brief Overview of Intermodal Transportation. *Logistics Engineering Handbook*, pages 28–1, 12 2007. doi: 10.1201/9780849330537.CH28.

[35] Viktor Gerdun, Tomaž Sedmak, Viktor Šinkovec, Igor Kovše, and Bojan Cene. Failures of bearings and axles in railway freight wagons. *Engineering Failure Analysis*, 14(5):884–894, 7 2007. ISSN 13506307. doi: 10.1016/J.ENGFAILANAL.2006.11.044.

[36] Giovanni Luca Giacco, Donato Carillo, Andrea D'Ariano, Dario Pacciarelli, and Ángel G. Marín. Short-term rail rolling stock rostering and maintenance scheduling. In *Transportation Research Procedia*, 2014. doi: 10.1016/j.trpro.2014.10.044.

[37] Brady Gilg, Torsten Klug, Rosemarie Martienssen, Joseph Paat, Thomas Schlechte, Christof Schulz, Senan Seymen, and Alexander Tesch. Conflict-free railway track assignment at depots. *Journal of Rail Transport Planning and Management*, 8(1):16–28, 2018. ISSN 22109706. doi: 10.1016/j.jrtpm.2017.12.004. URL https://doi.org/10.1016/j.jrtpm.2017.12.004.

[38] Paolo Guglielminetti, Cristiana Piccioni, Gaetano Fusco, Riccardo Licciardello, and Antonio Musso. Single Wagonload Traffic in Europe: Challenges, prospects and policy options. *Ingegneria Ferroviaria*, 70(11):927–948, 2015. ISSN 00200956.

[39] Chuijiang Guo and Dingyou Lei. Model of wagons' placing-in and taking-out problem in a railway station and its heuristic algorithm. *Mathematical Problems in Engineering*, 2014, 2014. ISSN 15635147. doi: 10.1155/2014/493809.

[40] Chuijiang Guo and Dingyou Lei. Model of wagons' placing-in and taking-out problem in a railway station and its heuristic algorithm. *Mathematical Problems in Engineering*, 2014, 2014. ISSN 15635147. doi: 10.1155/2014/493809.

[41] Jørgen Haahr and Richard M. Lusby. Integrating rolling stock scheduling with train unit shunting. *European Journal of Operational Research*, 2017. ISSN 03772217. doi: 10.1016/j.ejor.2016.10.053.

[42] Jørgen T. Haahr, Joris C. Wagenaar, Lucas P. Veelenturf, and Leo G. Kroon. A comparison of two exact methods for passenger railway rolling stock (re)scheduling. *Transportation Research Part E: Logistics and Transportation Review*, 91:15–32, 7 2016. ISSN 1366-5545. doi: 10.1016/J.TRE.2016.03.019.

[43] Jørgen Thorlund Haahr and Simon Henry Bull. Exact Methods for Solving the Train Departure Matching Problem. *Management Engineering*, 2015.

[44] Jørgen Thorlund Haahr, Richard M. Lusby, and Joris Camiel Wagenaar. Optimization methods for the Train Unit Shunting Problem. *European Journal of Operational Research*, 262(3):981–995, 11 2017. ISSN 0377-2217. doi: 10.1016/J.EJOR.2017.03.068.

[45] Nathalie Herr, Jean Nicod, Christophe Varnier, Noureddine Zerhouni, Malek Cherif, Farhat Fnaiech, Al Joint, Jean-Marc Nicod, and Nader Fnaiech. Joint optimization of train assignment and predictive maintenance scheduling. In *e 7 th International Conference on Railway Operations Modelling and Analysis (RailLille2017)*, 2017. URL https://hal.science/hal-02868240.

[46] Ping Huang, Chao Wen, Liping Fu, Qiyuan Peng, and Zhongcan Li. A hybrid model to improve the train running time prediction ability during high-speed railway disruptions. *Safety Science*, 122, 2 2020. ISSN 0925-7535. doi: 10.1016/J.SSCI.2019.104510.

[47] Florian Jaehn and Steffen Michaelis. Shunting of trains in succeeding yards. *Computers & industrial engineering*, 102:1–9, 12 2016. ISSN 03608352. doi: 10.1016/J.CIE.2016.10.006.

[48] Florian Jaehn, Johannes Rieder, and Andreas Wiehl. Minimizing delays in a shunting yard. *OR Spectrum*, 37(2):407–429, 2015. ISSN 14366304. doi: 10.1007/s00291-015-0391-1.

[49] Florian Jaehn, Johannes Rieder, and Andreas Wiehl. Single-stage shunting minimizing weighted departure times. *Omega*, 52:133–141, 4 2015. ISSN 0305-0483. doi: 10.1016/J.OMEGA.2014.11.001.

[50] Franck Kamenga, Paola Pellegrini, Joaquin Rogriguez, Boubekeur Merabet, and Bertrand Houzel. Train Unit Shunting : Integrating rolling stock maintenance and capacity management in passenger railway stations. *8th International Conference on Railway Operations Modelling and Analysis*, 2019.

[51] Franck Kamenga, Paola Pellegrini, Joaquin Rodriguez, and Boubekeur Merabet. Solution algorithms for the generalized train unit shunting problem. *EURO Journal on Transportation and Logistics*, 10:100042, 1 2021. ISSN 2192-4376. doi: 10.1016/J.EJTL.2021.100042.

[52] E R Kraft. A hump sequencing algorithm for real time management of train connection reliability. *Journal of the Transportation Research Forum*, 39(4), 2000. ISSN 1046-1469.

[53] Leo G. Kroon, Ramon M. Lentink, and Alexander Schrijver. Shunting of passenger train units: An integrated approach. *Transportation Science*, 42(4):436–449, 2008. ISSN 15265447. doi: 10.1287/trsc.1080.0243.

[54] Ramon Martijn Lentink. *Algorithmic Decision Support for Shunt Planning*. PhD thesis, School of Economics, Erasmus University Rotterdam, 2006.

[55] Bing Li, Wenyao Qu, Hua Xuan, Chunqiu Xu, Xueyuan Wang, and Yifei Geng. Shunting Strategy for Placing-In and Taking-Out Wagons on Branch-Shaped Freight Operation Network with Simulation Method. *Journal of Advanced Transportation*, 2020:1–25, 1 2020. ISSN 0197-6729. doi: 10.1155/2020/3410481. URL https://www.hindawi.com/journals/jat/2020/3410481/.

[56] Jian Li, Boliang Lin, Zhongkai Wang, Lei Chen, and Jiaxi Wang. A Pragmatic Optimization Method for Motor Train Set Assignment and Maintenance Scheduling Problem. *Discrete Dynamics in Nature and Society*, 2016, 2016. ISSN 1607887X. doi: 10.1155/2016/4540503.

[57] R Licciardello, N Adamko, S Deleplanque, P Hosteins, R Liu, P Pellegrini, A Peterson, M Wahlborg, and M Za'tko. Integrating yards, network and optimisation models towards real-time rail freight yard operations. 2020. ISSN 0020-0956. URL

http://www.ingegneriaferroviaria.it/web/en/content/integrazione-di-impianti-merci-rete-e-modelli-di-ottimizzazione-l%E2%80%99esercizio-tempo-reale.

[58] Boliang Lin and Yinan Zhao. Synchronized optimization of EMU train assignment and second-level preventive maintenance scheduling. *Reliability Engineering & System Safety*, 215:107893, 11 2021. ISSN 0951-8320. doi: 10.1016/J.RESS.2021.107893.

[59] Boliang Lin, Yaoming Shen, Zhongkai Wang, Shaoquan Ni, and Yinan Zhao. An iterative improvement approach for high-speed train maintenance scheduling. *Transportation Research Part B: Methodological*, 173:292–312, 7 2023. ISSN 01912615. doi: 10.1016/j.trb.2023.05.008.

[60] Xu Ying Liu, Jianxin Wu, and Zhi Hua Zhou. Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39(2): 539–550, 2009. ISSN 10834419. doi: 10.1109/TSMCB.2008.2007853.

[61] Richard M. Lusby, Jørgen Thorlund Haahr, Jesper Larsen, and David Pisinger. A Branch-and-Price algorithm for railway rolling stock rescheduling. *Transportation Research Part B: Methodological*, 99:228–250, 5 2017. ISSN 01912615. doi: 10.1016/J.TRB.2017.03.003.

[62] T. Makino, H. Sakai, C. Kozuka, Y. Yamazaki, M. Yamamoto, and K. Minoshima. Overview of fatigue damage evaluation rule for railway axles in Japan and fatigue property of railway axle made of medium carbon steel. *International Journal of Fatigue*, 132:105361, 3 2020. ISSN 0142-1123. doi: 10.1016/J.IJFATIGUE.2019.105361.

[63] P. D. McFadden and J. D. Smith. Vibration monitoring of rolling element bearings by the high-frequency resonance technique - a review. *Tribology International*, 17(1):3–10, 1984. ISSN 0301679X. doi: 10.1016/0301-679X(84)90076-8.

[64] Luís Mira, António R. Andrade, and Marta Castilho Gomes. Maintenance scheduling within rolling stock planning in railway operations under uncertain maintenance durations. *Journal of Rail Transport Planning & Management*, 14:100177, 6 2020. ISSN 2210-9706. doi: 10.1016/J.JRTPM.2020.100177.

[65] Jesse Mulderij, Bob Huisman, Denise Tönissen, and Koos Van Der Linden. Train Unit Shunting and Servicing: a Real-Life Application of Multi-Agent Path Finding.

[66] Lars Kjær Nielsen, Leo Kroon, and Gábor Maróti. A rolling horizon approach for disruption management of railway rolling stock. *European Journal of Operational Research*, 220(2): 496–509, 7 2012. ISSN 0377-2217. doi: 10.1016/J.EJOR.2012.01.037.

[67] Carsten J. Peters and Dietmar Eifler. Influence of service temperatures on the fatigue behaviour of railway wheel and tyre steels. *Materialpruefung/Materials Testing*, 51(11-12): 748–754, 11 2009. ISSN 00255300. doi: 10.3139/120.110094/MACHINEREADABLECITATION/RIS. URL https://www.degruyter.com/document/doi/10.3139/120.110094/html.

[68] M. F.E. Peters. Early warnings for failing train axle bearings based on temperature. *Proceedings of the Annual Conference of the Prognostics and Health Management Society, PHM*, pages 370–375, 2017. ISSN 23250178.

[69] Juan Pineda-Jaramillo and Francesco Viti. Identifying the rail operating features associated to intermodal freight rail operation delays. *Transportation Research Part C: Emerging Technologies*, 147, 2 2023. ISSN 0968-090X. doi: 10.1016/J.TRC.2022.103993.

[70] Xiang-chun Qi, Y. Yue, Jun-Hee Han, Leishan Zhou, and Hesham A Rakha. Integrated Modelling and Optimization of Train Scheduling and Shunting at Complex Railway Passenger Stations. 2017.

[71] Leo G. Kroon Dennis Huisman Richard Freling, Ramon M. Lentink and ERIM. Shunting of Passenger Train Units in a Railway Station. *ERIM Report Series Research in Management*, 2002.

[72] Moritz Ruf and Jean François Cordeau. Adaptive large neighborhood search for integrated planning in railroad classification yards. *Transportation Research Part B: Methodological*, 150:26–51, 8 2021. ISSN 01912615. doi: 10.1016/J.TRB.2021.05.012.

[73] D. Šemrov, R. Marsetič, M. Žura, L. Todorovski, and A. Srdic. Reinforcement learning approach for train rescheduling on a single-track railway. *Transportation Research Part B: Methodological*, 86:250–267, 4 2016. ISSN 0191-2615. doi: 10.1016/J.TRB.2016.01.004.

[74] Xinggao Shu. Wheel Profile Maintenance Guidelines. *Wheel Profile Maintenance Guidelines*, 4 2015. doi: 10.17226/22168.

[75] Di Loreto Simona, Malíková Ludmila, Tomekova Ivana, Mirkovic Tatjana, Ohlin Mikaela, and Maliaček Juraj. Overview of priority rules in operation, 2020. URL https://rne.eu/wp-content/uploads/RNE_OverviewOfthePriorityRulesInOperation_V9.1.pdf.

[76] Sky TG24. Firenze, deraglia Treno Merci. Abbattuti Pali di alimentazione., 2023. URL https://tg24.sky.it/cronaca/2023/04/20/firenze-treno-deragliato-foto.

[77] Alexandre Trilla, John Bob-Manuel, Benjamin Lamoureux, and Xavier Vilasis-Cardona. Integrated Multiple-Defect Detection and Evaluation of Rail Wheel Tread Images using Convolutional Neural Networks. *International Journal of Prognostics and Health Management*, 12(1), 2021. ISSN 2153-2648. doi: 10.36001/ijphm.2021.v12i1.2906.

[78] Urząd Transportu Kolejowego. Train punctuality in 3Q 2023, 2023. URL https://utk.gov.pl/en/new/20668,Train-punctuality-in-3Q-2023.html.

[79] D.A. van Cuilenborg. *Train Unit Shunting Problem, a Multi-Agent Pathfinding approach*. PhD thesis, 2020. URL https://repository.tudelft.nl/islandora/object/uuid%3Ac9a0f41c-de6f-4ef4-8d94-6d0a89df3eec.

[80] Roel van den Broek, Han Hoogeveen, Marjan van den Akker, and Bob Huisman. A Local Search Algorithm for Train Unit Shunting with Service Scheduling. *https://doi.org/10.1287/trsc.2021.1090*, 56(1):141–161, 11 2021. ISSN 15265447. doi: 10.1287/TRSC.2021.1090. URL https://pubsonline.informs.org/doi/abs/10.1287/trsc.2021.1090.

[81] Roel W. Van Den Broek. *Train Shunting and Service Scheduling: an integrated local search approach*. PhD thesis, 2016.

[82] Yihui Wang, Kangqi Zhao, Andrea D'ariano, Ru Niu, Shukai Li, and Xiaojie Luan. Real-time integrated train rescheduling and rolling stock circulation planning for a metro line under disruptions. *Transportation Research Part B*, 152, 2021. doi: 10.1016/j.trb.2021.08.003. URL https://doi.org/10.1016/j.trb.2021.08.003.

[83] Thomas Winter. Online and Real-Time Dispatching Problems. (May):250, 1999.

[84] Winter and Zimmermann. Real-Time Dispatch of Trams in Storage Yards. *Annals of Operations Research*, 97(1-4):131–141, 2000. ISSN 02545330. doi: 10.1023/A.

[85] Xiaoming Xu and Maged M. Dessouky. Train shunting with service scheduling in a high-speed railway depot. *Transportation Research Part C: Emerging Technologies*, 143: 103819, 10 2022. ISSN 0968-090X. doi: 10.1016/J.TRC.2022.103819.

[86] Optiyard Optimized Real-time Yard and Network Management. Collaborative project GA-777594 Yard Optimization Algorithm , Network Decision-Support Tool and Integration Framework. (2019):1–56, 2020.

[87] Jiateng Yin, Fan Pu, Lixing Yang, Andrea D'ariano, and Zhouhong Wang. Integrated optimization of rolling stock allocation and train timetables for urban rail transit networks: A benders decomposition approach. *Transportation Research Part B*, 176:102815, 2023. doi: 10.1016/j.trb.2023.102815. URL https://doi.org/10.1016/j.trb.2023.102815.

[88] Jianxin Yuan. Stochastic modelling of train delays and delay propagation in stations. 2006.

[89] Yixiang Yue, Juntao Han, Shifeng Wang, and Xiang Liu. Integrated Train Timetabling and Rolling Stock Scheduling Model Based on Time-Dependent Demand for Urban Rail Transit. *Computer-Aided Civil and Infrastructure Engineering*, 32:856–873, 2017. doi: 10.1111/mice.12300.

[90] A. A. Zagorskikh, O. K. Alsova, and A. V. Gavrilov. A simulation model of the process of organizing the movement and maintenance of trains at a Railway Station. *Journal of Physics: Conference Series*, 1661(1):012200, 11 2020. ISSN 1742-6596. doi: 10.1088/1742-6596/1661/1/012200. URL https://iopscience.iop.org/article/10.1088/1742-6596/1661/1/012200https://iopscience.iop.org/article/10.1088/1742-6596/1661/1/012200/meta.

[91] Peng Zhao, Yawei Li, Baoming Han, Ruixia Yang, and Zhiping Liu. Integrated Optimization of Rolling Stock Scheduling and Flexible Train Formation Based on Passenger Demand for an Intercity High-Speed Railway. *Sustainability 2022, Vol. 14, Page 5650*, 14(9):5650, 5 2022. ISSN 2071-1050. doi: 10.3390/SU14095650. URL https://www.mdpi.com/2071-1050/14/9/5650/htm.

[92] Yaqiong Zhao, Dewei Li, and Yonghao Yin. Integrated optimization of train formation plan and rolling stock scheduling under Time-dependent demand. *Computers & Operations Research*, 150:106049, 2 2023. ISSN 0305-0548. doi: 10.1016/J.COR.2022.106049.

[93] Qingwei Zhong, Richard M. Lusby, Jesper Larsen, Yongxiang Zhang, and Qiyuan Peng. Rolling stock scheduling with maintenance requirements at the Chinese High-Speed Railway. *Transportation Research Part B: Methodological*, 126:24–44, 8 2019. ISSN 0191-2615. doi: 10.1016/J.TRB.2019.05.013.

[94] Aurelius A. Zilko, Dorota Kurowicka, and Rob M.P. Goverde. Modeling railway disruption lengths with Copula Bayesian Networks. *Transportation Research Part C: Emerging Technologies*, 68:350–368, 7 2016. ISSN 0968090X. doi: 10.1016/J.TRC.2016.04.018.

Appendix A

Shunt-In/Shunt-Out (SISO) Model Formulation

$$\begin{aligned}
&\min \sum_{i \in \mathcal{T}} \gamma_i - \sum_{i=1}^{|\mathcal{T}|-1} adj_{i,i+1} + \frac{|\mathcal{T}|}{2}(\sigma_2 + \sigma_3) + \\
&+ \frac{|\mathcal{T}|}{4}(\sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{S}} (c_{s_{i,j}} \alpha + n_{ms_j} \beta) z_{i,j} + \sum_{i \in \mathcal{T}} y_i c_{u_i}
\end{aligned} \tag{A.1}$$

$$ad_T \ \leq \ d_T \ + \sigma_1 M + \sigma_2 M \tag{A.2}$$

$$dd_T \ + \ (1 \ - \ \sigma_1)M \ \geq \ ad_T \tag{A.3}$$

$$ad_T \ > \ d_T \ - \ (1 - \sigma_1)M \tag{A.4}$$

$$ad_T \ \leq \ dd_T \ + \ \sigma_2 M \tag{A.5}$$

$$ad_T \ > \ dd_T \ - \ (1 - \sigma_2)M \tag{A.6}$$

$$\sigma_1 \ + \ \sigma_2 \ \leq \ 1 \tag{A.7}$$

$$\sigma_3 \ \leq \ \sigma_1 M \tag{A.8}$$

$$\sigma_3 \ \geq \ \frac{a_T - d_T}{dd_T - d_T}\sigma_1 \tag{A.9}$$

$$\sigma_3 \ \leq \ \frac{ad_T - d_T}{dd_T - d_T} \ + \ (1 - \sigma_1)M \tag{A.10}$$

$$\sigma_3 \ \geq \ \frac{ad_T - d_T}{dd_T - d_T} \ - \ (1 - \sigma_1)M \tag{A.11}$$

$$a_T + (\sum_{i \in \mathcal{T}} \gamma_i - \sum_{i=1}^{|\mathcal{T}|-1} adj_{i,i+1})ts + \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{S}} c_{s_{i,j}} = ad_T \tag{A.12}$$

$$\frac{a_T \ + \ (\sum_{i\in\mathcal{T}} \gamma_i - \sum_{i=1}^{|\mathcal{T}|-1} adj_{i,i+1})ts}{dd_T} \ = \ \alpha \tag{A.13}$$

$$1-\alpha = \ \beta \tag{A.14}$$

$$y_i \geq \frac{m_i + r_T}{m_{max_i}} - 1 - (\sum_{k\in\mathcal{K}} x_{i,k})M \qquad \forall i\in\mathcal{T} \tag{A.15}$$

$$y_i \leq (1-\sum_{k\in\mathcal{K}} x_{i,k})\frac{m_i + r_T}{m_{min_i}} \qquad \forall i\in\mathcal{T} \tag{A.16}$$

$$y_i \leq \frac{m_i + r_T}{m_{max_i}} \qquad \forall i\in\mathcal{T} \tag{A.17}$$

$$\sum_{i\in\mathcal{T}:type_{in_i}\neq type_k} x_{i,type_k} = num_k \tag{A.18}$$

$$\sum_{k\in\mathcal{K}} x_{i,k} = 0 \qquad \forall i\in\mathcal{T}: type_{in_i} = type_k \tag{A.19}$$

$$\sum_{i\in\mathcal{T}} x_{i,k} = 0 \qquad \forall k\in\mathcal{K}: k\neq type_k \tag{A.20}$$

$$|type_{out_i} - type_{in_i}| = p_i \qquad \forall i\in\mathcal{T} \tag{A.21}$$

$$\sum_{j\in\mathcal{S}:type_{S_j}=type_{in_i}} z_{i,j} = y_i \qquad \forall i\in\mathcal{T} \tag{A.22}$$

$$\sum_{j\in\mathcal{S}:type_{S_j}\neq type_{in_i}} z_{i,j} = x_{i,type_k} \qquad \forall i\in\mathcal{T} \tag{A.23}$$

$$z_{i,j} \leq 2 - \frac{ms_j + r_T}{ms_{max_j}} \qquad \forall i\in\mathcal{T}, \forall j\in\mathcal{S} \tag{A.24}$$

$$\sum_{i\in\mathcal{T}} z_{i,j} \leq 1 \qquad \forall j\in\mathcal{S} \tag{A.25}$$

$$\sum_{k\in\mathcal{K}} x_{i,k} + y_i = \gamma_i \qquad \forall i\in\mathcal{T} \tag{A.26}$$

$$2adj_{i,i+1} \leq \gamma_i + \gamma_{i+1} \qquad \forall i = 1,...,|\mathcal{T}|-1 \tag{A.27}$$

$$\sum_{j\in\mathcal{S}} z_{i,j} code_{S_j} + (1-\gamma_i)code_{in_i} = code_{\mathrm{out}_i} \qquad \forall i\in\mathcal{T} \tag{A.28}$$

Appendix B

Event-Based Simulation Framework (EBSF) Flowchart

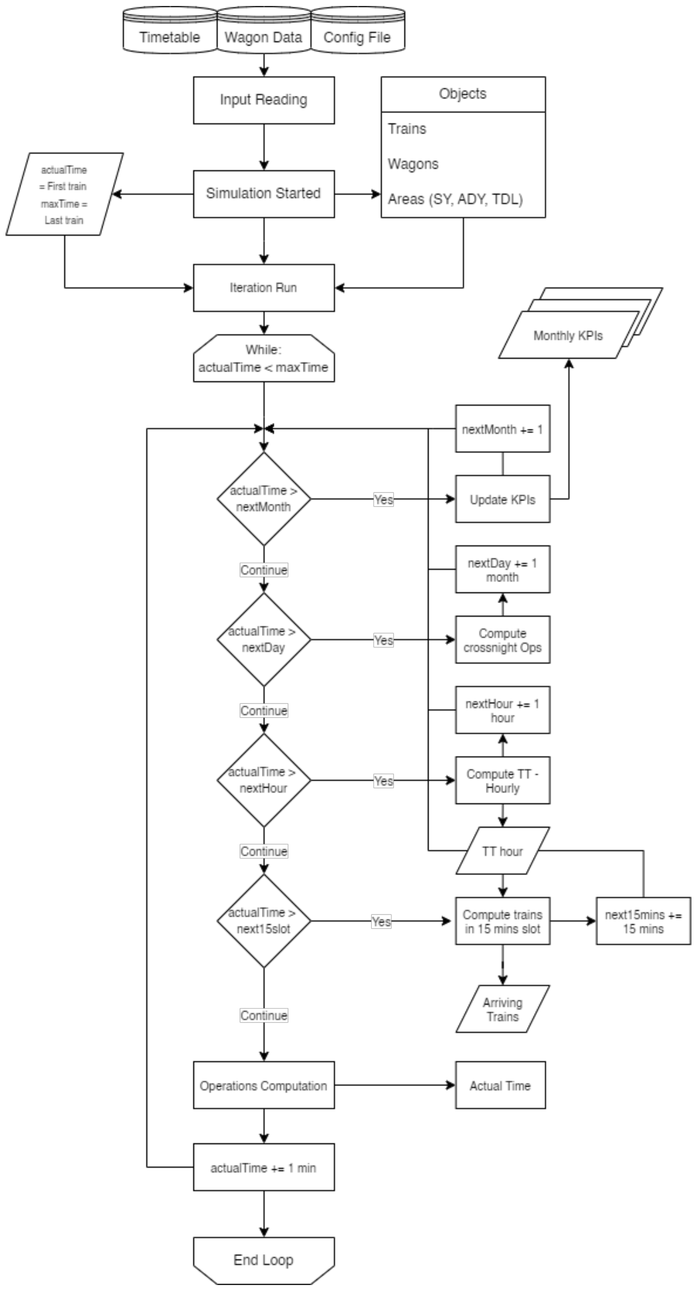


Figure B.1: First part of the EBSF flowchart providing a comprehensive illustration of all the pre- and post-processing operations that occur outside of the yards.

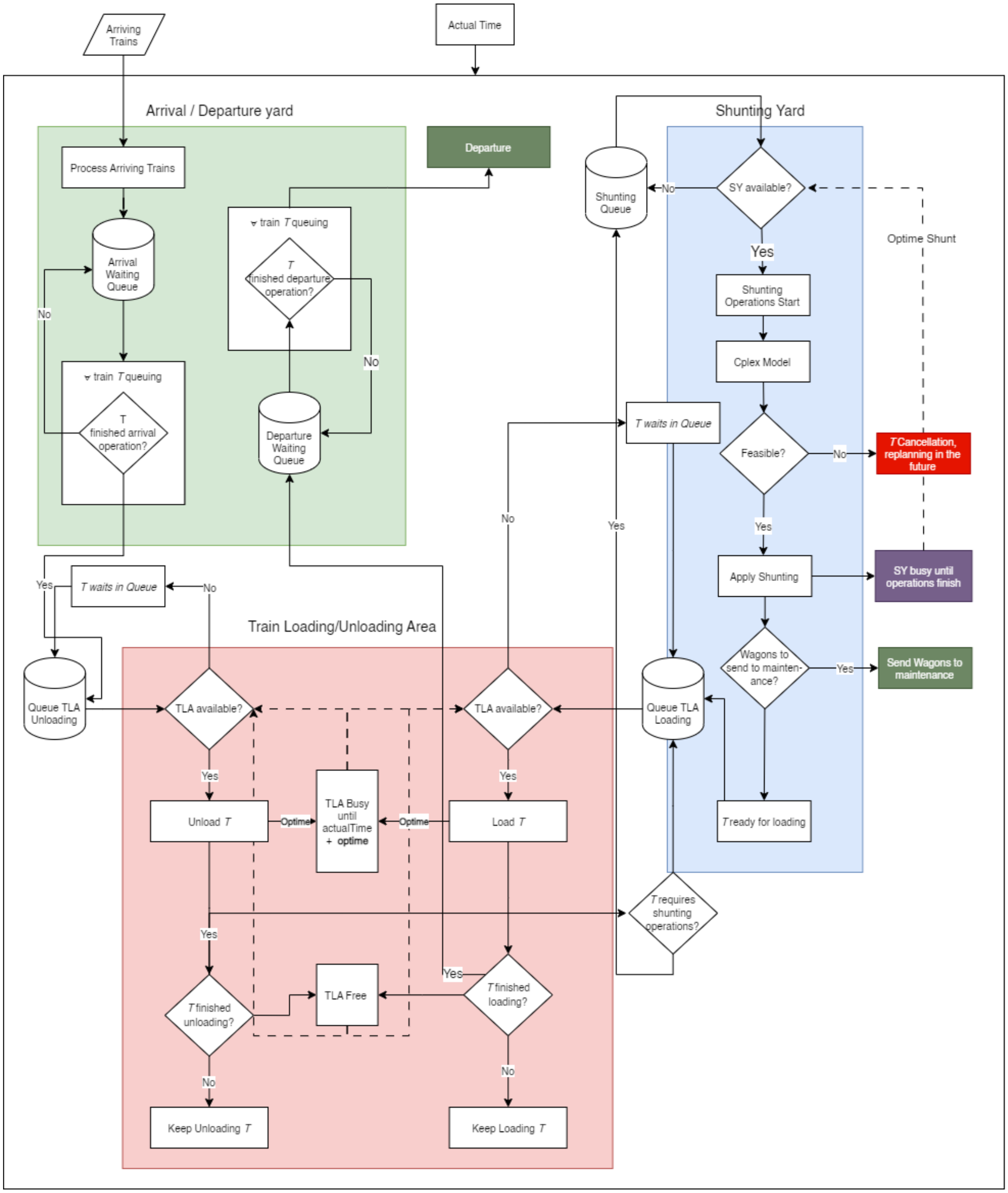


Figure B.2: Second part of the EBSF flowchart depicting all the operations carried out in the various areas within the station - Arrival/Departure Yard in Green, Train Loading/Unloading area in Red, Shunting yard in Blue.

Appendix C

Data Joining and Feature Engineering for the ML model - Flowcharts

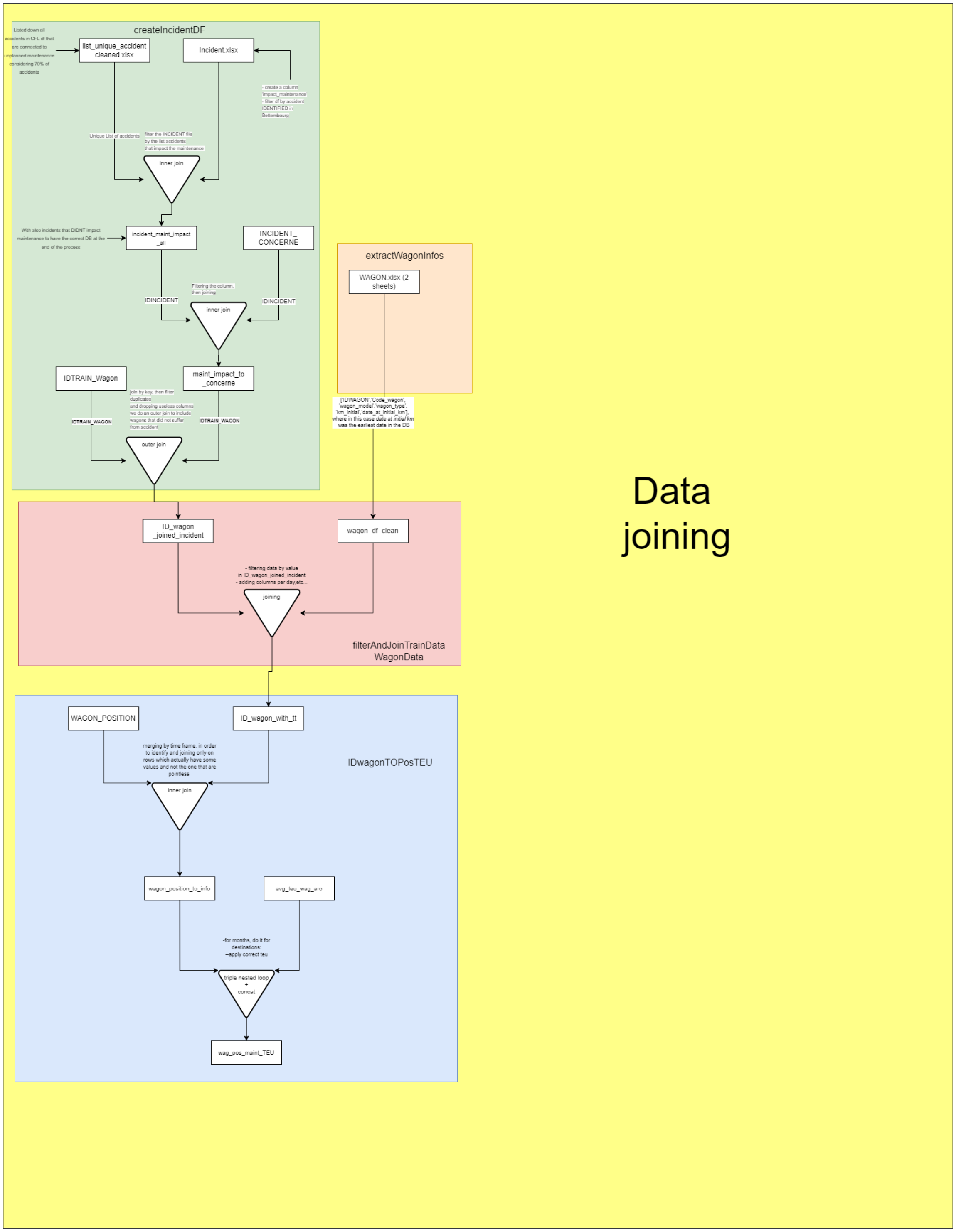


Figure C.1: Data joining phase related to the initial dataset.

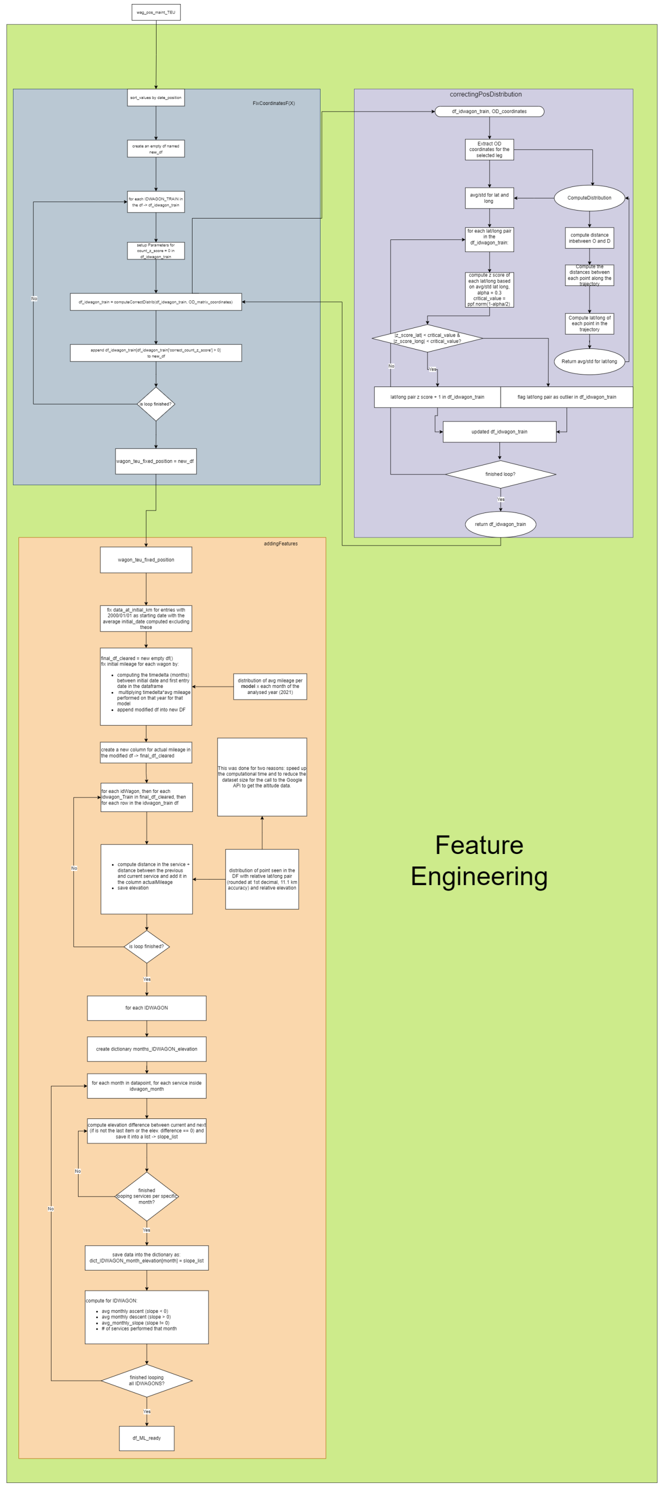


Figure C.2: Feature Engineering phase related to the initial dataset.

Appendix D

MILP Model with Condition-Based Maintenance Integration

$$\max \quad Rev_{rail} - (C^{\text{maint}} + shunts \cdot C_{\text{shunt}} + Risk + Canc^{trig} \cdot C^{\text{canc}}) \tag{D.1}$$

$$\sum_{i\in\mathcal{T}} (1 - w_i^{\text{SO}}) \cdot rev_i + \sum_{j\in TRK} \sum_{k\in PTRK} w_{i,j,k}^{\text{SI}} * rev_{j,k} = Rev_{rail} \tag{D.2}$$

$$\sum_{i\in\mathcal{T}} R_i^{\text{SI}} + R_i^{\text{SO}} = Risk \tag{D.3}$$

$$\begin{aligned}
& pred_i^{\text{ML}} \cdot \{TPR \cdot [\\
& (\sum_{g\in\text{gravity}} P_g \cdot R_{i,g}^{\text{disr}}) \cdot (1 - ML_i^{\text{maint}}) + (C_{\text{single}}^{\text{shunt}} + C^{\text{prev,maint}}) * ML_i^{\text{maint}} \\
&] + FPR * [(C_{\text{single}}^{\text{shunt}} + C^{\text{prev,maint}} + R^{\text{point}}) * ML_i^{\text{maint}}]\} \\
& +(1 - pred_i^{\text{ML}}) \cdot \\
& \{TNR \cdot (C_{\text{single}}^{\text{shunt}} + R^{\text{point}} + C^{\text{prev,maint}}) \cdot ML_i^{\text{maint}} \\
& +FNR \cdot [(\sum_{g\in\text{gravity}} P_g * R_{i,g}^{\text{disr}}) \cdot (1 - ML_i^{\text{maint}}) \\
& +(C_{\text{single}}^{\text{shunt}} + C^{\text{prev,maint}}) \cdot ML_i^{\text{maint}}]\} \\
& = R_i^{\text{SO}} \quad \forall i \in \mathcal{T}
\end{aligned} \tag{D.4}$$

$$\begin{aligned}
& \sum_{j\in TRK} \sum_{k\in PTRK} \{[pred_{j,k}^{\text{ML}} \cdot \\
& TPR \cdot (\sum_{g\in\text{gravity}} P_g \cdot R_{i,j,k}^{\text{disr}}) \cdot w_{i,j,k}^{\text{SI}}] + \\
& +(1 - pred_{j,k}^{\text{ML}}) \cdot FNR \cdot [(\sum_{g\in\text{gravity}} P_g \cdot R_{i,j,k}^{\text{disr}}) \cdot w_{i,j,k}^{\text{SI}}]\} \\
& = R_i^{\text{SI}} \qquad \forall i \in \mathcal{T}
\end{aligned} \tag{D.5}$$

$$\sum_{i\in\mathcal{T}}\sum_{g\in\text{gravity}}(P_g\cdot C_g^{\text{corr,maint}})\cdot maint_i + C^{\text{prev,maint}}\cdot ML_i^{\text{maint}} = C^{\text{maint}} \tag{D.6}$$

$$t_{\text{dep}} - t_{\text{op}} \leq M\cdot Canc^{trig} \tag{D.7}$$

$$shunts\cdot t_{\text{shunt}} = t_{\text{dep}} \tag{D.8}$$

$$shunts^{\text{SI}} + shunts^{\text{SO}} = shunts \tag{D.9}$$

$$\sum_{i\in\mathcal{T}} w_i^{\text{SO}} - \sum_{i=1}^{rail-1} adj_{i,i+1} = shunts^{\text{SO}} \tag{D.10}$$

$$2\cdot adj_{i,i+1}^{\text{SO}} \leq w_i^{\text{SO}} + w_{i+1}^{\text{SO}} \qquad \forall i\in[\mathcal{T},\mathcal{T}+1] \tag{D.11}$$

$$maint_i + ML_i^{\text{maint}} + dem_i \leq 1 \qquad \forall i\in\mathcal{T} \tag{D.12}$$

$$dem_i \leq (1.05 - prob_i^{\text{ML}}) \qquad \forall i\in\mathcal{T} \tag{D.13}$$

$$\sum_{i\in\mathcal{T}} dem_i \leq \sum_{l\in TYP}|diff_l| \tag{D.14}$$

$$maint_i + ML_i^{\text{maint}} + dem_i = w_i^{\text{SO}} \qquad \forall i\in\mathcal{T} \tag{D.15}$$

$$\sum_{l\in TYP} diff_l - \sum_{i\in\mathcal{T}}(w_i^{\text{SO}} - \sum_{j\in TRK}\sum_{k\in PTRK} w_{i,j,k}^{\text{SI}}) = 0 \tag{D.16}$$

$$\sum_{j\in TRK}\sum_{k\in PTRK} w_i^{\text{SO}}\cdot typ_{i,j,k} + (1-w_i^{\text{SO}})\cdot WI_{i,l} \geq w_{i,l}^{OUT} \qquad \forall i\in\mathcal{T},\forall l\in TYP \tag{D.17}$$

$$\sum_{i\in\mathcal{T}} w_{i,l}^{OUT} - WO_{i,l} = 0 \qquad \forall l\in TYP \tag{D.18}$$

$$\sum_{l\in TYP} w_{i,l}^{OUT} = 1 \qquad \forall i\in\mathcal{T} \tag{D.19}$$

$$\sum_{\hat{i}=1}^{\mathcal{T}-1} w_{\hat{i},j,k}^{\text{SI}}\cdot\sum_{\tilde{i}=\hat{i}+1}^{\mathcal{T}} w_{\tilde{i},j,k}^{\text{SI}}\cdot adj_{j,k} \geq adj_{j,k}^{\text{SI}} \qquad \forall j\in TRK,\forall k\in PTRK \tag{D.20}$$

$$\sum_{i\in\mathcal{T}}\sum_{j\in TRK}\sum_{k\in PTRK} w_{i,j,k}^{\text{SI}} - \sum_{k=1}^{PTRK-1} adj_{j,k}^{\text{SI}} = shunts^{\text{SI}} \tag{D.21}$$

$$\sum_{i\in\mathcal{T}} w_{i,j,k}^{\text{SI}} \leq 1 \qquad \forall j\in TRK,\forall k\in PTRK \tag{D.22}$$

$$\sum_{j\in TRK}\sum_{k\in PTRK} w_{i,j,k}^{\text{SI}} \leq 1 \qquad \forall i\in\mathcal{T} \tag{D.23}$$

$$w_{i,j,k}^{\text{SI}} \leq suit_{j,k} \qquad \forall i\in\mathcal{T},\forall j\in TRK,\forall k\in PTRK \tag{D.24}$$