Computers & Industrial Engineering 188 (2024) 109865

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

# Optimal management of full train load services in the shunting yard: A comprehensive study on Shunt-In Shunt-Out policies

Tommaso Bosi [a,*], Federico Bigi [b], Andrea D'Ariano [a], Francesco Viti [b], Juan Pineda-Jaramillo [b]

[a] Department of Civil, Computer Science and Aeronautical Technologies Engineering, Roma Tre University, Rome 00154, Italy
[b] Faculty of Science, Technology and Medicine (FSTM), University of Luxembourg, Esch-Sur-Alzette, L-4364, Luxembourg

ARTICLE INFO

*Keywords:*
Railway operations
Freight rail operation
Freight trains
Shunting operations
Shunting emissions
Rolling stock circulation

ABSTRACT

A key aim of the European Union is to double freight rail traffic by 2050 in order to reduce pollution emissions and alleviate congestion by shifting traffic from roads to rail networks. To accomplish this objective, it is crucial to minimize emissions and high costs associated with shunting yard operations while maintaining an acceptable level of service. This research paper introduces a new MINLP model that optimizes the shunt-in and shunt-out (SISO) operations of wagons in a shunting yard that handles full train load services. Additionally, an efficient Multi-Objective Dijkstra Algorithm (MDA) is proposed to handle simultaneous shunt-out operations in a multi-train scenario. The MINLP model takes a mesoscopic approach, aiming to minimize the number of SISO operations while satisfying strategic and tactical objectives such as wagon fleet size, operational costs, and shunting locomotive emissions. Several versions of the mathematical model are described, each employing different Shunt-In (SI) policies with varying criteria for wagon selection and strong goal orientation. The Multi-Objective Dijkstra Algorithm determines the Multi-Objective Shortest Path between mandatory shunts, considering both the time required for shunting and clustering costs. It provides information on the clusters of wagons that need to be shunted out. To assess the effectiveness of the MINLP model, real train timetables for freight trains in the Bettembourg Eurohub Sud Terminal (Luxembourg) are used, and various KPIs related to tactical and strategic objectives are evaluated. Furthermore, the performance of the Multi-Objective Dijkstra Algorithm is compared to the Shunt-Out sub-model, considering average computation time and solution quality in relation to the MINLP model. The computational results demonstrate that the criteria for wagon selection have a significant impact on the analyzed KPIs.

## 1. Introduction

Global trends in transport development show an ecological priority combined with energy efficiency. The data from the European Environment Agency (EEA) (Agency, 2021) proves how transport produces the largest of Europe's greenhouse gas emissions and is, therefore, the leading cause of air pollution in cities. Transportation today represents 27 % of the EU's total emissions, where 95 % of them come from cars, vans, trucks, and buses (i.e. road transport). In the context of green transition, freight rail transportation will play a key role. The promotion of freight rail transportation to relieve congested roads is, indeed, one of the current priorities in transport policy (Commission, 2015) (Kearns et al., 2016), as almost 78 % of goods are transported on tires. Therefore, the freight train traffic is expected to double in the next 30 years in order to help reach the carbon neutrality goal (Pagand et al., 2020). Nevertheless, freight transportation has costs that are unique to the mode, as well as logistic complexities that do not exist for road transport. A significant portion of these costs is related to operations performed inside shunting yards, namely, *shunting operations*.

### 1.1. Full train load services and shunting operations

A shunting yard is a railway facility used for the sorting, classification, and marshalling of railroad cars (Boysen et al., 2016) (insights in *Appendix C*). At intermodal stations, the shunting yard typically manages full train load services (Bohlin et al., 2016), which encompasses the organization and reconfiguration of entire trains, rather than just individual rail cars. This can include activities such as assembling new trains

* Corresponding author.
*E-mail addresses:* tommaso.bosi@uniroma3.it (T. Bosi), federico.bigi@uni.lu (F. Bigi), andrea.dariano@uniroma3.it (A. D'Ariano), francesco.viti@uni.lu (F. Viti), juan.pineda@uni.lu (J. Pineda-Jaramillo).

https://doi.org/10.1016/j.cie.2023.109865
Received 6 June 2023; Received in revised form 10 December 2023; Accepted 26 December 2023
Available online 30 December 2023

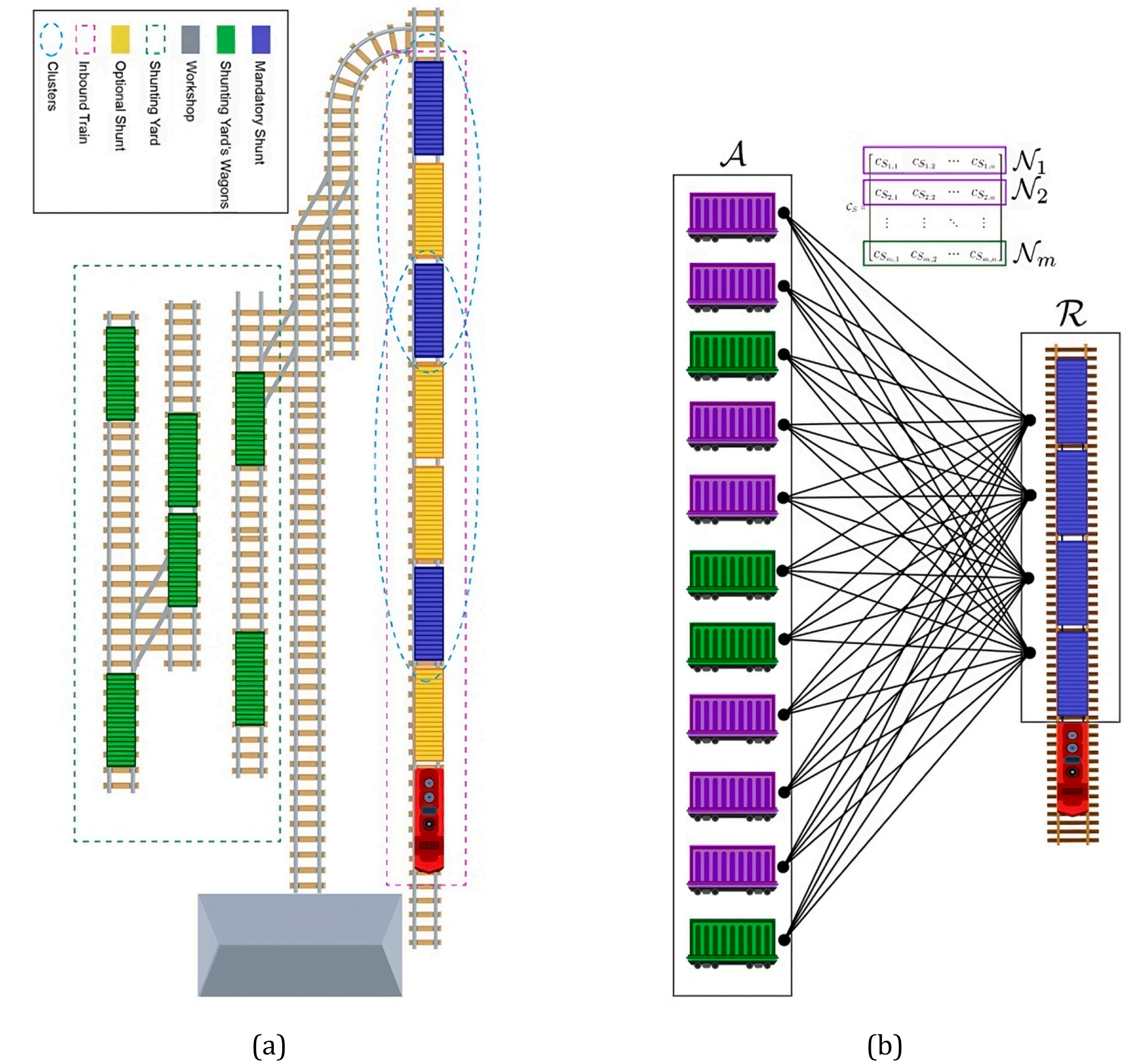


**Fig. 1.** SISO problem representation. For the Shunt-Out problem (a) the blue circles are the possible clusters of shunts and the decision falls on activating optional shunts or not. The Shunt-In problem (b) can be formulated as a Crew Scheduling Problem, where green and purple wagons are, respectively, of the Double and Simple types. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

by adding or removing rail cars from inbound trains for maintenance or demand-matching constraints (Guglielminetti et al., 2015). These yards are typically larger and have advanced infrastructures, such as cranes and container stacking areas, to support the handling of full train loads. In contrast, a shunting yard that operates car load service concentrates on individual railcars, rather than entire trains, which are sorted and reassembled into new train sets for different destinations. This may include switching railcars between tracks, moving them to different storage areas, or loading and unloading cargo (Bohlin and Hansmann, 2018) (Deleplanque et al., 2022). A shunting yard for managing full train load services can perform two types of shunt: demand shunts, which involve replacing inbound train wagons to conform to the type and number of wagons specified by the scheduled service; and maintenance shunts, which consist of moving railcars to workshops or tracks designated for inspection, repair, or maintenance. The maintenance shunt can be prompted by various factors, one of which is a mileage-based condition, and is usually done regularly to ensure that the railcars and the equipment of the shunting yard are in good condition and able to operate safely and efficiently. Regardless of the shunt type, the shunting activity is performed by shunting locomotives, which are small engines, usually diesel-driven, used to move railcars around the yard.

### 1.2. *Shunt-In Shunt-Out Operations*

In this context, we aim to optimize the shunting operations carried out in a shunting station operating full train load services (insights in Section 2.1), as currently this, and other shunting yard issues, are solved solely based on practitioners' experience. Given a timetable and a heterogeneous fleet of wagons as data input, the *Shunt-In Shunt-Out Problem* (SISO) aims to establish the selection criteria we take out wagons from

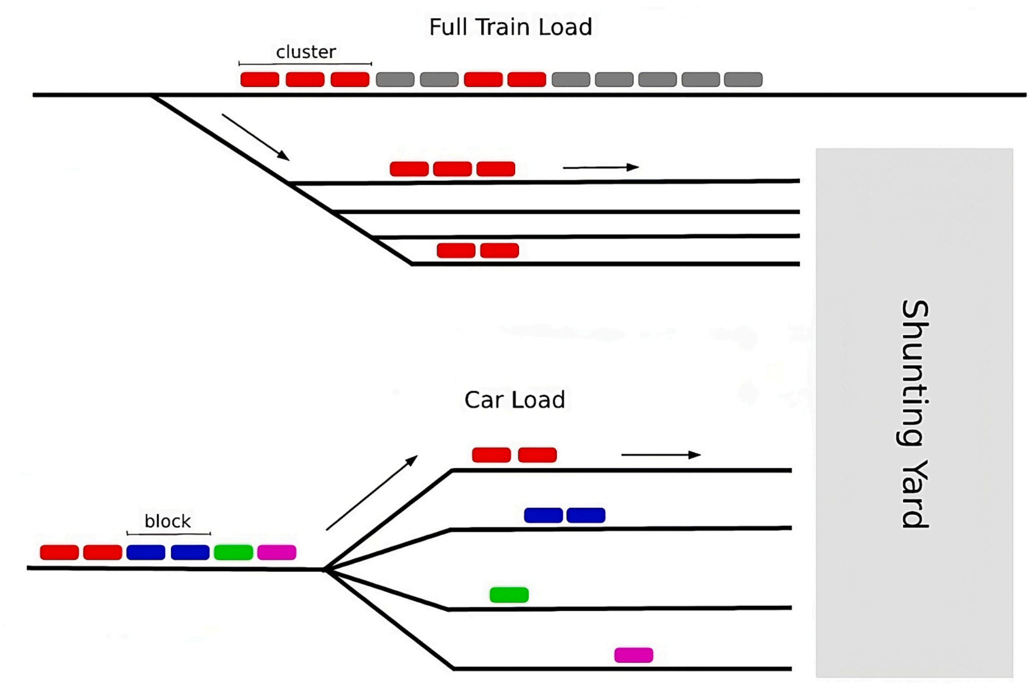


**Fig. 2.** Modelling Comparison: Shunting Differences Between Full Train Load and Car Load Services. Full train load services enable targeted shunting exclusively for clusters that necessitate it.

the inbound train due to condition-based maintenance and timetable constraints (*Shunt-Out*, SO), and we replace them with shunting yard's wagons in order to make up the outbound train (*Shunt-In*, SI). To further clarify, this problem deals with the assignment of wagons to timetabled services in order to meet requirements specified by leasing contracts and fulfil a given wagons demand, based on the composition and final destination of each departing train. The SISO problem is further bounded by the overall operational time required to shunt, which could lead to delays and train cancellations, and the shunting yard's supplies availability (Fig. 1a and 1b).

The selection criteria, also known as *shunting policy*, can be defined for both Shunt-Out and Shunt-In operations. In the first case, the policy aims to create clusters of wagons that will be removed due to demand-matching or maintenance constraints. Based on practice, a cluster of wagons, namely multiple adjacent wagons, can be considered a single shunt-out operation in terms of operational costs, regardless of the cluster size. However, it should be noted that the operational time required for shunting out the cluster is influenced by its size. Hence, a first trade-off arises between operational costs and the time taken to shunt out a cluster. In the second case, the shunt-in policy aims to identify shunting yard wagons to replace those that have been removed, in order to impact long-term KPIs. Indeed, this decision problem directly impacts both strategic and tactical objectives, such as minimizing shunting costs while complying with contractual maintenance constraints, maintaining an acceptable level of service by minimizing departure delays, avoiding cancellations due to the outbound train's deadline being exceeded, or a shortage of wagons in the shunting yard, minimizing the emission produced by diesel-driven shunting locomotives, and optimizing the management of the wagon fleet by reducing its size and the associated overhead costs.

### 1.3. MINLP model and MDA

Our research proposes a MINLP model dealing with the SISO problem for full train load services and a *Multi-Objective Dijkstra Algorithm* (MDA) for the Shunt-Out Multi-train case. These models are integrated into a Python-based, object-oriented, event-driven simulation framework. The MINLP model considers rolling stock maintenance and timetable constraints as well as a multicomponent objective function aiming for the minimization of the number of shunts performed. The objective function also includes factors such as delay penalties, SI policies, and wagon shortage prevention. The research also examines multiple SI policies, each of which is characterized by specific wagon assignment criteria. Due to their different assignment criteria, each SI policy shows pros and cons, therefore, they should be defined considering the goal that practitioners want to achieve. The MDA is designed to consider both the number of shunts performed and the time to shunt by finding the Multi-Objective Shortest Path (MOSP) between the first and last mandatory shunts of an inbound train. The MOSP gives us information on clusters of shunts activated and, therefore, on the economic costs incurred.

The main contribution of this paper is to propose a mesoscopic MINLP model that formalizes a new shunting yard issue, the Shunt-In Shunt-Out problem for full train load services, by considering both maintenance and demand shunting operations, which are typically treated separately. Additionally, the paper aims to evaluate the impact of multiple shunt-in policies on key performance indicators over a longer period. Indeed, as opposed to carload services, the nature of full train load services enables the train to retain its identity, which facilitates long-term simulations and allows for the collection of data on long-term KPIs. In addition to the MINLP model, an MDA is proposed to improve the computational time and effectively handle shunt-out operations on simultaneous inbound train arrivals, making it a valuable tool for practitioners in case of limited computational resources. Several simulations are carried out to validate the policies' usefulness, exploiting a real schedule from 2021 up to 2050 used by the Luxembourg National Railway Company with a particular focus on the Bettembourg Eurohub Sud Terminal which connects various EU countries. Luxembourg and its freight forwarding operator CFL play, indeed, a central role in Europe due to the location of its intermodal terminal. On the other hand, the MDA has been tested on a set of practical-size instances and compared with the SO sub-model results.

The rest of this paper is structured as follows: Section 2 reviews the relevant literature on shunting yard issues, full train load services, and maintenance operations; Section 3 provides a formal description of the SISO Problem and our assumptions; Section 4 describes the MINLP model for the SISO problem along with the MDA steps; Section 5 shows both the performance of each SI policy and the MDA taking into account multiple KPIs; Section 6 summarizes the conclusions and suggests where to focus future research.

## 2. Literature review

In the shunting yard, also called the classification or marshaling yard, inbound trains are disassembled and wagons are then assembled such that desired compositions of outbound trains are generated. However, these classification procedures are rather resources consuming. Indeed, shunting operations may take 10–50 % of trains' total transit time (Jaehn et al., 2015), while (Meinert et al., 2015) shows how Diesel-driven traction is still widely used in railway systems. Nevertheless, according to (Denari and Derossi, 2019), a mere 13 % of previous literature has dedicated attention to the issue of freight train maintenance. It is worth noting that, as far as we know, no study to date has ever explored the challenges associated with full train load services and shunting operations. The majority of the literature deals with the re-ordering of wagons, according to specified sequences on tracks classified by destinations (see (Deleplanque et al., 2022) (Van Den Broek et al., 2021) for surveys). The objective function is usually to minimize the number of classification tracks or classification stages. This highlights how additional research may improve the current state-of-the-art to understand and estimate the impact of freight train maintenance on emissions, costs, and delays in departure.

### 2.1. Full train and car load services

Railway freight transportation offers two distinct types of services. We focus on a type of yard operating *Full Train Load* service (Fig. 2), as opposed to *Carload* ones (Bohlin et al., 2016). The key differences, which can have an impact on the model, are outlined below:

- In full train load service, all wagons of a single train have the same origin and destination, while in the carload service trains must be split and wagons recombined to form new trains since they do not share the same origin and/or destination;
- Full trains are freight trains that function as a unified entity, efficiently transporting cargo from a departure terminal to a destination terminal without any intermediate stops for individual wagon pick-ups or drop-offs;
- In full train load services, it is not necessary to split each inbound train into blocks. Instead, only clusters of wagons requiring maintenance or demand shunting are selectively removed and replaced by shunting locomotives. This clear differentiation allows us to treat the SISO problem as distinct from the classification process.

Even though the literature has predominantly focused on car load services, it is essential to note that full train load services have garnered a positive reputation among market players, particularly for large shippers, due to their ability to efficiently transport massive volumes (Guglielminetti et al., 2015). For intermodal services, it is the most common service type, and we usually have regular full train load services connecting two intermodal terminals. One of the benefits of full train load services is the ability to efficiently handle large quantities of goods or materials. The train concept is the basis for the proposition of full train load services offering high capacity between origin and destination having direct access to rail, reasonable transit time and reliability, and acceptable price. In terms of maintenance activities, a marshaling station with full train load services can provide a centralized location for the repair and maintenance of rail cars. For example, instead of having to take individual rail cars out of service and send them to a separate maintenance facility, clusters of wagons of a full train can be brought into the marshaling station and the necessary maintenance can be carried out on the cars while they are still on the tracks. This can save time and improve efficiency by reducing the need for additional handling and transportation of rail cars.

### 2.2. Shunting yard strategic and tactical tasks

As (Boysen et al., 2016) states, the hierarchical decision problems at a classification yard focus on two major task typologies:

- *Strategic tasks* extend beyond shunting yards and encompass crucial aspects such as infrastructure investments and modifications. These tasks can be classified into two categories: superordinate strategic decisions and subordinate strategic decisions. Superordinate strategic decisions pertain to high-cost equipment investments, such as the implementation of new switch control systems, and are made at shunting yards. On the other hand, subordinate strategic decisions, such as determining the number of wagons and investing in medium-cost equipment, are also made at shunting yards (Lin et al., 2019);
- *Tactical tasks* need to be solved on a shorter planning horizon and they typically affect organizational processes and shunting policies. Super-ordinate tasks have a direct effect on network planning and sub-ordinate tasks only affect the yard organization, for instance, wagons priority rules, sorting, and blocking (Zhang et al., 2018) (Boysen et al., 2017) (Murali et al., 2016) (Bohlin and Hansmann, 2018) (Deleplanque et al., 2022).

As far as we know, most of the papers on shunting yard operations deal with tactical tasks, such as the *Wagon Classification Problem*, where the focus is on the optimal sequence of shunting steps in the marshaling area. The *Wagon Maintenance Scheduling* (WMS), configured as a sub-ordinate tactical task at shunting yards, plays a minor role in the literature. Indeed, this task is not specific to shunting yards. Nonetheless, it can be considered as a sub-problem of the *Train Makeup Problem* (TMP) (Falsafain and Tamannaei, 2019), i.e. the assignment of wagons to outbound trains. In the real world, the TMP may occur in a large number of variants depending on which particular constraints are added due to the shunting yard infrastructure architecture and how the integration into the hierarchical decision framework is elaborated.

### 2.3. Wagon maintenance

One of the most important cost drivers is train (Xu and Dessouky, 2022) and wagon maintenance (Jaehn et al., 2015), indeed, wagons spend significant portions of their downtime associated with maintenance and repair in the workshop, producing overhead (i.e. wagons pool required) and variable costs (i.e. storage costs). (Lin and Lin, 2017) explains how wagon maintenance follows a schedule driven by one of several triggers: mileage, time, or condition monitoring. While time-based methods were traditionally employed, nowadays a growing number of operators favor mileage or condition-based maintenance. This shift is primarily driven by the fact that, on average, rolling stock remains unused inside the shunting yard for 70 % of the time, resulting in additional and inefficient maintenance operations. Optimizing maintenance and demand shunting practices could significantly reduce the fleet size needed to meet the timetable requirements, thereby minimizing wagon idle time, as well as reducing fixed and storage costs.

(Luan et al., 2017) presents an integrated MILP model to optimize the *Preventive Maintenance Time Slots Problem* (PMTSs) by considering train routes, orders, and passing times at each station, as well as the work time of preventive maintenance tasks. (Herr et al., 2017) solves both the *Rolling Stock Problem* and the maintenance scheduling for passenger trains. They consider not only preventive maintenance scheduling but also a degradation level based on the distance performed by the rolling stock, namely, their objective function aims to maximize each train's useful life. (Gerum et al., 2019) develops data-driven

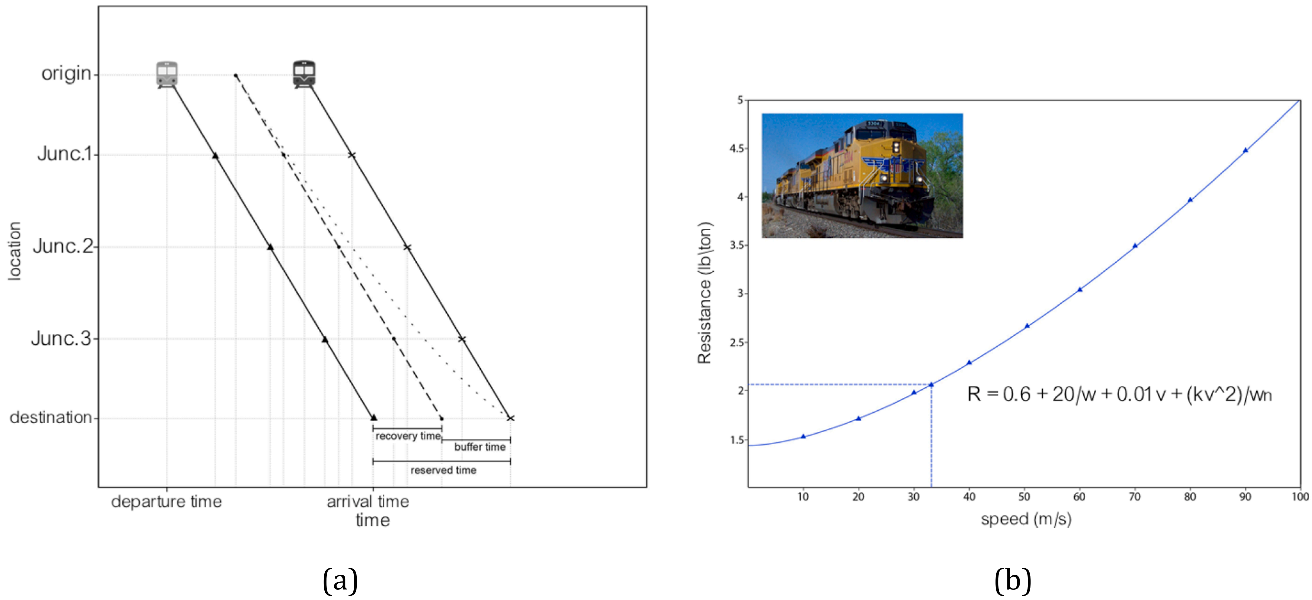


**Fig. 3.** A reserved time example (a) with two consecutive trains. This virtual additional time increases train punctuality. The Resistance curve (b) for a GE Locomotive is based on the Davis Equation.

**Table 1**
A summary of OR methods for operations in the shunting yard.

| Objective | Type of railway | Method | Main contribution | Study |
|---|---|---|---|---|
| Rolling Stock Rostering and Maintenance Scheduling | Passenger Trains | LP model | Joint optimization of rolling stock assignment and maintenance scheduling. | (Herr et al., 2020) (Herr et al., 2017) |
| Train Makeup Problem | Freight Trains: Carload services | Model | Optimizationof assignment of railcars of inbound freight trains to outbound trains considering routing. Provides basic shunting policies for optimizing operations based on cuts. | (Boysen et al., 2016) (Boysen et al., 2016) |
| Delay Minimization | Freight Trains: Carload services | MILP model + heuristic | Analysis model and heuristic for optimizing tardiness of outbound trains through the optimal humping sequence of the inbound trains. | (Jaehn et al., 2015) (Jaehn et al., 2015) |
| Loading/Unloading optimization | Freight Trains: Carload services | Time-discretize MILP | Optimization model and algorithm for the loaded train combination problem at heavy haul marshaling stations. | (Wang et al., 2022) (Wang et al., 2022) |
| Maintenance Scheduling | Passenger Trains | OR model + algorithm | Preventive maintenance time slots. | (Luan et al., 2017) (Luan et al., 2017) |
| Rolling Stock Rostering | Passenger Trains | OR model + algorithm | OR model which considers wagon arrival and departure, wagon availability, loading/unloading, and layout of operating sites. | (Chuijiang, 2020) (Chuijiang, 2021) |
| Rolling Stock Rostering | Freight Trains: Carload services | OR model | Optimization of the multistage train formation problem, which considers the shunting yard capacity and layout, freight cars location and is adapted for a rolling horizon planning. | (Bohlin et al., 2015) (Bohlin et al., 2016) |
| Rolling Stock Rostering and Maintenance Scheduling | Passenger Trains | ILP | Time windows ILP models to optimize rostering and maintenance schedules. | (Giacco et al., 2014) (Giacco et al., 2014) |
| Train Makeup Problem and Maintenance Scheduling | Freight Trains: Full Train load services | MINLP + Heuristic + Simulation | MINLP model for wagon shunt-in shunt-out operations exploiting multiple shunt-in policies to impact long-term KPIs | Present Study |

predictive maintenance scheduling policies by utilizing risk-averse prediction methods and a Markov decision process model for optimal scheduling. Nevertheless, these studies do not fully integrate the maintenance operations within the context of the shunting yard. When organizing wagon maintenance, practitioners must take into account two macro-issues: which inbound train's wagon needs shunting due to maintenance rules, and with which shunting yard's wagon must be replaced; how implementing a traffic schedule ensures traffic safety (Zagorskikh et al., 2020). Our study falls within the first decision problem, since, for most of the literature, how many and which wagons need to be replaced is an assumption (Chuijiang, 2021) (Adlbrecht et al., 2015).

### 2.4. Cascade effects on departure delays and train cancellations

Since freight rail transport has one of the lowest priorities in the railway network (Di Loreto et al., 2018), performing the strictly necessary shunting operations is one of the ways to compensate for cascade effects. Indeed, a late departure of a freight train can reduce its *reserved time* (Fig. 3a). The reserved time is given by the sum of two terms: the *recovery time*, namely, additional time included in train timetables over and above the minimum journey time necessary (Cabral et al., 2021); the *buffer time*, defined as the time to absorb the deviation from the scheduled trajectory and prevent delay propagation between consecutive trains (Jovanovic et al., 2017). Reserving larger time allows to increase train punctuality, but, on the other hand, reduces the capacity usage of the lines (Corman et al., 2017). In a context in which we expect to double the freight traffic by 2050, the optimization of shunting operations becomes relevant given that reserved times are destined to be reduced (Islam et al., 2016). This is a talking point for Europe since its railway network capacity is already at a level at which is getting harder to guarantee an acceptable service level (Navajas Cawood et al., 2016).

Moreover, due to the heavier weight carried, freight trains require wider headway to stop and reach their cruising speed compared to passenger trains. This necessity is directly connected to the recovery time, usually computed on the train's average cruising speed. The dilation of headway is clearly explained by the *Davis Equation*, stating that at higher resistance corresponds more time to stop and reach back the average speed (Cao et al., 2022) (Fig. 3b). This equation highlights how hard is for a freight train to reach back its average speed, and how crucial it is to allow it to get through its entire time window. Both railway priority rules and freight train resistance constraints express the magnitude of having a delay on schedule and the impact it may have on the entire network. Additionally, a delay in shunting operations could have an even greater impact if we consider shared-use corridors. These corridors, where passenger and freight traffic share the same rail tracks, present unique challenges due to the high cost of infrastructure construction (Wang et al., 2019). However, research has already suggested the use of consolidation policies that take into account the characteristics of both passenger and freight trains (Ursavas and Zhu, 2017). These policies could be combined with shunting policies to ensure demand-matching and maintenance activities while mitigating any potential negative effects from conflicts in resource sharing.

### 2.5. Emission related to shunting operations

Shunting operations are usually performed by diesel or hybrid locomotives, since, in the shunting yard, locomotives can't be powered by electric cables. The often very outdated shunting locomotives with combustion engines that meet the old and fairly liberal emission standards (Kurhan and Kurhan, 2018) (Daszkiewicz and Andrzejewski, 2017), further enhance the waste of fuel and energy resources, as most of it is consumed for the traction of trains. Therefore, any action that improves the efficiency of shunting operations or monitors locomotives in terms of fuel consumption and emission of harmful exhaust gases is valuable (Feo-Valero et al., 2016) (Mo et al., 2020), in order to reduce their negative environmental impact and maintenance costs (Merkisz et al., 2016) (Rymaniak et al., 2017). Increasing the *Clustering Rate*, namely, the number of wagons moved with a single shunting operation, improves the efficiency of fuel and energy management while reducing the movements per locomotive. The primary objective of the SO problem is indeed to create larger clusters.

### 2.6. Open issues and contributions

Table 1 presents a comprehensive overview of OR methods utilized in shunting yard operations. However, previous research has predominantly focused on carload service operations and short-term optimization aspects, such as minimizing departure delays, optimizing wagon sorting within trains, and locomotive assignment for outbound trains. Notably, there exists a significant research gap concerning the unique challenges associated with managing shunting operations for full train load services (e.g. SISO problem). Our study aims to bridge both a practical knowledge gap, as issues from full train load services have not been addressed in the literature, and a theoretical gap, as the shunt-out and shunt-in problem can be formulated and optimized using operations research tools. Additionally, it is worth mentioning that prior studies have often treated maintenance as a separate concern, routing, at most, wagons to workshops for maintenance purposes.

This paper addresses three major open issues, listed here below:

- The current state-of-the-art methods, as demonstrated in studies like (Chuijiang, 2021), assume to know which wagons must be removed from the inbound train and added to the outbound train. Usually, practitioners take this choice based on a single parameter, namely, the time to shunt required. A more realistic approach should take into account multiple indicators such as the mileage performed by each wagon (strictly related to leasing contracts and financial penalties);
- Previous studies, such as (Giacco et al., 2014), primarily focus on minimizing operational time and/or cost in the short term (daily) by considering shunting yard operating carload services. The train was never seen as a single entity itself which is why it is often neglected the long-term impacts of shunting operations. This research takes a different approach by performing a long-term analysis of shunting yard handling full train load services, including the estimation of the wagon fleet and the monitoring of the shunting yard capacity for a multi-year scenario;
- To the best of our knowledge, there are no existing methodologies that incorporate maintenance constraints into the optimization of shunting operations. Typically, the maintenance scheduling problem is treated as a separate, independent tactical problem. However, this approach can lead to sub-optimal solutions that do not fully consider key performance indicators such as the wagon fleet size (Giacco et al., 2014).

Our contributions to the state-of-the-art are as follows:

- We propose a novel problem for shunting yards that handles full train load services, called the Shunt-In Shunt-Out Problem, and demonstrate its significant impact on various short-term and long-term key performance indicators. The Shunt-In Shunt-Out problem aims to determine which wagons should be removed and replaced in order to satisfy both condition-based maintenance and demand-matching constraints. Current literature assumes to have direction on which wagons should be shunted and mostly addresses carload services problems like the Classification problem;
- We propose a mesoscopic approach for the shunting operation that simultaneously satisfies demand-matching constraints defined by timetabled services and adheres to the maintenance schedule. Traditionally, these two plans are treated as separate and mutually exclusive, our approach aims to find an optimal solution that comprehensively considers both constraints;
- We present a mixed-integer nonlinear programming (MINLP) model to solve the SISO problem that takes into account maintenance rules, demand matching constraints, and a multi-component objective function. The model aims to minimize the number of clusters of shunts activated and the potential departure delay caused by the time required for shunting. Additionally, it considers factors such as the shunt-in policy applied and the shunting convenience cost to ensure the feasibility of the shunting yard capacity;
- We present and evaluate three Shunt-In policies, which are implemented as additional terms in the MINLP objective function. Each SI policy exploits a wagon selection criteria and has been shown to have a significant impact on long-term key performance indicators, such as the wagon fleet size, emissions from shunting locomotives, delays, and train cancellations;
- We show and compare a Multi-Objective Dijkstra Algorithm (MDA) as a faster alternative to the SO sub-model. The MDA generates sub-optimal solutions by determining the Multi-Objective Shortest Path (MOSP) between the first and last mandatory shunts, providing insight into the clusters of shunts activated;
- We conduct long-term analysis on rail key performance indicators, utilizing the unique characteristics of full train load services that enable the identification and tracking of long-term attributes associated with individual trains. On the contrary, traditional studies focus on short-term KPIs for individual wagons since they typically optimize shunting yard operating carload services.

## 3. Problem description

The SO problem can be subjected to several constraints related to maintenance rules, operational costs, seasonal wagon demand, and so forth. For instance, since the most substantial part of a SO operation's cost lies in the make-ready stage of the shunting locomotive, two or

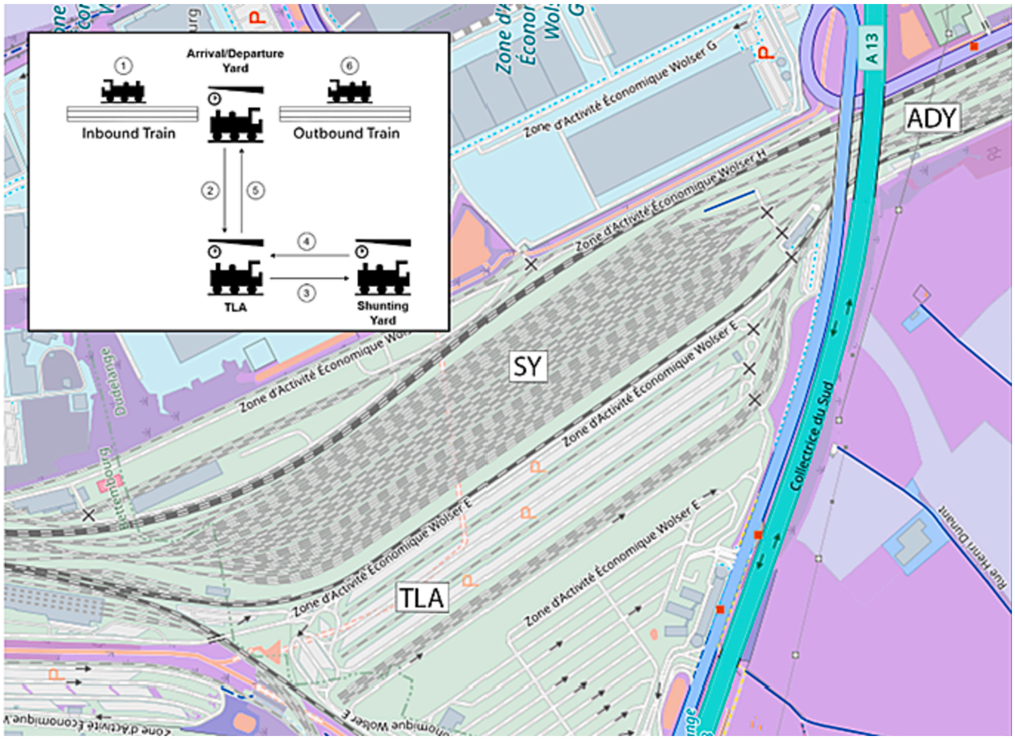


**Fig. 4.** The event flowchart concerning the SISO Problem.

more adjacent wagons requiring to be shunted out still produce a unit cost. This means we are interested in creating *clusters* of shunts by activating *optional* shunts, namely, shunts triggered neither by maintenance nor demand constraints but to bring clustering advantages. Nevertheless, the time to shunt out a cluster is equal to the number of its wagons multiplied by the unit time cost (in this study this is assumed to be a fixed time of 15 min). Deciding to create a cluster too large by activating too many optional shunts, may lead to a delay or even cancellation. A mathematical model can help practitioners decide when to cluster or not. Moreover, a feasible solution for the SO problem must assure both that the outbound train's composition is fulfilled and that a wagon is not moved to maintenance unless its current mileage falls within the lease contractual range. The whole process must take into account the resource level of the shunting yard and workshop, in order to avoid a quick shortage of wagons.

The SI problem is complementary to the SO one, as it aims to minimize time and economic costs by replacing each shunted-out wagon with shunting yard's *suitable* wagons. A suitable wagon must both have enough residual mileage to perform the outbound train's next trip and be of the type required by the new composition. The basic problem consists of replacing a shunted wagon with one having the lower shunt in time inside the shunting yard. Nevertheless, referring to a single parameter is a short-sighted approach. For instance, the shunted in wagons may have a mileage already close to their contract's threshold. That way when the train comes back to the station it will require another SO operation. This additional cost could have been avoided with a multi-component objective function considering a shunt in policy focused on economic costs.

The event flowchart concerning the SISO problem is described in Fig. 4. As soon as a given inbound train $\mathscr{T}$ enters the terminal, it is moved into the queue of the Arrival/Departure yard (ADY), where its wagons are inspected. Here, it has to wait until the Train Unloading/Loading area (TLA) is freed. The arrival and departure inspections require about 35 min. In the TLA, $\mathscr{T}$ starts the unloading operations based on the timetable, each of which requires 7,5 min. If $\mathscr{T}$ requires shunting operations due to demand matching or maintenance rules, it is moved into the shunting queue. Each SO operation requires 15 min, while the time to shunt in wagons from the shunting yard depends on the specific wagon, as it is related to its position and the maneuvers required to pull it out. Regardless of the shunt type, each shunting operation cost is estimated to be € 350 in this study and can only be performed if the shunting yard is not busy executing other shunting operations. The shunting operation cost is a weighted average of four primary factors: the shunting locomotive assignment and start-up costs, the crew scheduling costs, and the average fuel consumed per hour. This is the reason behind the constant and not statistical nature of this parameter, regardless of the cluster size. The maintenance for a single wagon is assumed to cost € 10,500 and the maintenance range is set from 150.000 km up to 172.500 km, based on the practice. Wagons that are shunted out due to demand matching could perform other trips before going to maintenance, therefore, they are sent inside the shunting yard, while the ones shunted out due to maintenance are sent to the workshop. In the workshop, the wagon remains down for 3 days. Once the 3 days have passed, the wagon is available again inside the shunting yard. Each wagon shunted out is then replaced with a wagon from the shunting yard through SI operations. Once all the SISO operations are performed on $\mathscr{T}$, the train is moved again to the TLA queue for the loading operations and then to the ADY ready for departure, unless it is canceled.

### 3.1. Assumptions

Several assumptions are considered according to practice:

- There are two types of wagon, *Simple* and *Double*, with different physical and contractual characteristics such as the length, capacity, maintenance range, and so on;
- The operational time to shunt out a wagon is on average 15 min, while the one to shunt in a wagon stored in the shunting yard changes from wagon to wagon;
- While a cluster of shunts, namely, two or more adjacent wagons requiring SO operations, is associated with a single economic cost, its

temporal cost is equal to the unitary time to shunt out multiplied by the number of wagons inside the cluster;
- There are mainly two types of SO operations, the *mandatory* and the *optional* ones. The first type is performed due to maintenance rules or demand-matching constraints. The second type is performed between successive mandatory shunts to create clusters and reduce shunting costs;
- The *maintenance* and optional shunt can be performed only when the wagon's *virtual mileage* ranges between the minimum and maximum mileage or exceeds the maximum mileage defined by the corresponding leasing contract. The virtual mileage is equal to the kilometers covered by the wagon *i*-th once it has performed the outbound train's next trip;
- 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 described in *Appendix C*;
- If the outbound train's departure delay exceeds three hours, the train is considered canceled;
- The inbound and outbound trains' sizes can be different by length, namely, by number of wagons;
- The SI operational time associated with each shunting yard's wagon is a stochastic value comprehending all the shunting operations times to pull out the wagon from the shunting yard. The distribution of this operational time is described in *Appendix C.2.1*;
- The demand matching does not consider a specific sequence of wagon types on the outbound train. Therefore, the only constraint concerns the mandatory number of each wagon type stated by the timetable.

## 4. Methodology

This section introduces the methodology and is structured as follows: Subsection 4.1 describes the nomenclature used to model the SISO problem; Subsection 4.2 details the pre-processing function developed to handle the assumption of the different sizes between inbound and outbound trains; Subsection 4.3 explains in-depth a basic version of the MINLP model, with its constraints and objective function, where it is applied as an example one of the SI policy, named MIN; Subsection 4.4 presents several SI policies translatable as different versions of the MINLP objective function; Subsection 4.5 deepens the shunting convenience cost used to avoid a quick shortage of the shunting yard's wagon pool; Subsection 4.6 illustrates the idea behind the MDA and each step performed by the algorithm.

### 4.1. Nomenclature

#### 4.1.1. Sets

- $\mathscr{T}$, set of inbound train's wagons
- $S$, set of shunting yard's wagons
- $K$, set of wagon types

#### 4.1.2. Parameters

- $a_T$, integer value expressing the inbound train's arrival time
- $d_T$, integer value expressing the outbound train's planned departure time
- $dd_T$, integer value expressing the outbound train deadline before its cancellation
- $ts$, integer value expressing the time required by a shunting locomotive to perform a single SO operation
- $r_T$, integer value expressing the kilometers the outbound train will perform during the next trip
- $m_i$, integer value expressing the current mileage of the wagon *i*-th on the inbound train $\mathscr{T}$
- $m_{S_j}$, integer value expressing the current mileage of the wagon *j*-th inside the shunting yard $S$
- $m_{max_i}$, integer value expressing the max mileage before the maintenance of the wagon *i*-th on the inbound train $\mathscr{T}$ based on the leasing contract
- $ms_{max_j}$, integer value expressing the max mileage before the maintenance of the wagon *j*-th inside the shunting yard $S$ based on the leasing contract
- $m_{min_i}$, integer value expressing the min mileage to shunt the wagon *i*-th on the inbound train $\mathscr{T}$
- $type_i$, integer value equal to 1 or 2 expressing the type of the wagon *i*-th on the inbound train $\mathscr{T}$
- $type_{S_j}$, integer value equal to 1 or 2 expressing the type of the wagon *j*-th in the shunting yard $S$
- $code_i$, integer value expressing the unique code associated with the wagon *i*-th on the inbound train $\mathscr{T}$
- $code_{S_j}$, integer value expressing the unique code associated with the wagon *j*-th inside the shunting yard $S$
- $type_r$, integer value equal to 1 or 2 expressing the wagon type on the outbound train that must rise due to the demand, compared to the inbound train $\mathscr{T}$
- $rise$, integer value expressing the surplus of wagons of the type $type_r$ in the outbound train new composition, compared to the inbound train $\mathscr{T}$
- $n_{m_i}$, float value expressing the *virtual rate* of the wagon *i*-th inside the inbound train $\mathscr{T}$, namely, the ratio between the kilometers covered once the outbound train's next trip has been performed (virtual mileage) and the max mileage $m_{max_i}$
- $n_{ms_j}$, float value expressing the *virtual rate* of the wagon *j*-th inside the shunting yard $S$, namely, the ratio between the kilometers covered once the outbound train's next trip has been performed (virtual mileage) and the max mileage $ms_{max_j}$
- $c_{u_i}$, float value expressing the shunting convenience cost used as a preemptive tool to avoid infeasibility of the shunting yard $S$
- $c_{s_{i,j}}$ float value expressing the temporal cost to replace the wagon *i*-th on the inbound train $\mathscr{T}$ with the wagon *j*-th inside the shunting yard $S$, normalized through the Min-Max normalization
- $M$, a Big-M coefficient
- $l_I$, integer value expressing the number of wagons on the inbound train $\mathscr{T}$, namely the"length" of the inbound train
- $l_O$, integer value expressing the number of wagons planned for the outbound train, namely the"length" of the outbound train

#### 4.1.3. Decision variables

- $ad_T$, integer value expressing the actual departure time of the outbound train once all the shunting operations are performed
- $code_{O_i}$, integer value expressing the unique code associated with the wagon *i*-th on the outbound train
- $\beta$, 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
- $\alpha$, float value equal to $1-\beta$
- $y_i = \begin{cases} 1, ifonthewagoni-thonthe \\ inboundtrainTamaintenance \\ oroptionalshuntisperformed \\ 0, otherwise \end{cases}$
- $x_{i,k} = \begin{cases} 1, ifonthewagoni-thonthe \\ inboundtrainTademand \\ shuntisperformedandis \\ replacedbyashuntingyard \\ wagonoftypek \\ 0, otherwise \end{cases}$
- $z_{i,j} = \begin{cases} 1, ifonthewagoni-thonthe \\ inboundtrainTisreplacedby \\ thewagonj-thinsidethe \\ shuntingyardS \\ 0, otherwise \end{cases}$

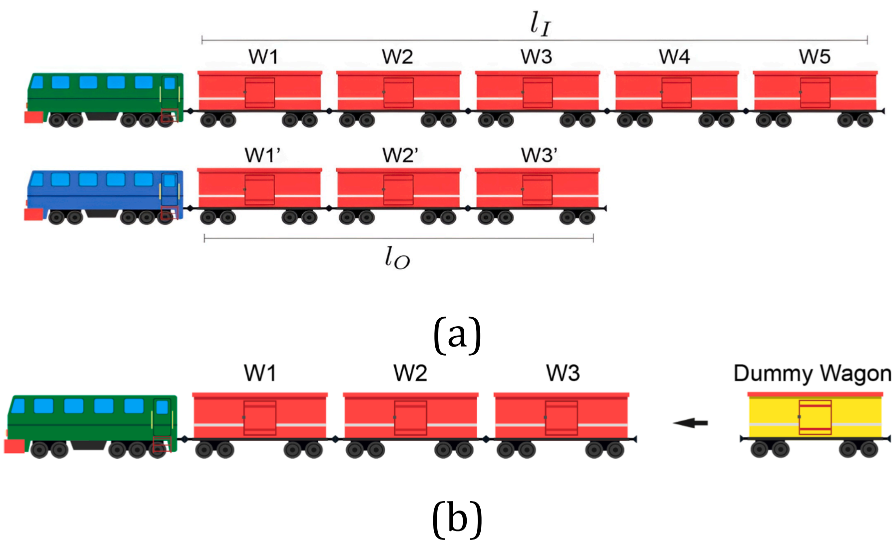


**Fig. 5.** Example of the first variable size case study, where green and blue trains are, respectively, the inbound and outbound trains. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

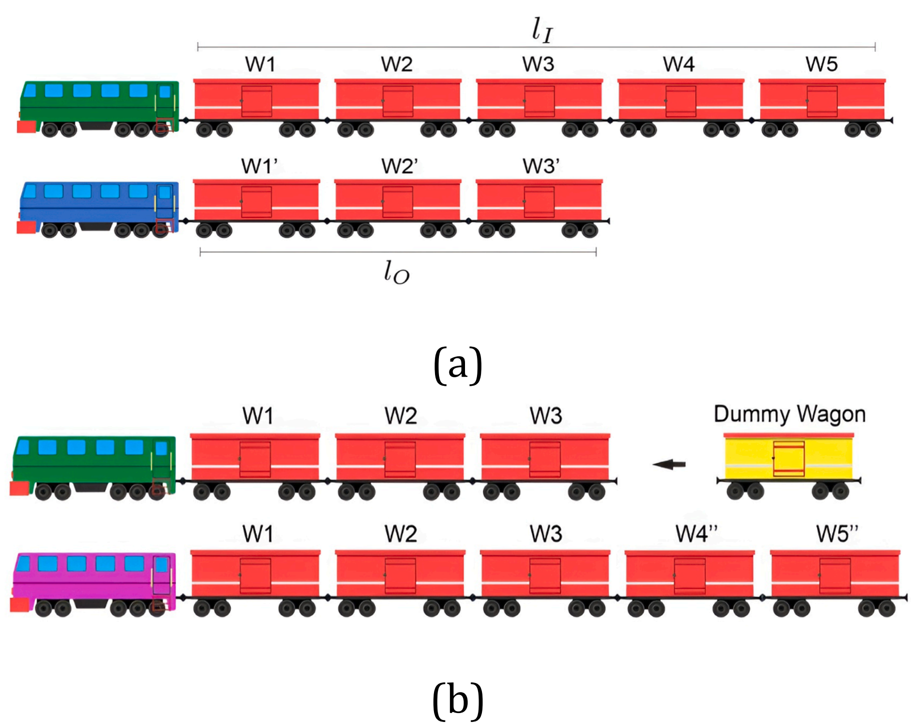


**Fig. 6.** Instance for the second variable size case study, where the green and blue trains are, respectively, the inbound and outbound trains and the purple train is the *virtual* inbound train. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

- $\gamma_i = \begin{cases} 1, if on the wagon i - th on the \\ inbound train T is shunted \\ out, regardless of the shunt type \\ 0, otherwise \end{cases}$
- $\sigma_1 = \begin{cases} 1, if dd_T \geq ad_T > d_T \\ 0, if d_T \geq ad_T \end{cases}$
- $\sigma_2 = \begin{cases} 1, if ad_T > dd_T \\ 0, if ad_T \leq dd_T \end{cases}$
- $\sigma_3 = \begin{cases} \dfrac{ad_T - d_T}{dd_T - d_T}, if \sigma_1 = 1 \\ 0, if \sigma_1 = 0 \end{cases}$
- $adj_{i,i+1} = \begin{cases} 1, if both the wagon i - th \\ and its adjacent wagon \\ i+1 - th on the inbound \\ train T must be shunted \\ out \\ 0, otherwise \end{cases}$

### 4.2. The resizing pre-processing function

As described in Sub-section 3.1, the inbound and outbound train's number of wagons could be different. Following it is explained the resizing pre-processing function that allows the MINLP model to handle a non-fixed length. We have two possible cases: the inbound train has several wagons, respectively, higher or lower than the outbound train. Moreover, one of the conditions to force the model to seek clustering, and then economic savings, is to have at least two mandatory shunts, whether these are demand or maintenance shunts. By exploiting this behavior, what we do is, based on the first or second case, to cut or add wagons on the inbound train and then insert a *dummy* wagon as a mandatory shunt (specifically a maintenance shunt).

That way the model will be forced to consider that shunt and try to cluster. In the first case, the inbound train's length $l_I$ is longer than the outbound train one $l_O$, as you can see in Fig. 5a. Therefore, we apply a cut from the $l_O$-th wagon on the inbound train, and we move the cut's wagons inside the shunting yard. The cut is counted as a single cost as it is a cluster of shunts. Then, we add the dummy wagon to position $l_O+1$ on the inbound train, and we activate a maintenance shunt for the latter by switching on the binary variable $y_{l_O+1}$. Similarly, in the second case, the inbound train's length $l_I$ is shorter than the outbound train one $l_O$, as you can see in Fig. 6a. Therefore, we append a number $l_O-l_I$ of wagons to the inbound train (based on the SI policy), treating them as a cluster of shunts. Then, we virtually add the dummy wagon to position $l_I+1$ with its $y_{l_I+1}$ activated. The time to shunt associated with the dummy wagon is computed based on the case: for the first case, it is equal to *ts* multiplied by the number of wagons cut; for the second case, it is equal to the sum of the shunt-in times of each wagon added. That way, the model will be forced to take into account the dummy wagon shunted and try clustering to minimize the economic costs.

### 4.3. Mathematical model

#### 4.3.1. Objective function

The MINLP model designed for the SISO problem is structured to minimize an objective function comprising three key terms. These terms serve the following purposes: minimize the number of clusters formed and, consequently, the operational costs of shunting; assess and minimize the delay produced by shunting operations; consider both the SI policy applied, along with its decision criteria, and the shunting yard state in terms of available wagons. To appropriately reflect a decreasing priority on delay, the chosen shunting policy, and the shunting yard state in the context of the clustering problem, weight coefficients $\frac{\mathcal{T}}{2}$ and $\frac{\mathcal{T}}{4}$ have been assigned to the last two terms of the objectives These coefficients are determined based on guidance from industry practitioners. Nevertheless, these weights can be customized to suit the specific requirements of the problem at hand. The clustering assumption is implemented to minimize the number of shunting operations by maximizing the width of clusters, thereby reducing the overall operational cost. Nonetheless, it is crucial to acknowledge that a broader cluster also extends the time needed for shunting out the cluster and bringing in wagons from the shunting yard, simultaneously increasing the risk of rendering the shunting yard infeasible.

$$\sum_{i\in\mathcal{T}}\gamma_i-\sum_{i=1}^{|\mathcal{T}|-1}adj_{i,i+1} \quad (4.1)$$

The initial term (4.1) within the objective function encapsulates the effective count of shunting operations, factoring in the clustering assumption. It represents the aggregate number of wagons shunted out, encompassing maintenance, optional, or demand shunts, while subtracting the activated adjacencies. This approach allows us to tally the clusters of shunts activated, adhering to the assumption that considers two or more adjacent wagons shunted out as a unified economic cost. Notably, shunting out operations are not subject to multiplication by a weight. This decision stems from the normalization of the other objective function terms, ensuring that they are all comparable in terms of the number of wagons involved.

$$\frac{|\mathcal{T}|}{2}\overbrace{(\sigma_2+\sigma_3)}^{\mu(ad_T)} \quad (4.2)$$

The second term (4.2) delineates the penalty associated with a diminished service level, attributed to the potential departure delay $\mu(ad_T)$ of the outbound train. To address the non-linear nature of $\mu(ad_T)$, three distinct binary variables, denoted as $\sigma_i$, have been introduced to govern the constraints (4.4)-(4.13). If $ad_T$ is smaller than $d_T$, $\sigma_2$ and $\sigma_3$ will be both equal to zero, and $\mu(ad_T)$ will be equal to zero as well; if $ad_T$ ranges between the outbound train's planned departure time $d_T$ and deadline $dd_T$, $\sigma_2$ will be equal to zero, while $\sigma_3$ will be equal to $\frac{ad_T-d_T}{dd_T-d_T}$, as well as $\mu(ad_T)$. A more in-depth analysis of the $\mu(ad_T)$ function is given in *Appendix C*. $\mu(ad_T)$ is multiplied by a weight proportionate to the number of wagons on the inbound train. It is noteworthy that this weight cannot be directly correlated with the number of wagons shunted out, as doing so would introduce a non-linear term and conflict with the clustering assumption. Such an approach would incentivize the solver to consistently cluster, as the clustering advantage and shunting time would become intertwined. By introducing the penalty, the solver is compelled to activate optional shunts only when the corresponding cluster avoids causing an excessive departure delay.

$$\frac{|\mathcal{T}|}{4}\left(\sum_{i\in\mathcal{T}}\sum_{j\in S}(\alpha W_1+\beta W_2)z_{i,j}+\sum_{i\in\mathcal{T}}y_i c_{u_i}\right) \quad (4.3)$$

In accordance with practice priority, the last term (4.3) is weighted with a halved value compared to the second term. This term considers two sub-terms: the SI policy applied, whose general structure is extensively detailed in Section 4.4, and the shunting convenience costs $c_{u_i}$. The SI policy's general structure consists of two components: $W_1$ potentially reflecting a temporal cost such as the time required for shunting-in, and $W_2$, representing a maintenance-related parameter like the virtual rate. $W_1$ and $W_2$ are then multiplied by the respective percentage of operational time needed to shunt out and shunt in left before the outbound train's deadline, as expounded in Section 4.4. This design ensures that if the time remaining after shunting out is insufficient ($\alpha$ higher than $\beta$), the model will prioritize wagons within the shunting yard with lower temporal costs rather than those dictated by the applied policy. Conversely, if $\beta$ exceeds $\alpha$, the model will adhere to the policy's guidance for shunting in wagons. The second sub-term functions as a preemptive measure to curb excessive optional shunts and prevent an impractical state of shunting yard capacity. With respect to the function of the value $c_{u_i}$ described in Section 4.5, this term serves as a crucial factor in compelling the model to carefully evaluate optional shunts. This precaution is necessary to mitigate the risk of depleting the shunting yard's wagon supply rapidly. Consequently, if the current mileage of the $i$-th wagon in the outbound train is low, and the number of wagons in the shunting yard is below a certain threshold, the $c_{u_i}$ cost escalates, prompting the model to refrain from shunting out the wagon.

#### 4.3.2. Time constraints

These constraints pertain to the delay function $\mu(ad_T)$ detailed in *Appendix C*. Predicated on the value of $ad_T$, these constraints determine whether the outbound train is punctual, delayed, or faces cancellation due to surpassing the stipulated deadline. This determination is facilitated through the utilization of three temporal variables, denoted as $\sigma_i$, as elucidated in Subsection 4.1. These variables contribute a delay penalty to the objective function.

$$ad_T \leqslant d_T + \sigma_1 M + \sigma_2 M \quad (4.4)$$

$$dd_T + (1-\sigma_1)M \geqslant ad_T \quad (4.5)$$

$$ad_T > d_T - (1-\sigma_1)M \quad (4.6)$$

$$ad_T \leqslant dd_T + \sigma_2 M \quad (4.7)$$

$$ad_T > dd_T - (1-\sigma_2)M \quad (4.8)$$

$$\sigma_1 + \sigma_2 \leqslant 1 \quad (4.9)$$

$$\sigma_3 \leqslant \sigma_1 M \quad (4.10)$$

$$\sigma_3 \geqslant \frac{a_T-d_T}{dd_T-d_T}\sigma_1 \quad (4.11)$$

$$\sigma_3 \leqslant \frac{ad_T - d_T}{dd_T - d_T} + (1-\sigma_1)M \quad (4.12)$$

$$\sigma_3 \geqslant \frac{ad_T - d_T}{dd_T - d_T} - (1-\sigma_1)M \quad (4.13)$$

Constraints (4.4)-(4.6) ensure the fulfillment of $\sigma_1$ conditions, indicating that if both $\sigma_1$ and $\sigma_2$ are equal to 0, the outbound train must be on time. Conversely, if $\sigma_1$ equals 1, then $ad_T$ must fall within the range from $d_T$ (not included) to $dd_T$. On the other hand, constraints (4.7) and (4.8) express the conditions for $\sigma_2$, saying that if $\sigma_2$ is 0, $ad_T$ has not yet reached the outbound train's deadline; otherwise, the outbound train faces cancellation. While constraint (4.9) establishes a connection between $\sigma_1$ to $\sigma_2$ by ensuring their mutual exclusivity, constraints (4.10)-(4.13) establish the relationship between $\sigma_1$ and $\sigma_3$, delineating the conditions for $\sigma_3$. Specifically, if $\sigma_1$ is 0, then $\sigma_3$ must also be 0. Conversely, if $\sigma_1$ equals 1, $ad_T$ falls within the range of $d_T$ to $dd_T$, and $\sigma_3$ is calculated as $\frac{ad_T-d_T}{dd_T-d_T}$.

$$a_T + \sum_{i\in\mathcal{T}} \gamma_i ts + \sum_{i\in\mathcal{T}}\sum_{j\in S} c_{s_{ij}} z_{i,j} = ad_T \quad (4.14)$$

Constraint (4.14) establishes that $ad_T$ is equivalent to the arrival time of the inbound train, plus the cumulative time needed to execute both shunt out and shunt in operations.

$$\frac{(a_T + \sum_{i\in\mathcal{T}} \gamma_i ts)}{dd_T} = \alpha \quad (4.15)$$

$$1-\alpha = \beta \quad (4.16)$$

Incorporating constraints (4.15) and (4.16) into the model introduces nonlinearity, as these constraints define the values of α and β, as explained in Section 4.1. The parameter α s restricted to a range from 0 to 1, derived by summing the arrival time and the duration needed for shunting out, then dividing this total by the outbound train's deadline.

**SO Constraints.** The subsequent constraints guarantee the execution of shunting-out operations in accordance with specified assumptions. These assumptions take into account the maximum and minimum mileage for each wagon, employing the big-M method.

$$y_i \geq \frac{m_i + r_T}{m_{max_i}} - 1 - \left(\sum_{k\in K} x_{i,k}\right) M \quad \forall i \in \mathcal{T} \quad (4.17)$$

$$y_i \leq \left(1 - \sum_{k\in K} x_{i,k}\right) \frac{m_i + r_T}{m_{min_i}} \quad \forall i \in \mathcal{T} \quad (4.18)$$

Constraints (4.17) and (4.18) outlined above encapsulate three specific conditions: (i) if the next trip exceeds the maximum mileage, denoted as $m_{max_i}$, then the wagon $i$-th must be shunted out, leading to $y_i$ being set to 1; (ii) the wagon $i$-th should only be shunted out if it surpasses the required maximum mileage $m_{max_i}$, otherwise, $y_i$ is equal to 0; and (iii) in the scenario where a demand shunt is already activated for the $i$-th wagon, performing either a maintenance or an optional shunt is not allowed.

$$y_i \leq \frac{m_i + r_T}{m_{max_i}} \quad \forall i \in \mathcal{T} \quad (4.19)$$

Alternatively, constraint (4.19) should be taken into account if the model is required to activate mandatory shunts only when the maximum mileage $m_{max_i}$ is exceeded.

$$\sum_{i\in\mathcal{T}:type_i \neq type_r} x_{i,type_r} = rise \quad (4.20)$$

$$\sum_{k\in K} x_{i,k} = 0 \quad \forall i \in \mathcal{T} : type_i = type_r \quad (4.21)$$

$$\sum_{i\in\mathcal{T}} x_{i,k} = 0 \quad \forall k \in K : k \neq type_r \quad (4.22)$$

Constraints (4.20)-(4.22) ensure the adherence to the new composition for the outbound train. Specifically, the combined value of $x_{i,k}$ for $k$ differing from the designated type $type_r$ (representing the wagon type that needs to increase on the outbound train) must equal the variable *rise*. This variable signifies the additional wagons of type $type_r$ required by the new composition. Furthermore, the sum of $x_{i,k}$ for $type_r$ must be 0. This condition guarantees that the contribution of type $type_r$ does not alter the composition of the other wagon types in the process.

$$|type_{O_i} - type_i| = p_i \quad \forall i \in \mathcal{T} \quad (4.23)$$

Under specific conditions, operational practices may necessitate a systematic arrangement for the departure train. This implies that for each wagon position, a specific type is mandated, as stated by the timetable. To articulate this requirement, constraints (4.20)-(4.22) need to be substituted with constraint (4.23), where $type_{O_i}$ signifies the required wagon type for position $i$ on the outbound train. Additionally, the binary variable $p_i$ equals 1 when the inbound train's wagon $i$ is shunted out due to demand matching. However, constraint (4.23) is only valid when there are exactly two wagon types.

**SI Constraints.** Constraints (4.24)-(4.27) allow the model to perform the shunt-in operations.

$$\sum_{j\in S:types_j = type_i} z_{i,j} = y_i \qquad \forall i \in \mathcal{T} \quad (4.24)$$

$$\sum_{j\in S:types_j \neq type_i} z_{i,j} = x_{i,type_r} \qquad \forall i \in \mathcal{T} \quad (4.25)$$

$$z_{i,j} \leq 2 - \frac{ms_j + r_T}{ms_{max_j}} \qquad \forall i \in \mathcal{T}, \forall j \in S \quad (4.26)$$

$$\sum_{i\in\mathcal{T}} z_{i,j} \leq 1 \qquad \forall j \in S \quad (4.27)$$

Constraints (4.24) and (4.25) force the model to activate $z_{i,j}$ with the appropriate type $j$. This ensures that wagons shunted out with maintenance or optional shunt are replaced by shunting yard wagons of the same type, while those shunted out with demand shunt are replaced by wagons of the opposite type. Conversely, constraints (4.26) and (4.27) guarantee two crucial conditions: first, that no wagon with insufficient residual mileage for the next trip $r_T$ will replace the inbound train's wagons, and second, that a single shunting yard wagon will not replace more than one inbound train's wagon. These constraints delineate the feasible region where the applied SI policy will select shunting yard wagons.

**Adjacency Constraints.** The following constraints enable the model to adhere to the clustering assumption, ensuring that a unit cost is assigned to a cluster of adjacent wagons shunted out.

$$\sum_{k\in K} x_{i,k} + y_i = \gamma_i \qquad \forall i \in \mathcal{T} \quad (4.28)$$

$$2adj_{i,i+1} \leq \gamma_i + \forall i = 1, \cdots, |\mathcal{T}| - 1 \quad (4.29)$$

Constraints (4.28) and (4.29) a ensure that when two or more wagons on the inbound train are shunted out, the corresponding adjacency variables will be activated and included in the objective function. This is achieved by consolidating the demand, optional, and maintenance shunts performed into a single variable $\gamma_i$. Furthermore, the activation of $adj_{i,i+1}$ is enforced only when both $\gamma_i$ and $\gamma_{i+1}$ are equal to 1.

$$\sum_{j\in S} z_{i,j} code_{S_j} + (1-\gamma_i) code_i = code_{O_i} \qquad \forall i \in \mathcal{T} \quad (4.30)$$

In conclusion, constraint (4.30) is an optional but beneficial addition for problem optimization. While not strictly necessary, it proves useful

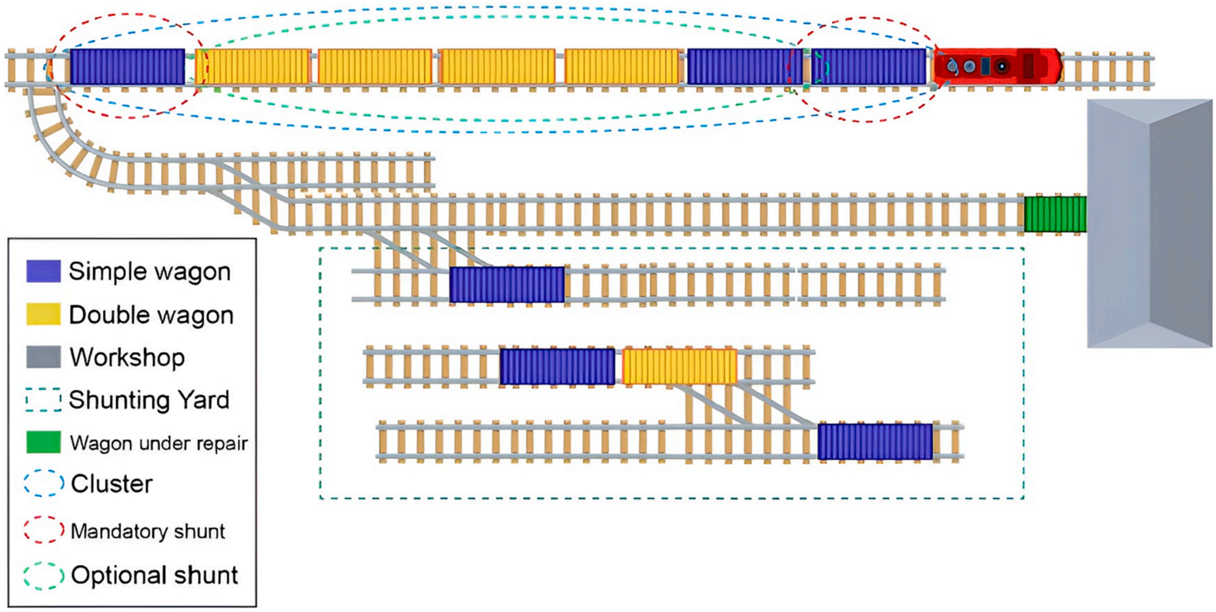


**Fig. 7.** Inbound train instance explaining the wrong behaviour of an objective function without a pre-emptive tool term. The red and green circles identify, respectively, mandatory and optional shunts that will be performed. As we can see, the shunting yard can not replace all the wagons of Double type, thus the solution is unfeasible. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in maintaining a record of wagon codes that will be on the outbound train after completing all the SO and SI operations. If wagon $i$-th is replaced by activating $z_{i,j}$, this constraint links the code of wagon $j$-th to position $i$-th; otherwise, if no replacement occurs, the code of wagon $i$-th remains unchanged.

### 4.4. Shunt-In policies

In this section, we introduce several SI policies, serving as add-ons to the multi-component objective function (4.3). For a comprehensive understanding of both the SI problem and the distribution of costs $c_{S_{i,j}}$, refer to *Appendix C* for detailed insights.

#### 4.4.1. Policy modelling

The foundational version of the SI model considers only the time cost associated with relocating a wagon from position $i$-th to position $j$-th. However, this approach may be limiting from a strategic perspective, as it focuses solely on short-term decisions without incorporating predictive insights into current and future shunting costs. SI policies provide a framework for the model to leverage various factors that influence the number of upcoming shunting operations. These factors are intricately tied to assumptions outlined in Section 3.1, including those related to clustering, and have demonstrated significant impacts on wagon fleet size, departure delays, and the average mileage covered by each wagon.

The general structure of an SI policy is presented as (4.31). Depending on the policy criteria, the weights $W_1$ and $W_2$ can assume different meanings, enabling the objective function to align with specific tactical and strategic objectives.

$$\sum_{i\in A}\sum_{j\in R}(\alpha W_1 + \beta W_2)z_{i,j} \tag{4.31}$$

In our case study, we will refer to $W_1$ as the cost $c_{s_{i,j}}$, while $W_2$ will vary based on the applied policy. The operational time available for shunting is defined as the duration between the arrival of the inbound train in the shunting yard and the deadline for the outbound train's departure, set at 180 min after the scheduled departure time. According to this definition, $\alpha$ and $\beta$ are complementary parameters describing the temporal state of the system, and they are dynamically determined in real-time by the MINLP constraints. Specifically, $\beta$ represents the remaining percentage of operational time available for shunting once all the SO operations have been performed (the latter is expressed by $\alpha$). In practical terms, $\beta$ signifies the fraction of operational time allotted for shunt-in operations before the train faces cancellation. Depending on the values assumed by $\alpha$ and, consequently, by $\beta$, the solver will decide whether to prioritize the cost $c_{s_{i,j}}$ or the policy criteria $W_2$. Since we are dealing with two distinct measurement units, it becomes essential to normalize both $W_1$ and $W_2$ (in our case, utilizing *Min-Max Normalization* (Patro and Sahu, 2015).

#### 4.4.2. MIN policy

This policy aims to shunt in the shunting yard's wagon with the minimum virtual rate $n_{ms_j} = \frac{ms_j + r_T}{ms_{max_j}}$. Therefore, the objective function is the following:

$$\sum_{i\in A}\sum_{j\in R}\left(\alpha c_{S_{i,j}} + \beta n_{ms_j}\right)z_{i,j} \tag{4.32}$$

Based on (4.32), if the percentage of operational time required by the SO operations is predominant, the solver will choose to shunt in wagons with a lower cost $c_{s_{i,j}}$. Otherwise, if there is enough operational time remaining for the SI operations, the solver will be directed to choose wagons with the lowest $n_{ms_j}$.

#### 4.4.3. AVG L-S policy

This policy is designed to optimize the utilization of a wagon's mileage capacity based on its service assignment history. When a wagon $i$ has been assigned numerous long-trip services ($n_{long_i}$), it is reasonable to allocate it to short-trip services ($n_{short_i}$), to fully exploit its max mileage $m_{max_i}$. The assignment record is quantified by the *degree of unbalance*, expressed in equation (4.33). We aim to select the suitable $i$-th wagon $\in S$ with the highest degree of unbalance, defined as:

$$\theta_i = n_{long_i} - n_{short_i} \tag{4.33}$$

The objective function will therefore become:

$$\sum_{i\in A}\sum_{j\in R}\left(\alpha c_{S_{i,j}} - \beta\left((-1)^{long}\theta_i\right)\right)z_{i,j} \tag{4.34}$$

Where, by assuming $r_{threshold}$ as the 25$^{th}$ percentile of the distribution

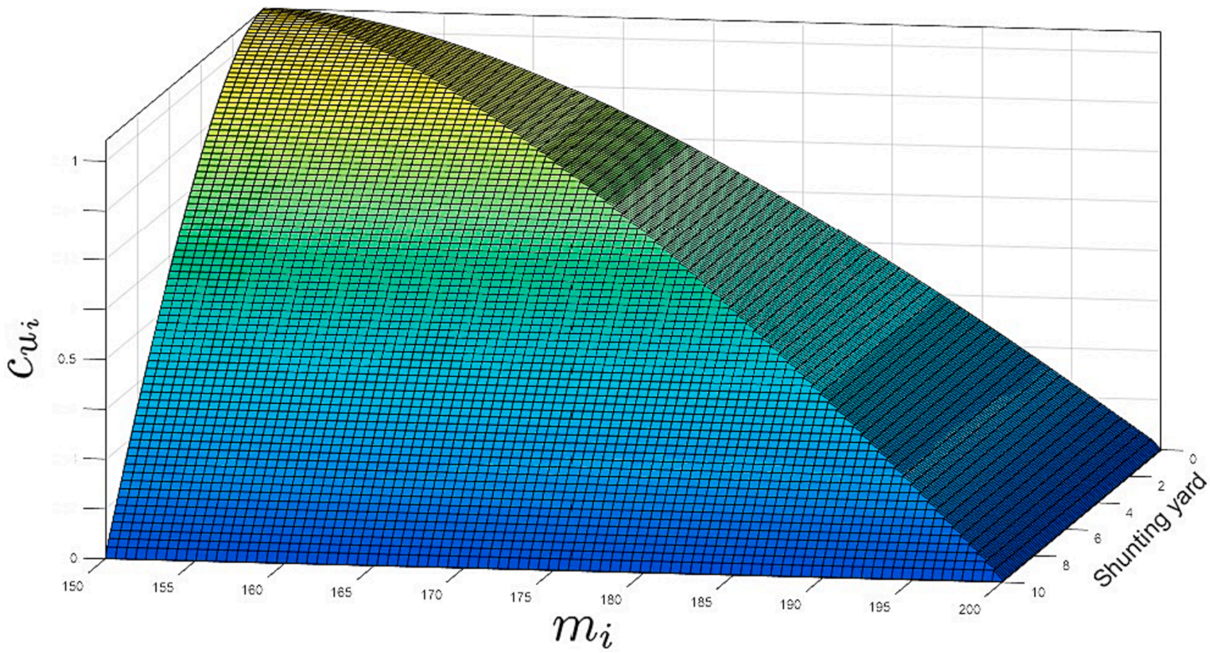


**Fig. 8.** Example of the $c_u$ function expressing the convenience of shunting out with $\lambda_k = 10$, $m_{min_i} = 150.000$ Km and $m_{max_i} = 200.000$ Km.

of the trips considered, the binary data *long* is equal to 1 if $r_T$ for the outbound train exceeds $r_{threshold}$, and 0 otherwise. If the upcoming service involves a long trip, the solver will prioritize wagons that have undertaken more short trips than long ones; otherwise, if the next service is a short trip, the solver will select wagons with more long trips than short ones.

#### 4.4.4. NCLD policy

The NCLD policy assists the solver in forming clusters with maximum width. This strategy is closely tied to the clustering assumption and aims to generate clusters of the greatest possible size. We define the virtual rate for both inbound train and shunting yard wagons as follows:

$$n_{m_i} = \frac{m_i + r_T}{m_{max_i}} \tag{4.35}$$

$$n_{ms_j} = \frac{ms_j + r_T}{ms_{max_j}} \tag{4.36}$$

If wagons with homogeneous virtual rates are shunted in, there is a high likelihood that, during subsequent trips, the wagons of the inbound train will be shunted out collectively. The add-on in the multi-component objective function (4.1) will be:

$$\sum_{i\in A}\sum_{j\in R}\left(\alpha c_{S_{i,j}} + \beta|A_{SO} - A_{SI}|\right) z_{i,j} \tag{4.37}$$

With $A_{SO}$ and $A_{SI}$ being, respectively, the average virtual rate on the inbound train after completion of all the SO operations and the average virtual rate of the wagons shunted in, as defined by equations (4.38) and (4.39).

$$A_{SO} = \frac{\sum_{i\in A}(1-\gamma_i)n_{m_i}}{\sum_{i\in A}(1-\gamma_i)} \tag{4.38}$$

$$A_{SI} = \frac{\sum_{i\in A}\sum_{j\in R} z_{i,j} n_{ms_j}}{\sum_{i\in A}\sum_{j\in R} z_{i,j}} \tag{4.39}$$

That way, if $\beta$ is higher than $\alpha$, (4.37) will minimize the distance between the average virtual rate of the left wagons on the outbound train and the one of the wagons shunted in.

### 4.5. Shunting convenience costs

The sum (4.40) in the third term of the objective function (4.3) was added as a pre-emptive tool to avoid a quick shortage of wagons as resources inside the shunting yard. In this subsection, we explain the motivation behind this choice.

$$\sum_{i\in\mathcal{T}} y_i c_{u_i} \tag{4.40}$$

Let's assume we have not added (4.40) and, for instance, the inbound train has arrived much earlier than the outbound train's departure time planned. Moreover, we have only two wagons requiring maintenance while the others have a $m_i + r_T$ lower than their $m_{max_i}$ and the inbound and outbound trains' compositions are the same (Fig. 7). Then, in order to minimize the number of shunts performed, and consequently the economic costs associated with it, the MINLP model will shunt out the entire train as it has enough time till $d_T$. Nevertheless, this approach is quite greedy in terms of resources consumed and it is much more likely that the shunting yard will quickly run out of wagons, causing infeasible solutions. (4.40) represents a limitation to that behavior. To consider both the wagon's current mileage $m_i$ and the shunting yard state, we insert new real-valued data $c_{u_i}$, associated with each inbound train's wagon variable $y_i$, and a threshold $\lambda_k$ on the shunting yard's number of wagons for each wagon type $k$.

Our aim is to preferably shunt out a wagon, with a maintenance or an optional shunt, when its current mileage is quite high and the shunting yard is not suffering, i.e. the current number of wagons of type $k$ is not below its corresponding threshold. To determine the $c_{u_i}$ values, we have considered the following functions:

$$v_i = \begin{cases} 0, m_i + r_T \le m_{min_i} \\ \dfrac{m_{max_i} - (m_i + r_T)}{m_{max_i} - m_{min_i}}, m_{max_i} \ge m_i + r_T \ge m_{min_i} \end{cases} \tag{4.41}$$

$$v_{S_k} = \begin{cases} 0, an_{S_k} \ge \lambda_k \\ \dfrac{\lambda_k - an_{S_k}}{\lambda_k}, \lambda_k \ge an_{S_k} \ge 0 \end{cases} \tag{4.42}$$

$$c_{u_i} = v_i v_{S_k} \tag{4.42}$$

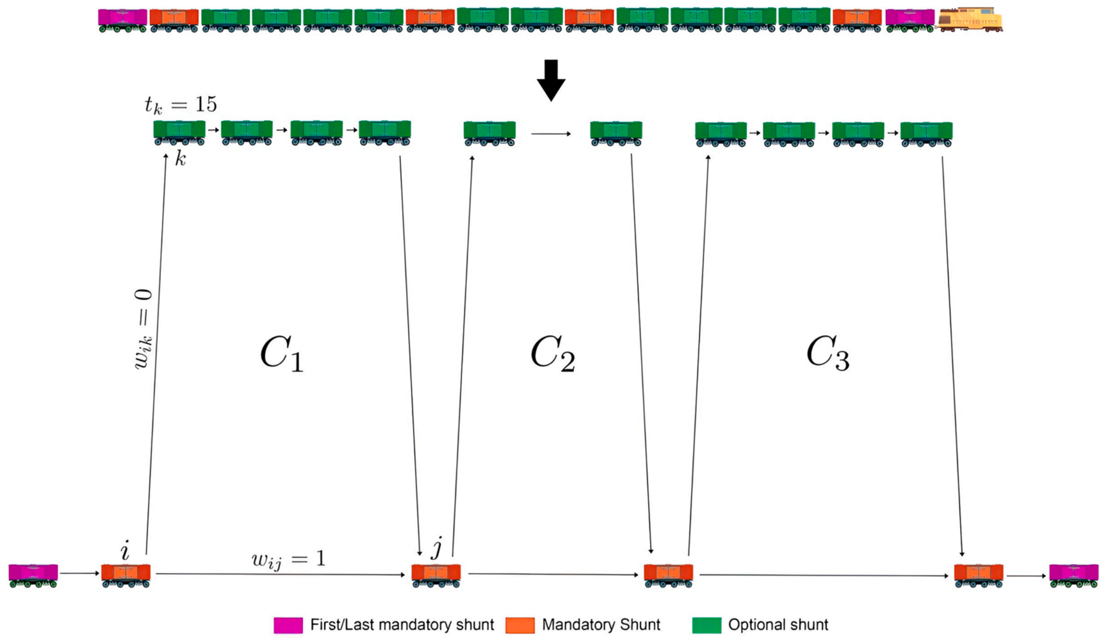


**Fig. 9.** Graph representation of an inbound train requiring SISO operations. $C_1$, $C_2$ and $C_3$ are the potential clusters to be activated.

Where $an_{S_k}$ is the current number of wagons inside the shunting yard and $v_{S_k}$, since it is calculated based on $\lambda_k$, depends on the type $k$ of the wagon $i$-th. As shown in Fig. 8, when the $m_i$ of wagon $i$-th is high, the weight will be closer to 1. In the same way, the more the shunting yard threshold is exceeded the higher the weight will be.

### 4.6. Multi-objective dijkstra algorithm

As we can see from Section 4.3, the SO sub-model is populated by a higher number of variables and constraints compared to the SI one. This is confirmed by the results, which prove how the average computation time consists of 85 % SO sub-model and the remaining percentage by SI sub-model. Moreover, in practice, there may be limited resources on which DSS tools can be run and multiple inbound trains might require SISO operations at once. In fact, practitioners drew attention to the possibility that because of parallel tracks, two or more incoming trains could require"simultaneous" SISO operations (*Multi-train* case), namely, considering an arbitrarily short time window. This means that we should find methods to solve the single-train case for the SO problem as fast as possible. These are the motivations behind the *Multi-Objective Dijkstra Algorithm* (MDA) described in this subsection. An essay on the Dijkstra Algorithm is given in Appendi*ices B and C*.

#### 4.6.1. The Multi-Objective Shortest Path

In many applications of routing problems as well as in our case, a single attribute is not sufficient to express the preference between paths. For instance, the SO problem mainly focuses on creating as many clusters as possible, but also on maintaining an acceptable service level. This means we have to work on the *Multi-Objective Shortest Path Problem* (MOSP) (Wang et al., 2019) (Casas et al., 2021), in which cost vectors are defined for arcs and/or nodes. In our case, we can opt to associate a cost with both arcs and nodes: the unitary time to shunt out, equal to 15 min, associated with each node (wagon shunted out); a binary data equal to 0 or 1, associated with each arc (expressing the choice to cluster or not).

#### 4.6.2. MDA assumptions

For the MDA the following assumptions are considered:

- For each wagon, we consider only a max mileage $m_{max_i}$, defined by the leasing contract, after which the wagon must be shunted out with a maintenance shunt;
- We can have three different instance cases, namely, the only demand, only maintenance, and maintenance&demand cases. The MDA is run only if we have two or more maintenance shunts, thus in the last two cases;
- The MDA does not consider the shunting yard wagons available, focusing only on economic and temporal costs. Indeed, the algorithm will choose whether or not to activate optional shunts between two consecutive mandatory shunts, in order to both achieve clustering advantages plus avoid delay and train cancellations;
- Since the MDA only focuses on the SO problem, the time available for SO operations is set at half the overall operational time. This is assumed to allow the SI sub-model to have enough time for the SI operations.

#### 4.6.3. Graphs

Before running the MDA, we handle inbound train data such that they can populate graphs. Let us assume we have a graph $G(N, A)$ where: each node $v_i \in N$ represents the $i$-th inbound train's wagon; each arc $(v_i, v_j) \in A$ represents the adjacency relation between two consecutive nodes $v_i$ and $v_j$. Each node and arc are associated with, respectively, costs equal to 15 and 0. These costs express the unitary time to shunt out a wagon (15 min), and the choice to cluster or not. The MOSP for this first graph consists of clustering all the wagons between the first and last mandatory shunts. At this stage, we identify all the maintenance shunts that we have to perform, due to leasing contract constraints on the max mileage $m_{max_i}$. A graph $H(N_h, A_h)$ is then created by both cutting $G$'s nodes not included in the sub-graph that goes from the first to the last mandatory shunts and adding new arcs between consecutive mandatory shunts (Fig. 9). Costs equal to 1 are associated with these arcs, in order to express the interruption between clusters. When one of these arcs belongs to the MOSP, the algorithm has opted not to cluster between consecutive mandatory shunts due to the impact of optional shunts on the departure delay. $A_h$ will be a set of arcs containing both *optional* arcs of cost 0, activating clusters, and *interruption* arcs of cost 1, interrupting clusters.

#### 4.6.4. A MOSP for the SO problem

The MDA aims to find the Multi-Objective Shortest Path between the

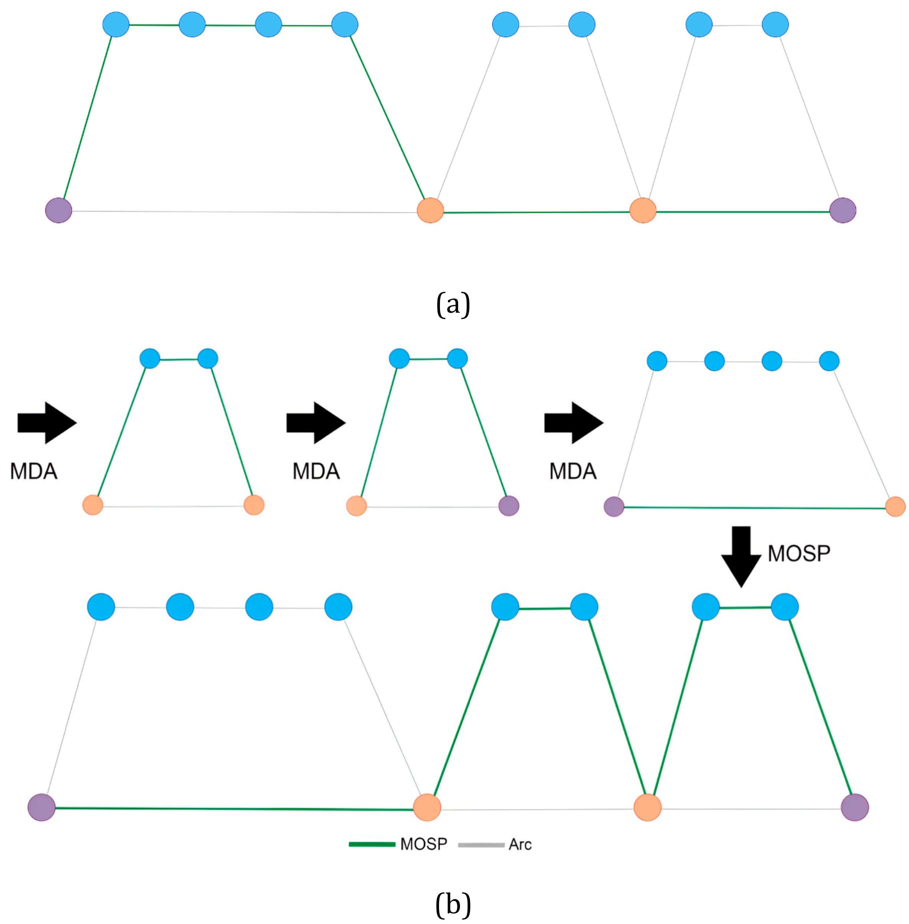


**Fig. 10.** Comparison between the MDA with and without splitting.

first and last nodes bounding $H$. For each mandatory shunt encountered, the MDA will choose whether to continue with the outgoing optional (to cluster) or interruption arc (not to cluster). To do this, the algorithm considers both the economic cost and the impact of the cluster on the value of $\mu(ad_T)$(*Appendix C*).

If the delay is zero or quite low, the MDA will choose to cluster. The decision the MDA has to take for each consecutive maintenance shunt is based on the comparison between two costs: the cost of the interruption arc between two consecutive mandatory shunts, equal to 1; the additional delay produced by the sum of the costs of nodes between consecutive optional shunts, multiplied by the weight $\frac{|\mathcal{T}|}{2}$, the same weight used for the delay penalty in the objective function 4.2. Considering the potential cluster $C_1$ in Fig. 9, the MDA checks if $\mu(ad_T)\frac{|\mathcal{T}|}{2} - w_{i,j} < 0$ with $\mu(ad_T)$ computed based on the following estimated time:

$$t_e = t_c + 15 n_{C_1} \tag{4.44}$$

where $t_c$ is the time at the current mandatory shunt $i$ while $n_{C_1}$ is the number of wagons between the mandatory shunts $i$ and $j$, potentially subjected to optional shunting.

As explained previously: $\mu(ad_T) = 0$ till $t_e < d_T$; if $d_T < t_e \leq dd_T$ then $\mu(ad_T) = \frac{ad_T - d_T}{dd_T - d_T}$; otherwise $\mu(ad_T) = 1$. If $\mu(ad_T)\frac{|\mathcal{T}|}{2} - w_{i,j} < 0$, it is convenient to activate the cluster $C_1$, as the choice of an interruption arc would produce a higher value of the objective function. Nevertheless, considering the inbound train's arrival sequence of clusters as it is, might lead to opting for sub-optimal solutions. Let us assume both to have the instance in Fig. 10 and that we have enough time to shunt 8 wagons before $d_T$. If we apply the algorithm as it is explained so far, the MDA will choose option 10a.

Indeed, $\mu(ad_T)$ is not a local value, like costs on arcs, but an overall value computed based on the wagons shunted out. The best choice is therefore to pass the first cluster and activate the second and third ones. This way the outbound train will be not delayed, and two clusters will be activated instead of one (as you can see in Fig. 10b). To do this, we split the original graph into sub-graphs and run the MDA on each of them, consecutively, from the smallest to the largest. The multi-objective Dijkstra algorithm is proposed as a faster alternative solution method for the SO sub-model in the MINLP model. The MDA is able to provide a feasible solution to the SO sub-problem, namely, the clusters created to meet both maintenance and demand-matching constraints for the inbound train being processed. The selection of clusters is determined by the Multi-Objective Shortest Path found between the first required shunt and the last one. The MDA has shown promising results in terms of average computation time and optimality loss.

## 5. Computational results

### 5.1. MINLP model results

The simulations refer to two time windows: 2021 and 2022–2050. This is done according to CFL's short-term objectives and Europe's carbon neutrality deadline (Birol, 2021). The short-term analysis compares the MINLP model performance with real data provided by practitioners. In lack of forecast data on the demand, the long-term analysis's benchmark consists of running the simulator with the activation of

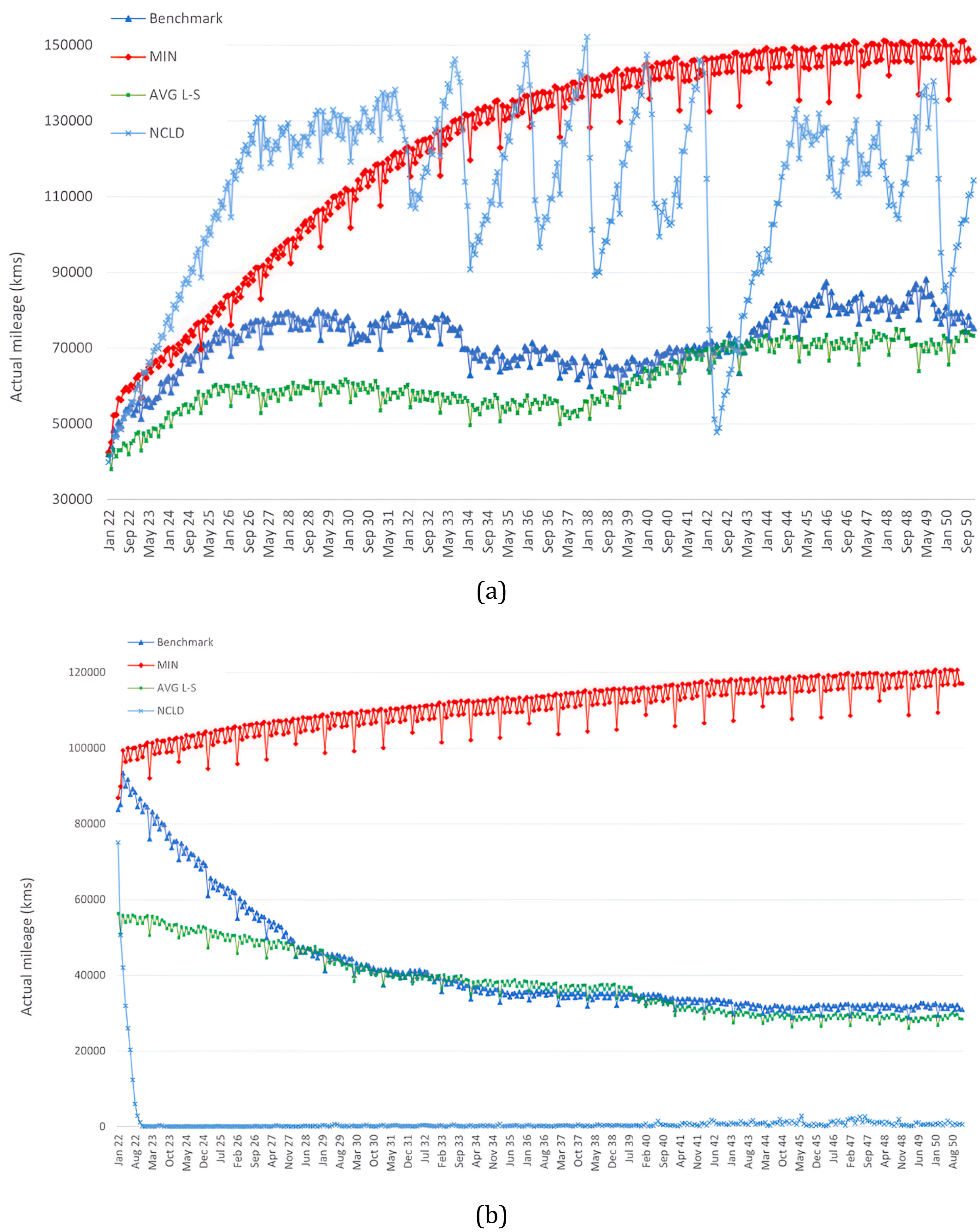


**Fig. 11.** Average mileage for the Simple (a) and Double (b) wagons available in the shunting yard.

maintenance shunt only when the $m_{max_i}$ is exceeded while choosing randomly suitable shunting yard wagons for the SI operations. Indeed, this is the approach applied to date by CFL practitioners. At the end of each month, data are collected for the shunting yard, workshop, shunting operations, average departure delay, and so on. We focus on three key points, intricately tied to the tactical and strategic goals influenced by SISO operations (as expounded in Section 2):

- *Wagon fleet size management*. The size of the wagon fleet emerges as a critical consideration due to its profound impact on various facets, including wagon idle time, overhead expenses, and storage costs. This is particularly significant given that the cost of rolling stock stands out as one of the most influential factors in the competitiveness of a railway company (Giacco et al., 2014);
- *Shunting operations optimization*. The minimization of shunting operations bears a direct linear correlation to the reduction of emissions from shunting locomotives and a subsequent decrease in operational costs. Addressing this aspect not only aligns with environmental sustainability but also offers a tangible avenue for cost savings, thereby enhancing the overall economic viability of railway operations;
- *Mitigating Delays* and *train cancellations*. The avoidance of delays and train cancellations stands out as a primary objective aimed at ensuring a consistently high service level. Delays not only disrupt schedules but can also have cascading effects on customer satisfaction and the overall reliability of the rail network.

In essence, our focus on these three key elements reflects a holistic

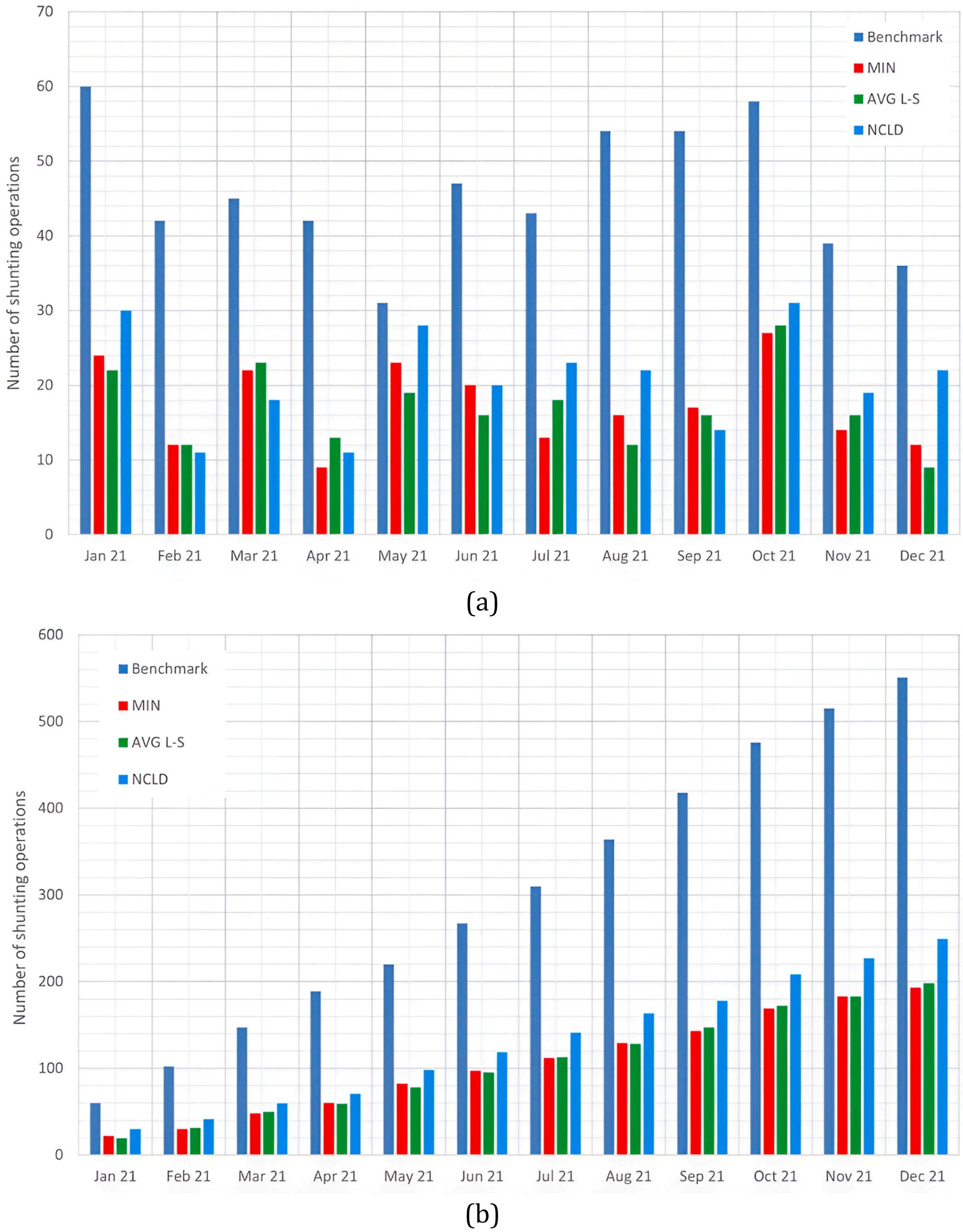


**Fig. 12.** Shunting operations performed monthly (a) and cumulative (b) for the 2021 simulation.

strategy encompassing fleet management, environmental responsibility, and service reliability.

*5.1.1. Wagon fleet*

A feasible solution for the SISO problem is the one satisfying maintenance and demand-matching constraints. Therefore, in the shunting yard, the number of suitable wagons must be enough to replace the wagons shunted out due to maintenance and to fulfill the outbound train planned composition. In our case study, the distribution of wagon requests by type exhibits a notable imbalance, with 18 % allocated to the Simple type and a predominant 82 % to the Double type. This asymmetry underscores the necessity for precise wagon management to circumvent potential challenges. Specifically, mismanaging the allocation of Simple wagons could introduce a bottleneck in the system, potentially resulting in critical operational disruptions such as train cancellations. For the simulations, we based our model on a fleet of 1100 wagons. This quantity is deemed reasonable and reflective of the available wagon fleet size according to the actual operational context observed in CFL's datasets. If a wagon is used, during the run, for at least one service, it will be counted as part of the solution. Each wagon is held under a lease contract defining a cost per day based on the wagon type: Simple wagons cost € 23 per day, while the Double ones cost around € 54 per day. In Table A.1 of *Appendix A* data relative to the mileage distribution on wagons available and the fleet used resulting from the 2021 simulation are gathered. Our analysis goes beyond solely assessing the fleet size; we delve into the saturation levels of mileage before maintenance for each wagon. This additional dimension is crucial as it provides insights into the potential unavailability of both Simple and Double

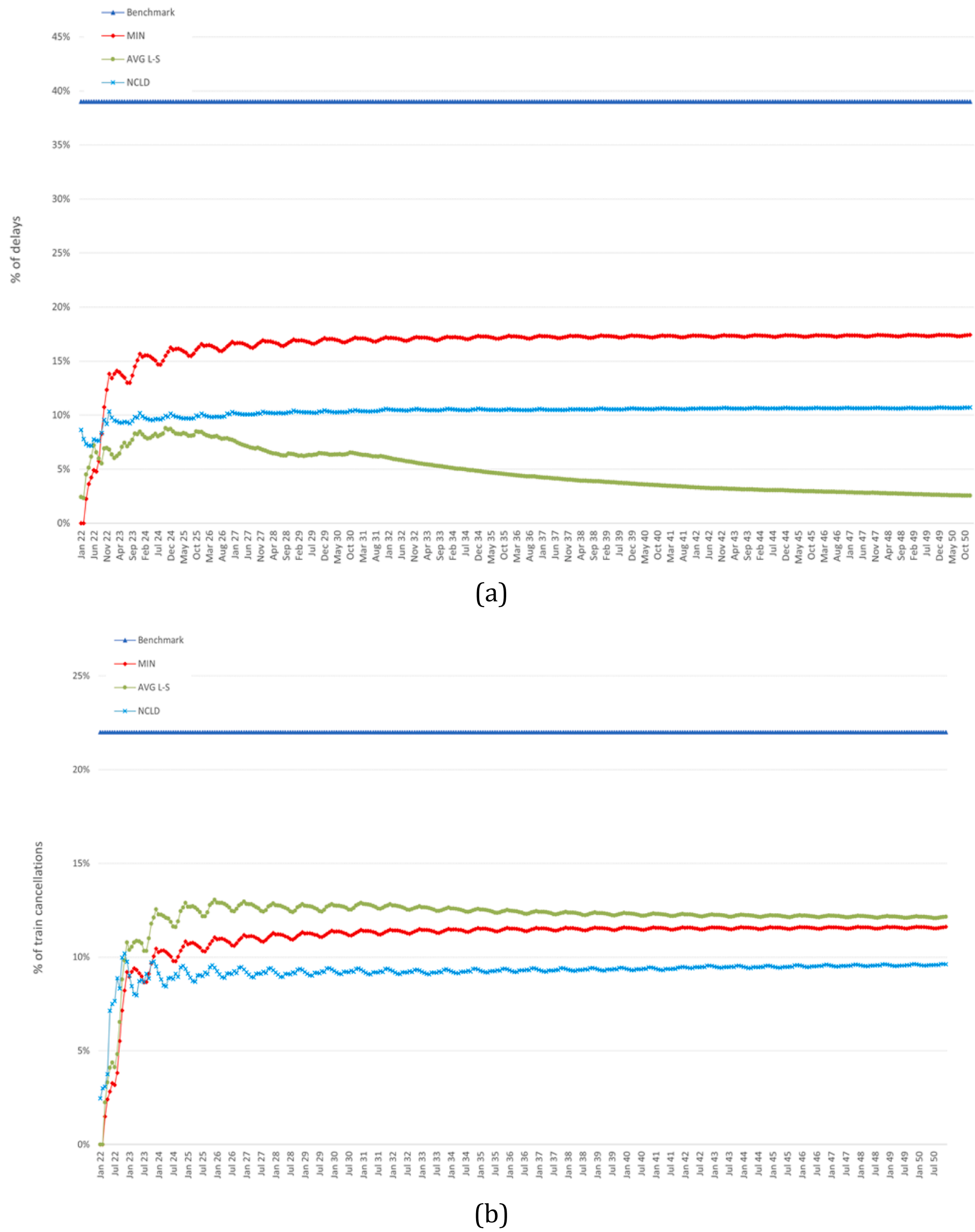


**Fig. 13.** Percentages of delays (a) and train cancellations (b) for the 2022–2050 simulation.

wagons. Specifically, a lower average residual mileage is directly correlated with an increased likelihood of unavailability, thereby elevating the risk of train cancellations. We observe that, for Simple wagons, the MIN policy demonstrates the most effective management. The higher standard deviation, median, and mean, along with a smaller number of required wagons, translate into a more strategic use of the wagon pool. On average, each wagon covers more kilometers compared to the benchmark, optimizing resource utilization. Yet, these advantages may be accompanied by potential implications, notably a heightened probability of breakdowns or technical failures due to increased wear and tear from extended usage. In the case of Double wagons, both the MIN and AVG L-S policies exhibit superior overall management, showcasing more than double the average kilometers performed per wagon and a higher standard deviation compared to the benchmark. This efficiency allows the company to allocate fewer wagons to the terminal for service fulfillment, resulting in significant economic savings ranging from €2 to €3.6 million. Given the significantly higher percentage of Double wagons compared to Simple ones, both MIN and AVG L-S policies emerge as viable alternatives when considering only operational and storage costs. However, as previously highlighted, the scarcity of Simple wagons introduces the risk of unfeasible solutions, such as train cancellations. In this context, the MIN policy could be considered the optimal approach for managing both Simple and Double wagons in the short term, ensuring a balanced and sustainable operational strategy. For the 2022–2050 simulation, we have also added data on the average current mileage of wagons available in the shunting yard (Fig. 11a-11b and Table A.2 in *Appendix A*). This is a good index of fleet reliability. Although MIN performs better in the short term, its short-sighted

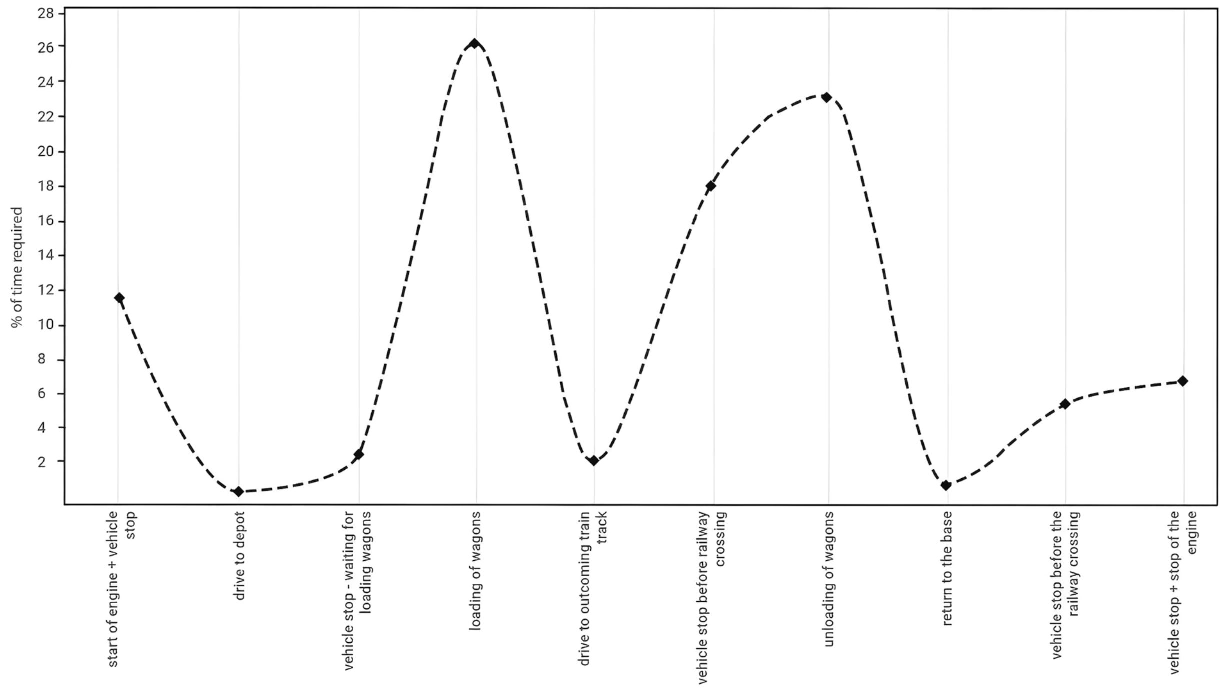


**Fig. 14.** The share of the individual activities in the whole cycle of work of an SM42 locomotive.

approach pays in the long term. Especially for the Simple wagons, MIN tends to use the newest wagons by leaving parked the ones with the highest mileage. This behaviour could lead to infeasibility, since it may happen that all available wagons can not perform the outbound train's next trip. Instead, AVG L-S shows positive behavior by keeping, for the Simple wagons, the average mileage of around 70,000 km. An average mileage curve too low paired with a high standard deviation could be the product of a significant number of shunts performed. This translates into a fleet that is underused and a rapid deterioration of the service, which goes against our objective. NCLD shows an interesting trend typical of supply chain/inventory management, called the *Bullwhip Effect* (Wang and Disney, 2016). This effect can result from the inclination of this policy to send in maintenance wide clusters of Simple wagons altogether. Once these wagons return from the workshop with a mileage zeroed, they create the drops depicted in Fig. 11a. Data regarding the wagon fleet used and the relative management costs show how, compared to the benchmark, AVG L-S uses less than half of the fleet while providing the same reliability in terms of the wagons available in the shunting yard. Even if the benchmark and MIN manage better the Simple type, AVG L-S shows a better usage of the fleet overall, with a higher median, mean, and standard deviation. This translates into fewer wagons used to the fullest, and savings of € 203 mln over 30 years.

### 5.1.2. *Shunting operations*

In this subsection, we discuss the impact of the SISO model optimization on the shunting operations performed, directly correlating to both shunting locomotive emissions and operational costs.

Fig. 12 illustrates the favorable impact of an optimization approach on the number of operations conducted in the short term. Specifically, while the clustering rate for the benchmark is approximately 2 wagons per shunt, both MIN and AVG L-S exhibit significantly improved rates at 6 wagons per operation (5.95 and 5.6, respectively). However, an excessively high clustering rate can potentially strain the availability of wagons in the shunting yard. This, in turn, heightens the risk of encountering unfeasible solutions, particularly when the clusters are populated by a substantial number of Simple wagons. Therefore, maintaining a balanced clustering rate is crucial to prevent operational constraints and ensure the feasibility of the shunting process. In the simulation spanning from 2022 to 2050, as the benchmark incorporates the optimization model for SO operations, the cumulative gap in shunting operations with different policies becomes narrower (refer to Table A.3 in *Appendix A*). Looking at the clustering rate (Tables A.4-A.5 in *Appendix A*), the benchmark stands at 2 wagons/operation, while MIN drops from 5.95 to 1.85 wagons/operation. In the long term, MIN is not encouraged to perform more optional shunting. Instead, AVG L-S increases its wagon/operation ratio from 5.6 to 6.6. Such behavior allows a fleet flexible rotation throughout the services by distributing mileages efficiently. It is worth mentioning that, even if we aim to raise the clustering rate due to economic benefit, it can negatively impact the delays and train cancellation rate, the time required to operate is directly proportional to the number of wagons shunted out.

### 5.1.3. *Service level*

The punctuality rate is a critical point for freight train operation, since, on average, shunting operations can affect up to 20 % of delays and train cancellations, according to CFL. As stated in Section 2.4, the expected increase in railway traffic will force practitioners to reduce the reserved time, moreover, freight trains could suffer from the additional delay caused by the lower priority compared to passenger trains. A freight train is considered delayed if it departs between 60 and 180 min later than the scheduled departure time. When the train exceeds the deadline of 180 min, it is canceled except in case that carries high-value goods. In our instances, arrival delays may be due to the combination of trips and shunting operations.

To comprehensively address delays arising not just from SISO operations but also from trip attributes and train composition, we incorporated a Machine Learning model developed by (Pineda-Jaramillo and Viti, 2023) into our framework. This model calculates travel times (Barbour et al., 2018) based on a range of wagon attributes, including weight, length, volume, and other relevant factors. The cyclical delay trend is strongly related to the demand seasonality, as there are usually periods when the demand for goods is high, requesting a larger number of demand shunts to be fulfilled. In Table A.6 of *Appendix A* data comparing SISO policies and benchmark delay and train cancellation rates for the 2021 simulation are gathered. Policies provide a significant reduction in the percentage of delays and train cancellations, and the

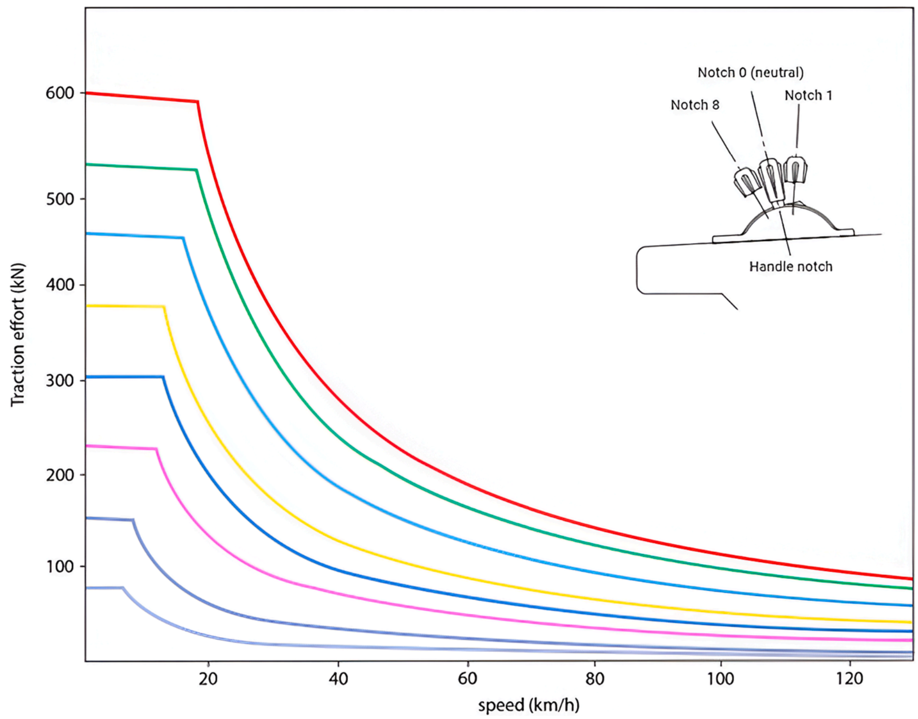


**Fig. 15.** The traction effort characteristics for the GT46C-Ace locomotive.

Trieste service line is the most representative. Trieste involves 33 % of the services and requires a considerable number of wagons, which translates into a high number of shunting operations. The SISO model here proves its strength by reducing the delays and cancellations rate up to 4 % for MIN and AVG L-S compared to 40 % of the benchmark. Furthermore, results suggest that tailored SI policies could exhibit enhanced effectiveness on specific routes. The efficacy of these policies is intricately linked to factors like wagon demand and trip attributes that characterize the unique features of each origin–destination section. This observation underscores the potential for customized SI strategies, providing an opportunity to optimize performance with greater precision on distinct routes. The 2022–2050 analysis does not look at single destinations but at the overall percentage of delays and train cancellations (Fig. 13 and Table A.7 in *Appendix A*). This approach enables us to identify systemic patterns, trends, and potential areas of concern that might not be apparent when looking solely at single destinations. In this case, our benchmark consists of the average rate of delays and cancellations provided by the practitioners set at, respectively, 38 % and 22 %. Results show how NCLD and MIN keep a steady cumulative delay throughout the simulation while AVG L-S, once passed a warm-up phase where the shunting activity is heavier due to the setup of the *degree of unbalance*, reduces the delay rate over time by increasing the clustering rate. Fig. 13b shows how all the SI policies managed to reduce by around 10 % the cumulative cancellation rate compared to CFL data.

#### 5.1.4. *Emissions*

In this subsection, we collect and analyze data regarding the fuel consumption and corresponding emissions generated by the shunting activity in our simulations. It is important to note that there exists a linear positive correlation between carbon emissions and the number of shunting operations. This relationship arises from the fact that emissions are directly derived from the fuel consumed by shunting locomotives during their maneuvers. For this purpose, we need to record each operation carried out by a diesel locomotive in the shunting yard as well as the fuel consumed (as we can see in Fig. 14 for an SM42 locomotive).

Furthermore, the fuel consumption is directly related to *Notch positions*, which control locomotive operations from the 8-notches control panel. It was proven that during these operations notch changes can occur more than 400 times per hour, which is around 20 times the number of notch changes performed when traveling (Rymaniak et al., 2019). The emission of greenhouse gas increases proportionally with fuel consumption, as it results from the traction effort and transient operating conditions of the internal combustion engine (Fig. 15).

With respect to the data reported by (Agency, 2021), the European average fuel consumption of a diesel shunting locomotive can be derived by using the cumulative hours of shunting activity per year, applying the fuel consumption factors in Table A.8 in *Appendix A*. Therefore, by defining a standard operational time to shunt given by the sum of the shunt out and shunt in average times (55 min), the cumulative number of hours of shunting activity per month/year, and the fuel consumed, is computed as follows:

$$TOT_f = \left(N_s \frac{55}{60}\right) f_s \tag{5.1}$$

Where:

- $TOT_f$ is the total fuel consumed in *kg*;
- $N_s$ is the number of shunts performed;
- $f_s$ is the *kg* of fuel consumed per hour.

From here, we can gather the volumes of the main gases produced per kg/tonne of fuel (Rymaniak et al., 2019). Emission factors have been derived from data in the Diesel Railway research by (Biemann and Notter, 2023). This study provides an assessment of the diesel locomotive fleet in Europe and average emission factors. The most significant pollutants produced are reported in Table A.9 of *Appendix A*, with a particular focus on $CO_2$, $NO_x$, and *PM*, and of lesser, but still significant importance, are emissions of *CO*, NMVOC$_s$ (non-methane volatile organic compounds), and some metals (Agency, 2021). Indeed, the above gases are the major contributors to the greenhouse effect, and therefore to global warming. Moreover, $NO_x$ is responsible for the so-called *acid rain* while *CO* is usually produced by the exhaust of engines and is highly toxic for human beings. By using fuel data as the primary activity indicator (Table A.10-A.11 in *Appendix A*), we can

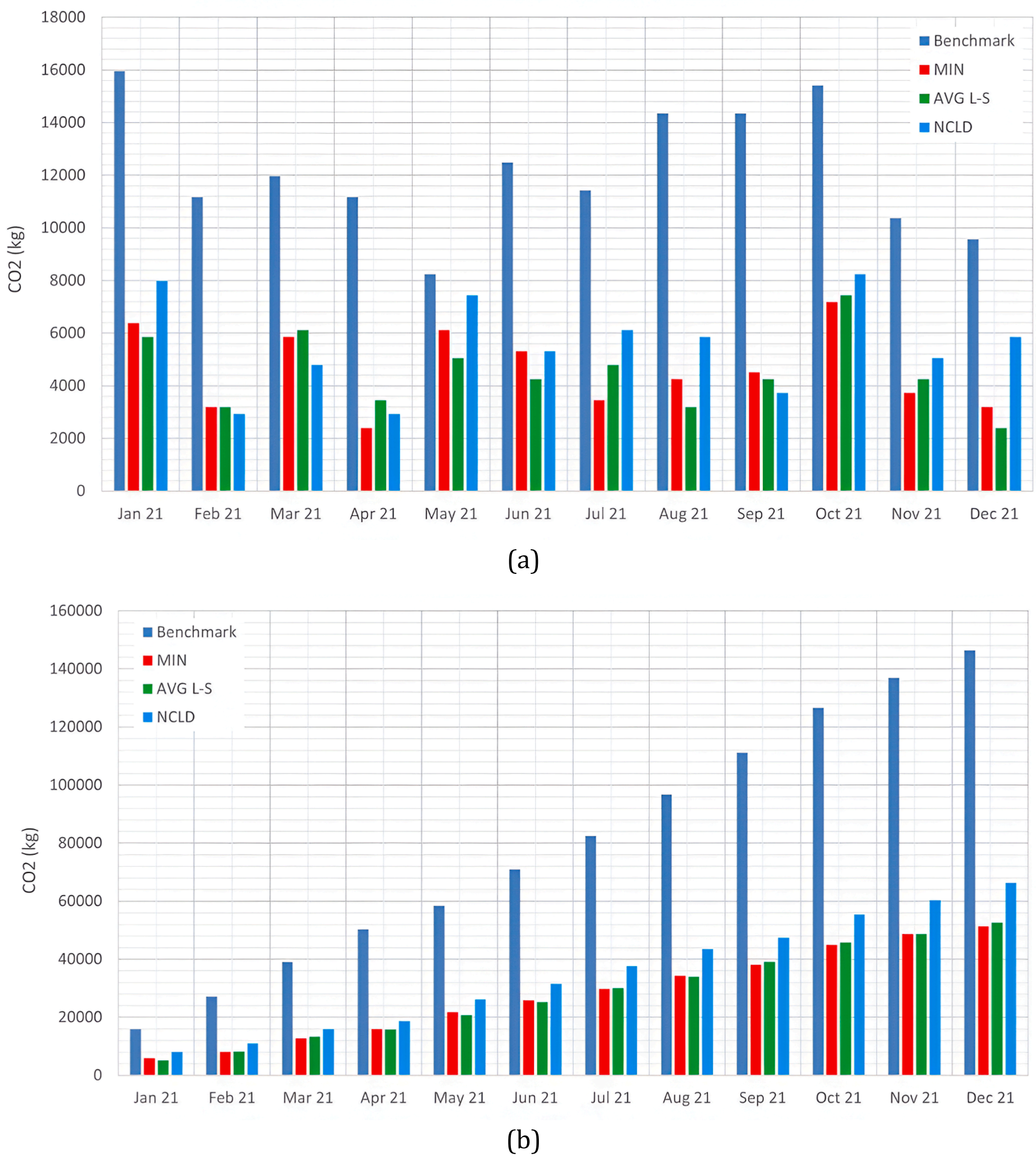


**Fig. 16.** Average and cumulative kilograms of $CO_2$ produced by the benchmark and each policy for the 2021 simulation.

extract the annual volumes produced for each gas. In order not to burden the analysis, the following formula has been used to compute the average and cumulative tonnes of emissions per month/year from the fuel consumed:

$$E_i = \sum_m \sum_j \left(FC_{j,m} \times EF_{i,j,m}\right) \quad (5.2)$$

Where:

- $E_i$ represents the emissions of the pollutant $i$-th for the period concerned in the inventory (kg or g);
- $FC_{j,m}$ is the fuel consumption of the fuel type $m$ used by the category $j$ for the period and area considered (tonnes);
- $EF_{i,j,m}$ is the emission factor of pollutant $i$-th for each unit of the fuel type $m$ used by the
- category $j$ (kg/tonnes);
- $m$ is the fuel type (diesel, gas oil);
- $j$ is the locomotive category (shunting, rail-car, line-haul).

We could also have looked at a weighted objective function, by assigning each gas a weight, based on the greater or lesser environmental impact. As we can see in Tables A.12-A.13 in *Appendix* A and Fig. 16, there are significant effects in terms of fuel consumption and emissions produced both in the short and long term. Considering the best SI policy, namely MIN, the reduction of monthly and cumulative emissions produced with respect to the benchmark is around 65 %. This is translatable as 95 tons of emissions subtracted from the atmosphere. The results show an initial similar trend for each policy, with a growing branch between the benchmark and MIN. In the context of the 2022–2050 simulation, it is crucial to note that these outcomes should be considered in light of the anticipated increase in freight rail traffic aligned with the EU carbon-neutrality goals. The total reduction in $CO_2$ achieved by implementing the AVG L-S policy instead of the fundamental MINLP model amounts to around 8 %, equivalent to a reduction of 1.52 tons.

### 5.2. MDA results

The MDA performance has been tested on 70 real-size instances both to focus only on the SO problem and to stress-test the algorithm with more complex situations. Indeed, these instances consider the worst case

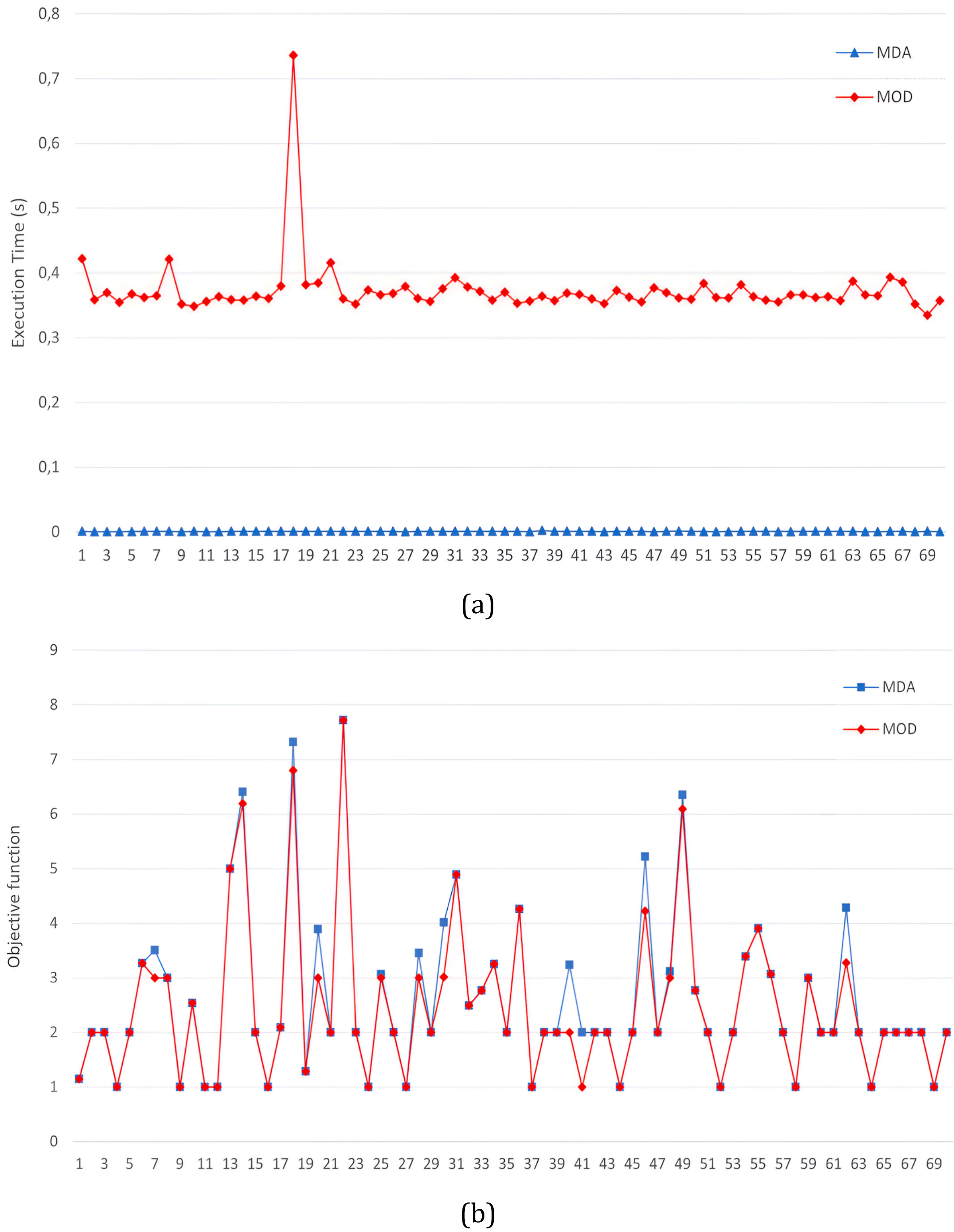


**Fig. 17.** Comparison between the MDA and the SO sub-model computational time performance.

for both the inbound and outbound train sizes (from 20 to 30 wagons) and the number of shunting operations required (+75 % compared to practitioners' data). The SO sub-model comes from the MINLP model in Section 4.3 by extracting constraints (4.4)-(4.13), (4.14) without the last sum expressing the time to SI, (4.17)-(4.18) and (4.28)-(4.29). Based on the assumptions in Section 3.1, the objective function is then reduced to the first two terms by removing the SI policy and shunting cost convenience terms.

As shown in Fig. 17a, the algorithm is on average 357 times faster than the sub-model, proving basically a constant time. Moreover, the optimality loss is around 4,7% on average, which translates into 0,015 extra shunts (Fig. 17b). Recognizing the effectiveness of the MDA in handling the SO, strategic integration into the framework becomes an envisaged possibility. The MDA could function as an initial and swift tool for SO, paving the way for the framework to subsequently invoke the SI policy by calling upon the MILP model for further refinement in the SI problem. Indeed, the non-linearity of the SISO programming model is mitigated through the proactive calculation of the operational time required for shunting out wagons. However, it is essential to note that the sequential computation of $\alpha$ and $\beta$ might pose the risk of sub-optimal solutions. Specifically, this approach could potentially have adverse effects on delays and, in extreme cases, lead to train cancellations. Consequently, while the MDA proves to be a valuable and efficient tool, careful consideration of the sequential computation of shunting operational times is imperative to mitigate the potential for sub-optimal outcomes.

**Table A1**
Distribution of mileage on rolling stocks and wagon fleet usage for the 2021 simulation.

| | Wagon Usage Distribution (Km)Wagon Fleet Used | | | | |
|---|---|---|---|---|---|
| | STD | Median | Mean | Number of Wagons | Cost (€ Mln) |
| | | | | Benchmark | |
| **SIMPLE** | 18,244 | 6380 | 19,810 | 128 | 1.1 |
| **DOUBLE** | 20,020 | 18,223 | 36,367 | 295 | 5.8 |
| | | | | AVG L-S | |
| **SIMPLE** | 11,010 | 11,376 | 14,539 | 120 | 1.0 |
| **DOUBLE** | 55,677 | 69,844 | 68,981 | 285 | 3.2 |
| | | | | MIN | |
| **SIMPLE** | 20,080 | 25,280 | 28,158 | 67 | 0.6 |
| **DOUBLE** | 69,939 | 57,840 | 81,903 | 138 | 2.7 |
| | | | | NCLD | |
| **SIMPLE** | 10,112 | 14,147 | 14,848 | 115 | 1.0 |
| **DOUBLE** | 54,328 | 46,614 | 63,136 | 192 | 3.7 |

**Table A2**
Distribution of mileage on rolling stocks and wagon fleet usage for the 2022–2050 simulation.

| | Wagon Usage Distribution (Km)Wagon Fleet Used | | | | |
|---|---|---|---|---|---|
| | STD | Median | Mean | Number of Wagons | Cost (€ Mln) |
| | | | | Benchmark | |
| **SIMPLE** | 326,368 | 125,530 | 329,237 | 170 | 28 |
| **DOUBLE** | 402,884 | 301,088 | 437,303 | 755 | 297 |
| | | | | AVG L-S | |
| **SIMPLE** | 218,118 | 152,137 | 267,500 | 209 | 34 |
| **DOUBLE** | 1,094,474 | 1,683,600 | 1,469,344 | 224 | 88 |
| | | | | MIN | |
| **SIMPLE** | 286,682 | 250,105 | 331,470 | 169 | 28 |
| **DOUBLE** | 854,882 | 662,145 | 939,078 | 351 | 138 |
| | | | | NCLD | |
| **SIMPLE** | 388,391 | 265,365 | 388,710 | 122 | 20 |
| **DOUBLE** | 678,124 | 1,388,309 | 1,082,490 | 313 | 123 |

**Table A3**
Number of shunting operations for the benchmark and each SISO policy for the 2022–2050 simulation.

| Year | Benchmark | MIN | AVG L-S | NCLD |
|---|---|---|---|---|
| 2022 | 252 | 193 | 198 | 249 |
| 2023 | 498 | 394 | 379 | 443 |
| 2024 | 717 | 585 | 569 | 639 |
| 2025 | 953 | 777 | 757 | 836 |
| 2026 | 1180 | 974 | 946 | 1033 |
| 2027 | 1416 | 1172 | 1137 | 1237 |
| 2028 | 1637 | 1372 | 1335 | 1448 |
| 2029 | 1857 | 1575 | 1557 | 1653 |
| 2030 | 2082 | 1777 | 1754 | 1863 |
| 2031 | 2293 | 1974 | 1948 | 2076 |
| 2032 | 2501 | 2171 | 2158 | 2289 |
| 2033 | 2721 | 2370 | 2375 | 2505 |
| 2034 | 2940 | 2566 | 2581 | 2727 |
| 2035 | 3161 | 2773 | 2798 | 2930 |
| 2036 | 3370 | 2976 | 2998 | 3151 |
| 2037 | 3574 | 3193 | 3202 | 3364 |
| 2038 | 3793 | 3396 | 3400 | 3597 |
| 2039 | 3999 | 3610 | 3620 | 3816 |
| 2040 | 4218 | 3807 | 3811 | 4046 |
| 2041 | 4433 | 4013 | 4010 | 4274 |
| 2042 | 4635 | 4214 | 4212 | 4517 |
| 2043 | 4850 | 4429 | 4415 | 4744 |
| 2044 | 5066 | 4634 | 4609 | 4969 |
| 2045 | 5271 | 4836 | 4809 | 5205 |
| 2046 | 5480 | 5030 | 5011 | 5433 |
| 2047 | 5692 | 5235 | 5213 | 5664 |
| 2048 | 5908 | 5434 | 5420 | 5898 |
| 2049 | 6109 | 5644 | 5621 | 6134 |
| 2050 | 6332 | 5841 | 5824 | 6363 |

**Table A4**
Average Clustering rate per month for the 2021 simulation.

| Month | Benchmark | MIN | AVG L-S | NCLD |
|---|---|---|---|---|
| Jan 21 | 1,17 | 4,18 | 3,8 | 2,22 |
| Feb 21 | 1,7 | 5,44 | 4 | 2,59 |
| Mar 21 | 1,30 | 2,56 | 2,35 | 2,41 |
| Apr 21 | 1,71 | 2,63 | 3,83 | 2,69 |
| May 21 | 2,5 | 9,15 | 8,55 | 3 |
| Jun 21 | 1,28 | 5,24 | 5 | 2,41 |
| Jul 21 | 1,45 | 7,64 | 6,13 | 2,18 |
| Aug 21 | 2,25 | 10,6 | 10,29 | 2,60 |
| Sep 21 | 1,47 | 4 | 3,75 | 2,69 |
| Oct 21 | 1,97 | 8,79 | 8,17 | 2,90 |
| Nov 21 | 2,26 | 5,75 | 5,63 | 2,84 |
| Dec 21 | 1,69 | 5,38 | 5 | 2,41 |

**Table A5**
Average Clustering rate per year for the 2022–2050 simulation.

| Year | Benchmark | MIN | AVG L-S | NCLD |
|---|---|---|---|---|
| 2022 | 1,73 | 1,85 | 12,19 | 1,36 |
| 2023 | 1,76 | 1,77 | 11,98 | 1,36 |
| 2024 | 1,75 | 1,80 | 10,81 | 1,36 |
| 2025 | 1,79 | 1,78 | 9,71 | 1,36 |
| 2026 | 1,71 | 1,83 | 7,38 | 1,34 |
| 2027 | 1,78 | 1,74 | 7,16 | 1,34 |
| 2028 | 1,77 | 1,83 | 9,10 | 1,35 |
| 2029 | 1,73 | 2,53 | 9,29 | 1,33 |
| 2030 | 1,79 | 1,84 | 7,26 | 1,35 |
| 2031 | 1,82 | 1,83 | 5,22 | 1,33 |
| 2032 | 1,77 | 1,90 | 5,27 | 1,33 |
| 2033 | 1,76 | 1,72 | 5,67 | 1,32 |
| 2034 | 1,78 | 1,78 | 4,73 | 1,35 |
| 2035 | 1,79 | 2,22 | 5,44 | 1,33 |
| 2036 | 1,78 | 1,68 | 4,96 | 1,33 |
| 2037 | 1,71 | 2,23 | 5,83 | 1,32 |
| 2038 | 1,87 | 1,80 | 4,98 | 1,32 |
| 2039 | 1,78 | 1,83 | 5,39 | 1,31 |
| 2040 | 1,80 | 1,81 | 4,96 | 1,32 |
| 2041 | 1,88 | 1,81 | 5,41 | 1,31 |
| 2042 | 1,76 | 2,11 | 5,41 | 1,31 |
| 2043 | 1,80 | 1,79 | 6,03 | 1,32 |
| 2044 | 1,78 | 1,76 | 5,11 | 1,31 |
| 2045 | 1,78 | 1,84 | 4,57 | 1,31 |
| 2046 | 1,82 | 1,85 | 5,12 | 1,31 |
| 2047 | 1,79 | 1,76 | 5,22 | 1,31 |
| 2048 | 1,83 | 1,82 | 4,64 | 1,31 |
| 2049 | 1,76 | 1,82 | 4,79 | 1,31 |
| 2050 | 1,36 | 1,33 | 2,40 | 1,19 |

## 6. Conclusions

In this paper, we present a novel methodology for optimizing the Shunt-In Shunt-Out (SISO) problem in shunting yards that operate full train load services, as this and other shunting yard issues are solved based solely on CFL practitioners' experience. The research objective is to propose shunting policies to remove wagons from the inbound trains, due to maintenance and demand-matching constraints, and replace them with suitable shunting yard wagons, in order to impact rail long-term KPIs, such as the wagon fleet size, the number of shunting operations, shunting locomotives emissions and delays and train cancellations. Our methodology involves integrating a mixed-integer nonlinear programming (MINLP) model into a simulation environment to process each inbound train requiring SISO operations iteratively. The MINLP model considers the technical feasibility given by maintenance rules and demand matching while allowing for the application of different SI policies, each of which is characterized by specific selection criteria. We also propose a Multi-Objective Dijkstra Algorithm (MDA) as a faster alternative to the Shunt-Out sub-model, which is capable of handling the multi-train case where two or more inbound trains may require SISO operations simultaneously. The MDA finds the shortest path between the

**Table A6**
Average delay and train cancellation rates per year for the 2022–2050 simulation.

| | Benchmark | | MIN | | AVG L-S | | NCLD | |
|---|---|---|---|---|---|---|---|---|
| Year | % Delays | % Canc. | %Delays | %Canc. | % Delays | % Canc. | %Delays | % Canc. |
| Antwerp | 34 % | 19 % | 21 % | 19 % | 22 % | 19 % | 25 % | 23 % |
| Champigneulles | 27 % | 26 % | 5 % | 4 % | 6 % | 5 % | 5 % | 3 % |
| Kiel | 35 % | 11 % | 18 % | 15 % | 22 % | 19 % | 12 % | 3 % |
| Lyon | 11 % | 18 % | 17 % | 16 % | 17 % | 16 % | 13 % | 18 % |
| Trieste | 44 % | 46 % | 4 % | 4 % | 4 % | 4 % | 10 % | 6 % |

**Table A7**
Average delay and train cancellation rates per year for the 2022–2050 simulation.

| | Benchmark | | MIN | | AVG L-S | | NCLD | |
|---|---|---|---|---|---|---|---|---|
| Year | % Delays | % Cancell. | %Delays | %Cancell. | % Delays | % Cancell. | %Delays | % Cancell. |
| 2022 | 39 % | 22 % | 6 % | 4 % | 6 % | 5 % | 8 % | 7 % |
| 2023 | 39 % | 22 % | 14 % | 9 % | 7 % | 11 % | 10 % | 9 % |
| 2024 | 39 % | 22 % | 15 % | 10 % | 8 % | 12 % | 10 % | 9 % |
| 2025 | 39 % | 22 % | 16 % | 11 % | 8 % | 13 % | 10 % | 9 % |
| 2026 | 39 % | 22 % | 16 % | 11 % | 8 % | 13 % | 10 % | 9 % |
| 2027 | 39 % | 22 % | 17 % | 11 % | 7 % | 13 % | 10 % | 9 % |
| 2028 | 39 % | 22 % | 17 % | 11 % | 6 % | 13 % | 10 % | 9 % |
| 2029 | 39 % | 22 % | 17 % | 11 % | 6 % | 13 % | 10 % | 9 % |
| 2030 | 39 % | 22 % | 17 % | 11 % | 6 % | 13 % | 10 % | 9 % |
| 2031 | 39 % | 22 % | 17 % | 11 % | 6 % | 13 % | 10 % | 9 % |
| 2032 | 39 % | 22 % | 17 % | 11 % | 6 % | 13 % | 10 % | 9 % |
| 2033 | 39 % | 22 % | 17 % | 11 % | 5 % | 13 % | 10 % | 9 % |
| 2034 | 39 % | 22 % | 17 % | 11 % | 5 % | 13 % | 11 % | 9 % |
| 2035 | 39 % | 22 % | 17 % | 11 % | 5 % | 12 % | 11 % | 9 % |
| 2036 | 39 % | 22 % | 17 % | 11 % | 4 % | 12 % | 10 % | 9 % |
| 2037 | 39 % | 22 % | 17 % | 11 % | 4 % | 12 % | 10 % | 9 % |
| 2038 | 39 % | 22 % | 17 % | 12 % | 4 % | 12 % | 11 % | 9 % |
| 2039 | 39 % | 22 % | 17 % | 12 % | 4 % | 12 % | 11 % | 9 % |
| 2040 | 39 % | 22 % | 17 % | 12 % | 4 % | 12 % | 11 % | 9 % |
| 2041 | 39 % | 22 % | 17 % | 12 % | 3 % | 12 % | 11 % | 9 % |
| 2042 | 39 % | 22 % | 17 % | 12 % | 3 % | 12 % | 11 % | 9 % |
| 2043 | 39 % | 22 % | 17 % | 12 % | 3 % | 12 % | 11 % | 9 % |
| 2044 | 39 % | 22 % | 17 % | 12 % | 3 % | 12 % | 11 % | 9 % |
| 2045 | 39 % | 22 % | 17 % | 12 % | 3 % | 12 % | 11 % | 9 % |
| 2046 | 39 % | 22 % | 17 % | 12 % | 3 % | 12 % | 11 % | 10 % |
| 2047 | 39 % | 22 % | 17 % | 12 % | 3 % | 12 % | 11 % | 10 % |
| 2048 | 39 % | 22 % | 17 % | 12 % | 3 % | 12 % | 11 % | 10 % |
| 2049 | 39 % | 22 % | 17 % | 12 % | 3 % | 12 % | 11 % | 10 % |
| 2050 | 39 % | 22 % | 17 % | 12 % | 3 % | 12 % | 11 % | 10 % |

**Table A8**
Fuel consumption per hour for different locomotive types.

| Category | Fuel consumption | Unit |
|---|---|---|
| Line-haul locomotives | 21.9 | kg/h |
| Shunting locomotives | 90.9 | kg/h |
| Railcars | 53.6 | kg/h |

**Table A9**
Kilograms and grams of pollutants produced per tonne of fuel consumed.

| Pollutant | Nomenclature | Value | Unit |
|---|---|---|---|
| $NO_x$ | Nitrogen oxides | 54.4 | kg/tonne |
| CO | Carbon dioxide | 10.8 | kg/tonne |
| NMVOC | Non-methane volatile organic compounds | 4.6 | kg/tonne |
| $NH_3$ | Ammonia | 10 | g/tonne |
| $PM_{10}$ | Particulate matter | 2.1 | kg/tonne |
| $N_2O$ | Nitrous oxide | 24 | g/tonne |
| $CO_2$ | Carbon oxide | 3190 | kg/tonne |
| $CH_4$ | methane | 176 | g/tonne |

first and the last mandatory shunts, providing data on the resulting clusters of shunts created. Results from our experiments show that our SISO policies have a significant long-term impact on overall costs, emissions, delays, and train cancellations. Based on data from our case study, our best-performing policies result in an overall cost reduction of € 3.8 million, a fuel consumption and emissions reduction of 65 %, and up to −36 % of trains delayed for the 2021 simulation. Based on experiments, the MDA has proven to be 357 times faster than the SO submodel with an optimality loss of only 4.7 %. Our research can be configured as a first step towards the optimization of shunting yard operations for full train load services. By mathematically formalizing the SISO problem, which has traditionally been addressed based on experience alone, we aim to reduce cognitive bias and improve the overall efficiency of shunting yards. The nature of full train load services enables the identification and tracking of attributes associated with individual trains, allowing for a long-term analysis of key performance indicators. Rather than simply waiting for the trigger of the condition-based maintenance or intuitively choosing the shunting yard replacement wagons, the application of SISO policies can significantly impact long-term KPIs as proven by results. The development and implementation of effective, cost-free shunting policies can considerably improve the rail freight industry's attractiveness, supported by an infrastructural effort.

Future research will delve deeper into the SISO multi-train case and explore new heuristics and policies for the SI problem. One potential solution for the SI problem could be a *Column Generation Algorithm*, which has shown promise for solving NP-Hard problems like this.

**Table A10**
Average and cumulative fuel consumption in tonnes per month for the 2021 simulation.

| | Average Fuel (ton) | | | | Cumulative Fuel (ton) | | | |
|---|---|---|---|---|---|---|---|---|
| Month | Benchmark | MIN | AVG L-S | NCLD | Benchmark | MIN | AVG L-S | NCLD |
| Jan 21 | 5,00 | 2,00 | 1,83 | 2,50 | 5,00 | 1,83 | 1,58 | 2,50 |
| Feb 21 | 3,50 | 1,00 | 1,00 | 0,92 | 8,50 | 2,50 | 2,58 | 3,44 |
| Mar 21 | 3,75 | 1,83 | 1,92 | 1,50 | 12,25 | 4,00 | 4,17 | 4,96 |
| Apr 21 | 3,50 | 0,75 | 1,08 | 0,92 | 15,75 | 5,00 | 4,92 | 5,87 |
| May 21 | 2,58 | 1,92 | 1,58 | 2,33 | 18,33 | 6,83 | 6,50 | 8,16 |
| Jun 21 | 3,92 | 1,67 | 1,33 | 1,67 | 22,25 | 8,08 | 7,92 | 9,86 |
| Jul 21 | 3,58 | 1,08 | 1,50 | 1,92 | 25,83 | 9,33 | 9,42 | 11,77 |
| Aug 21 | 4,50 | 1,33 | 1,00 | 1,83 | 30,33 | 10,75 | 10,67 | 13,61 |
| Sep 21 | 4,50 | 1,42 | 1,33 | 1,17 | 34,83 | 11,92 | 12,25 | 14,82 |
| Oct 21 | 4,83 | 2,25 | 2,33 | 2,58 | 39,66 | 14,08 | 14,33 | 17,36 |
| Nov 21 | 3,25 | 1,17 | 1,33 | 1,58 | 42,91 | 15,25 | 15,25 | 18,92 |
| Dec 21 | 3,00 | 1,00 | 0,75 | 1,83 | 45,91 | 16,08 | 16,50 | 20,76 |

**Table A11**
Cumulative fuel consumption in tonnes for the 2022–2050 simulation.

| Year | Benchmark | MIN | AVG L-S | NCLD |
|---|---|---|---|---|
| 2022 | 137 | 106 | 106 | 132 |
| 2023 | 378 | 295 | 292 | 340 |
| 2024 | 611 | 493 | 473 | 534 |
| 2025 | 842 | 683 | 666 | 729 |
| 2026 | 1070 | 878 | 855 | 928 |
| 2027 | 1301 | 1071 | 1046 | 1130 |
| 2028 | 1534 | 1274 | 1240 | 1337 |
| 2029 | 1750 | 1472 | 1441 | 1543 |
| 2030 | 1970 | 1679 | 1653 | 1753 |
| 2031 | 2190 | 1877 | 1855 | 1962 |
| 2032 | 2398 | 2075 | 2060 | 2178 |
| 2033 | 2611 | 2270 | 2269 | 2389 |
| 2034 | 2837 | 2471 | 2476 | 2612 |
| 2035 | 3055 | 2672 | 2686 | 2822 |
| 2036 | 3266 | 2876 | 2897 | 3038 |
| 2037 | 3476 | 3086 | 3102 | 3253 |
| 2038 | 3685 | 3294 | 3302 | 3483 |
| 2039 | 3897 | 3503 | 3513 | 3701 |
| 2040 | 4111 | 3712 | 3717 | 3930 |
| 2041 | 4325 | 3913 | 3914 | 4154 |
| 2042 | 4533 | 4115 | 4113 | 4396 |
| 2043 | 4744 | 4320 | 4318 | 4626 |
| 2044 | 4959 | 4532 | 4511 | 4850 |
| 2045 | 5169 | 4736 | 4712 | 5083 |
| 2046 | 5377 | 4938 | 4913 | 5315 |
| 2047 | 5587 | 5134 | 5112 | 5543 |
| 2048 | 5799 | 5338 | 5316 | 5778 |
| 2049 | 6008 | 5540 | 5524 | 6009 |
| 2050 | 6222 | 5743 | 5725 | 6245 |

**Table A12**
Kg of $CO_2$ and total emissions produced by SISO operations for the 2021 simulation.

| | CO2 (kg) | | Total Emissions (kg) | |
|---|---|---|---|---|
| Policy | Monthly | Cumulative | Monthly | Cumulative |
| Benchmark | 12,205 | 146,460 | 12,499 | 149,008 |
| MIN | 4629 | 51,301 | 4741 | 52,536 |
| AVG L-S | 4519 | 52,630 | 4628 | 53,897 |
| NCLD | 5515 | 66,230 | 5648 | 67,825 |

Additionally, the distinct behavior and goal orientation of each SI policy suggests that they may be able to be combined and further optimized through the use of machine learning techniques. One potential enhancement could be also the incorporation of forecasting tools to accurately calculate shunt-in time based on the layout of the shunting yard. Furthermore, implementing effective communication channels that focus on optimizing short-term operations such as sorting and assembling railcars while avoiding scalability problems can help improve the effectiveness of the policies in a real-world setting.

**Table A13**
Cumulative tonnes of $CO_2$ per year for the 2022–2050 simulation.

| Year | Benchmark | MIN | AVG L-S | NCLD |
|---|---|---|---|---|
| 2022 | 438 | 337 | 338 | 421 |
| 2023 | 1206 | 941 | 932 | 1084 |
| 2024 | 1949 | 1571 | 1507 | 1704 |
| 2025 | 2687 | 2178 | 2123 | 2327 |
| 2026 | 3413 | 2801 | 2727 | 2960 |
| 2027 | 4149 | 3417 | 3336 | 3605 |
| 2028 | 4892 | 4063 | 3956 | 4264 |
| 2029 | 5584 | 4697 | 4596 | 4924 |
| 2030 | 6283 | 5356 | 5274 | 5591 |
| 2031 | 6986 | 5988 | 5918 | 6259 |
| 2032 | 7650 | 6620 | 6572 | 6949 |
| 2033 | 8329 | 7241 | 7239 | 7620 |
| 2034 | 9049 | 7881 | 7897 | 8332 |
| 2035 | 9746 | 8524 | 8568 | 9002 |
| 2036 | 10,417 | 9175 | 9243 | 9690 |
| 2037 | 11,087 | 9844 | 9894 | 10,376 |
| 2038 | 11,754 | 10,509 | 10,532 | 11,112 |
| 2039 | 12,432 | 11,175 | 11,207 | 11,806 |
| 2040 | 13,114 | 11,841 | 11,859 | 12,535 |
| 2041 | 13,796 | 12,481 | 12,487 | 13,250 |
| 2042 | 14,460 | 13,127 | 13,120 | 14,025 |
| 2043 | 15,132 | 13,782 | 13,775 | 14,757 |
| 2044 | 15,818 | 14,457 | 14,391 | 15,470 |
| 2045 | 16,488 | 15,107 | 15,031 | 16,214 |
| 2046 | 17,154 | 15,753 | 15,671 | 16,955 |
| 2047 | 17,823 | 16,377 | 16,306 | 17,683 |
| 2048 | 18,498 | 17,028 | 16,959 | 18,432 |
| 2049 | 19,164 | 17,672 | 17,622 | 19,170 |
| 2050 | 19,849 | 18,320 | 18,263 | 19,921 |

## CRediT authorship contribution statement

**Tommaso Bosi:** . **Federico Bigi:** Conceptualization, Formal analysis, Methodology, Validation, Writing – original draft, Writing – review & editing. **Andrea D'Ariano:** Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Francesco Viti:** Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Juan Pineda-Jaramillo:** .

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

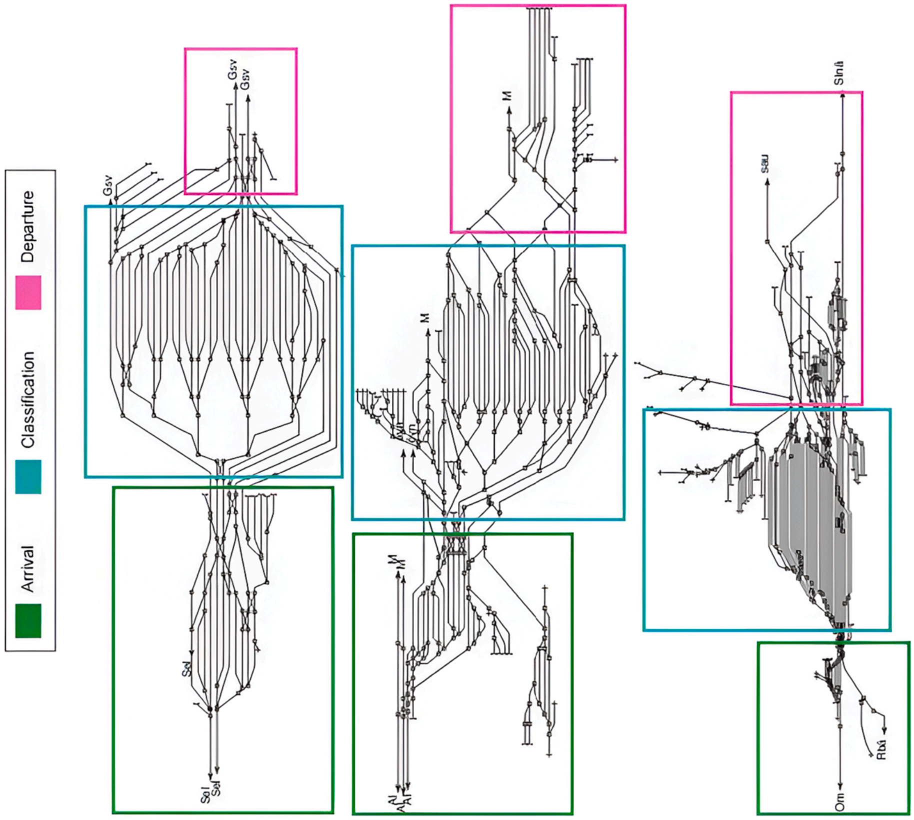


**Fig. C1.** Some examples of shunting yard layouts with the main areas highlighted (Bohlin et al., 2016).

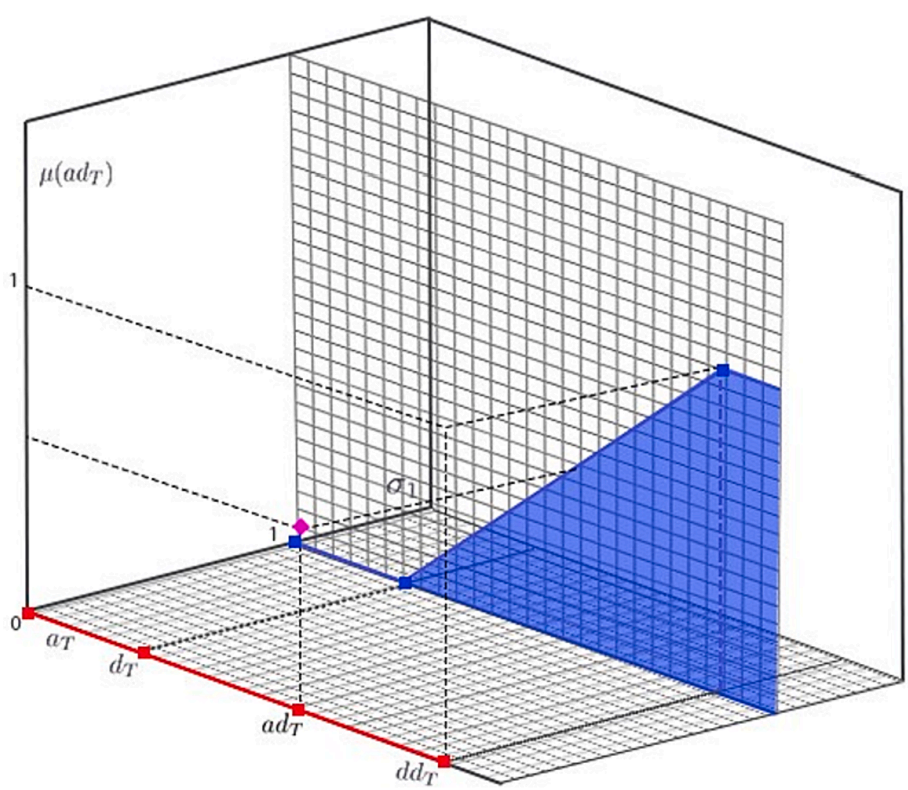


**Fig. C2.** Non-linear function of μ($ad_T$). Based on $\sigma_1$ the value will lay on the red or blue function.

**Table C1**
Propositions and logical connectives explaining the cause-and-effect relationship between the MINLP output and the long-term KPIs.

| Proposition | Logical Connective |
|---|---|
| If the clustering rate $C_r$ is high, the total number of shunts $N_s$ and emissions $E_{tot}$ are low. | $C_r \rightarrow (\ N_s \wedge\ E_{tot})$ |
| If the clustering rate $C_r$ is high, it is possible that the average delay per month/year $D_a$ will be high. | $C_r \rightarrow \diamond D_a$ |
| If the clustering rate $C_r$ is high, it is possible that the average percentage of train cancellations $C_t$ and/or the probability of an unfeasible state of the shunting yard $p(U_s)$ will be high. | $C_r \rightarrow \diamond (C_t \vee p(U_s)\ )$ |
| The shunting policy $S_p$ affects both the clustering rate $C_r$, the wagon fleet size $F_s$, the average delay per month/year $D_a$ and the average percentage of train cancellations $C_t$. | $S_p \rightarrow (C_r \wedge F_s \wedge D_a \wedge C_t)$ |

## Data availability

The data that has been used is confidential.

## Acknowledgements

This research was supported by the Société's nationale des chemins de fer luxembourgeois (CFL), Roma Tre and Luxembourg Universities. We thank supervisors and colleagues from our respective Engineering Departments who provided insight and expertise that greatly assisted the research.

## Appendix A

Tables

### *A.1 Data and results*

This section presents a comprehensive overview of the data and results obtained from the simulations. In particular, the fleet size and usage distribution of the wagons are detailed in Tables A.1 and A.2. The performance of the number of shunting operations and clustering rate trends are presented in Tables A.3, A.4, and A.5. Tables A.6 and A.7 provides the results on delays and train cancellations, while Tables A.8 and A.9 provide detailed information on the key performance indicators related to emissions. Tables A.10, A.11, A.12, and A.13 provide information on the average and total emissions produced by shunting locomotives.

## Appendix B. DIJKSTRA ALGORITHM

### *B.1 Pseudo-Code*

This section presents the pseudo-code behind the Dijkstra algorithm. The Dijkstra algorithm is a widely used algorithm for finding the shortest path between two nodes in a graph. The algorithm uses a priority queue to select the next node to visit, where the priority is determined by the distance from the starting node. The algorithm starts at the starting node and visits each neighboring node, updating the distance to each neighbor if a shorter path is found. In the context of the paper, the Dijkstra algorithm is applied to find the shortest routes, based on weighted arcs and nodes, between two consecutive mandatory shunts. The shortest route gives information on the clusters of shunts created.

Algorithm 1. Dijkstra Algorithm.

```
distance[s] ← 0
for v in V do
distance[v] ← +∞
previous[v] ← NULL
add v to priority queue Q
end for
while Q is not Empty do
u ← extract MIN{dist(v)} from Q
for each unvisited neighbor v of u do
tempDistance ← distance[u] + weight(u,v)
if tempDistance < distance[v] then
distance[V] ← tempDistance
previous[V ] ← u
end if
end for
end while
return distance[], previous[]
```

## Appendix C

Insights.

### C.1 The shunting yard

The typical layout of a shunting yard can be roughly separated into three main areas, which all consist of a set of parallel tracks (Boysen et al., 2016):

- the receiving area;
- the classification area;
- the departure area.

In the receiving area, inbound trains are temporarily parked while awaiting humping and disassembly. The arrival train is queued on one of the tracks, afterwards, its wagons are inspected and each of them is labeled with the proper destination code. Based on the priority policy, each inbound train, or a subset of its wagons, is pushed by a shunting locomotive over the hump hill. While rolling down the hill, switches direct the subset of wagons to their planned classification tracks. Outbound trains are then pulled to departure tracks to be inspected and, once road engines have been attached, they are ready for departure (See Fig. C1).

### C.2 The SI model

The SI problem can be formulated as a *General Set Partitioning* problem (GSPP), specifically, as a *Crew Scheduling* version, where each job node $\in R$ (wagon shunted out by the SO sub-model) must be assigned with a crew node $\in A$ (shunting yard's wagon suitable for the outbound train's next trip) to minimize the overall cost on arcs while covering all the jobs (as shown in Fig. 1b). Furthermore, each inbound train's wagon shunted out with a demand or an optional shunt will occupy the position of the wagon *i*-th replacing it, while the ones shunted out with a maintenance shunt will be moved to the workshop.

#### C.2.1 Matrix of arc Costs$C_S$

For the SI problem, we define a cost matrix $C_S$, where each entry $c_{S_{i,j}}$ represents the cost to replace the inbound train's wagon *j*-th with the shunting yard's wagon *i*-th. This cost can be assumed as either a time, an economic cost, or a combination of different parameters based on the user's convenience. The construction of the $C_S$ matrix comes from the topology of the shunting yard and each cost $c_{S_{i,j}}$ is computed based on a *Train Unit Shunting* problem (TUSP) model (Kamenga et al., 2021). Since the TUSP is NP-Hard, a quite effective solution (that also allows us to avoid holistic models) is to exploit a number $l_I$ of Gaussian distributions $N_j\left(\mu_j, \sigma_j^2\right)$ based on means and standard deviations provided by practitioners.

*Nomenclature*

#### 4.1.1. Sets

- $R \subseteq T$, sub-set of inbound train's wagons which must be shunted out
- $A \subseteq S$, set of shunting yard's wagons suitable to perform $r^T$, namely, the kilometers the outbound train will perform during its next trip
- $K$, set of wagon types
- $C_S$, matrix of cost to shunt in wagons $\in A$
- $c_{S_{i,j}}$, integer value expressing the cost to replace the wagon *j*-th $\in R$ with the wagon *i*-th $\in A$
- $type_k$, integer value expressing the number of wagons of type k required by the outbound train's composition and related to sub-set $A$
- $type_{S_{i,k}} = \begin{cases} 1, & \textit{if the wagon } i-th \in A \\ & \textit{is of type } k \\ 0, & \textit{otherwise} \end{cases}$
- $z_{i,j} = \begin{cases} 1, & \textit{if on the wagon } j-th \in R \\ & \textit{is replaced by the wagon} \\ & i-th \in A \\ 0, & \textit{otherwise} \end{cases}$

#### C.2.2 Mathematical model

The SI model aims to minimize the costs given by replacing each shunted-out wagon with suitable wagons inside the shunting yard, as represented by the objective function (C.1).

$$\sum\nolimits_{i\in A}\sum\nolimits_{j\in R} c_{S_{i,j}} z_{i,j} \tag{C.1}$$

The following constraints (C.2)-(C.4) consider both the classical ones of a Set Partitioning problem and other constraints useful to fit the demand requested by the SO problem.

$$\sum\nolimits_{j\in R} z_{i,j} = 1 \forall i \in A \tag{C.2}$$

Constraint (C.2) states that each shunted-out wagon must be replaced by exactly one shunting yard's suitable wagon.

$$\sum\nolimits_{i\in A} z_{i,j} \leq 1 \; \forall j \in R \tag{C.3}$$

Constraint (C.3) ensures that each suitable wagon *i*-th in the shunting yard can replace no more than one shunted-out wagon.

$$\sum\nolimits_{i\in A}\sum\nolimits_{j\in R} type_{S_{i,k}} z_{i,j} - type_k = 0 \; \forall k \in K \tag{C.4}$$

The last constraint (C.4) expresses the necessity to fit the composition of the outbound train. We need to ensure the demand is correctly fulfilled regardless of the position in which we put $i$.

*C.3 Delay function and linearization*

In this section, we deal with the outbound train's departure delay function, expressed by $\mu(ad_T)$ (Fig. C2). The delay function is a piece-wise linear function described by the following equations:

$$\mu(ad_T) = \begin{cases} 0, if ad_T \leq d_T \\ \dfrac{ad_T - d_T}{dd_T - d_T}, if dd_T \geq ad_T > d_T \\ 1, if ad_T > dd_T \end{cases} \tag{C.5}$$

Based on the actual outbound train's departure time $ad_T$, μ will assume values ranging between 0 and 1: it will be equal to 0 if the outbound train is on time and equal to $\frac{ad_T - d_T}{dd_T - d_T}$ if the latter is late, till its cancellation, where μ will assume a value equal to 1. To consider this function in the model, at first, we have used the $\sigma$ variables as follows:

$$\mu(ad_T) = \frac{ad_T - d_T}{dd_T - d_T}\sigma_1 + \sigma_2 \tag{C.6}$$

That way, as described in section 4.1, $\sigma_1$ would activate or cancel the fraction based on $\sigma_2$ value. Nevertheless, the term $\frac{ad_T - d_T}{dd_T - d_T}\sigma_1$ is a non-linear convex function that transforms the MILP model into a MINLP one (Asghari et al., 2022) (Sahinidis, 2018) (Fig. C2).

Thus, to handle the non-linearity we have moved the equation from constraints to the objective function, introducing $\sigma_3$ and constraints (4.4)-(4.13) related. That way we have linearized the function.

*C.4 Comprehensive view of the MINLP model*

The objective function and all the constraints of the MINLP model are listed below.

$$min\sum_{i\in\mathcal{T}}\gamma_i - \sum_{i=1}^{|\mathcal{T}|-1} adj_{i,i+1} + \frac{|\mathcal{T}|}{2}(\sigma_2 + \sigma_3) + \frac{|\mathcal{T}|}{4}\left(\sum_{i\in\mathcal{T}}\sum_{j\in S}(\alpha W_1 + \beta W_2)z_{i,j} + \sum_{i\in\mathcal{T}} y_i c_{u_i}\right) \tag{4.3}$$

$$ad_T \leq d_T + \sigma_1 M + \sigma_2 M \tag{4.4}$$

$$dd_T + (1-\sigma_1)M \geqslant ad_T \tag{4.5}$$

$$ad_T > d_T - (1-\sigma_1)M \tag{4.6}$$

$$ad_T \leqslant dd_T + \sigma_2 M \tag{4.7}$$

$$ad_T > dd_T - (1-\sigma_2)M \tag{4.8}$$

$$\sigma_1 + \sigma_2 \leq 1 \tag{4.9}$$

$$\sigma_3 \leq \sigma_1 M \tag{4.10}$$

$$\sigma_3 \geq \frac{a_T - d_T}{dd_T - d_T}\sigma_1 \tag{4.11}$$

$$\sigma_3 \leq \frac{ad_T - d_T}{dd_T - d_T} + (1-\sigma_1)M \tag{4.12}$$

$$\sigma_3 \geq \frac{ad_T - d_T}{dd_T - d_T} - (1-\sigma_1)M \tag{4.13}$$

$$a_T + \sum_{i\in\mathcal{T}}\gamma_i ts + \sum_{i\in\mathcal{T}}\sum_{j\in S} c_{s_{i,j}} z_{i,j} = ad_T \tag{4.14}$$

$$\frac{\left(a_T + \sum_{i\in\mathcal{T}}\gamma_i ts\right)}{dd_T} = \alpha \tag{4.15}$$

$$1 - \alpha = \beta \tag{4.16}$$

$$y_i \geq \frac{m_i + r_T}{m_{max_i}} - 1 - \left(\sum_{k\in K} x_{i,k}\right) M \forall i \in \mathcal{T} \tag{4.17}$$

$$y_i \leq \left(1 - \sum_{k\in K} x_{i,k}\right)\frac{m_i + r_T}{m_{min_i}} \ \forall i \in \mathcal{T} \tag{4.18}$$

$$\sum_{i\in\mathcal{T}:type_i\neq type_r} x_{i,type_r} = rise \tag{4.19}$$

$$\sum_{k\in K} x_{i,k} = 0 \forall i \in \mathscr{T} : type_i = type_r \tag{4.21}$$

$$\sum_{i\in \mathscr{T}} x_{i,k} = 0 \forall k \in K : k \neq type_r \tag{4.22}$$

$$\sum_{j\in S: type_{S_j} = type_i} z_{i,j} = y_i \quad \forall i \in \mathscr{T} \tag{4.24}$$

$$\sum_{j\in S: type_{S_j} \neq type_i} z_{i,j} = x_{i,type_r} \quad \forall i \in \mathscr{T} \tag{4.25}$$

$$z_{i,j} \leq 2 - \frac{ms_j + r_T}{ms_{max_j}} \quad \forall i \in \mathscr{T}, \forall j \in S \tag{4.26}$$

$$\sum_{i\in \mathscr{T}} z_{i,j} \leq 1 \forall j \in S \tag{4.27}$$

$$\sum_{k\in K} x_{i,k} + y_i = \gamma_i \quad \forall i \in \mathscr{T} \tag{4.28}$$

$$2adj_{i,i+1} \leq \gamma_i + \gamma_{i+1} \quad \forall i = 1, \cdots, |\mathscr{T}| - 1 \tag{4.29}$$

$$\sum_{j\in S} z_{i,j} code_{S_j} + (1-\gamma_i) code_i = code_{O_i} \quad \forall i \in \mathscr{T} \tag{4.30}$$

### C.5 Dijkstra algorithm description

The Dijkstra algorithm is one of the most famous algorithms for the *Shortest Path Problem* (SP). Given a graph $G(N,A)$ with positive weights and a source node $s$, it finds the shortest paths from $s$ to all the other nodes within $G$ (*Shortest Path Tree*, SPT). The initialization considers three terms:

- *dist* is the array of distances from the source node $s$ to each node in the graph, where $dist(s) = 0$ and for all other nodes $v$, $dist(v) = +\infty$;
- $Q$ is the queue of all nodes in the graph. At the end of the algorithm's run, $Q$ will be empty;
- $S$ is the set of nodes the algorithm has visited. At the end of the algorithm's run, $S$ will contain all the nodes of the graph.

Then, the algorithm performs the following steps:

- While $Q$ is not empty, pop the node $v$, which is not already in $S$, from $Q$ with the smallest distance from the source node;
- Add the node $v$ to $S$, and update the distance values of its adjacent nodes as follows;
- For each new adjacent node $u$, if $dist(v) + weight(u,v) < dist(u)$, update $dist(u)$ with the new minimal distance value. Otherwise, no updates are made to $dist(u)$.

The Dijkstra Algorithm's pseudocode can be found in *Appendix B*.

### C.6 Input data, model solution structure, and its relation with the long-term KPIs

The event-based simulation environment is responsible for managing the time, space, and objects within the shunting yard, and it performs the following tasks:

1. One of the shunt-in policies is selected and a class is created based on the policy's requirements;
2. The total operational time is computed, which is the amount of time available for the shunting operations between the train's current arrival time in the yard and the scheduled deadline before the train is canceled;
3. The following inputs are gathered and used to create an instance for the model:

- Inbound train composition
- – Train length (in terms of wagons)
- – Wagon type
- – Wagon current mileage
- – Wagon maximum mileage constraint before maintenance
- Outbound train
- – Train length (in terms of wagons)
- – Number of wagons requested for each type
- Total operational time
- – Considering all the operations to be performed on the way
- * All the dispatching operations provided by the practitioners
- * Estimated time for loading the train based on the forecast demand
- Wagons available in the Shunting yard

– Wagon type
– Wagon current mileage
– Wagon maximum mileage constraint before maintenance
– Wagon position and shunt-in time

The MINLP model solution structure consists of multiple arrays, matrices, and scalars expressing the decision on the wagons to be shunted out and shunted in and the delays or train cancellations that occurred. Each of these outputs is useful to perform events inside the virtual environment as well as to keep track of the impact on the long-term KPIs. Once the MINLP model is run the following outputs are provided to the simulator:

- The array of dimension $|\mathcal{T}|$ of the positions of the wagons to be shunted out due to maintenance or optional shunts (e.g. $y = [0, 1, 0, 1]$). On this array, a further filter is performed, by looking at the actual mileage and the leasing contract constraints, in order to distinguish which wagon should go in the workshop and which one on the shunting yard;
- The matrix of dimension $|\mathcal{T}| \times |K|$ of the *i*-th position of the wagon to be shunted out due to demand-matching shunt and the *k*-th wagon type that will replace it (e.g. $X^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$);
- The matrix of dimension $|\mathcal{T}| \times |S|$ of the *i*-th position of the wagon shunted out, regardless of the shunt type, and the *j*-th shunting yard's wagon that will replace it (e.g. if there are three wagons in the shunting yard $Z = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$);
- The array of dimension $|\mathcal{T}|$ of the wagons' identification codes with respect to the wagon position ($code_O = [101, 180, 107, 113]$);
- The float value, expressing the shunting time required;
- The integer value, expressing the total number of shunts to be performed, considering the clustering assumption for the shunt-out operations and the sum of the times required by each shunting yard wagons shunted in for the shunt-in operations.

Furthermore, if the MINLP model can't find a feasible solution, this means that either the operational time available is not enough or the shunting yard is running out of suitable wagons. Then, the train requests a cancellation, the train cancellation counter is updated and the wagons to be sent for maintenance don't go into maintenance until the next service, where the model is run again with the same train but a new demand instance. Each wagon object can have attributes such as the distance covered before maintenance, the number of maintenance performed, and so on. These attributes can be updated whenever new data is available, providing a comprehensive view of the performance of the railway system. The collected data can be stored in databases, allowing for further analysis and decision-making.

Object-oriented programming (OOP) allows us to manage each train, track, service, and so on as objects with multiple attributes and methods. These attributes store data that can be used to track KPIs over time. For each train processed by the MINLP model, unless a train cancellation has occurred, the outputs are stored in databases:

- The shunting yard object tracks the total number of shunting operations performed both on a monthly and cumulative basis. Additionally, whenever a previously unused wagon is activated to cover a service, the shunting yard object updates the attribute related to the size of the active wagon fleet;
- Similarly, the train object has attributes such as the average delay performed and the number of delays and cancellations, both on a monthly and cumulative basis. Every time a train is shunted, these values are updated based on the output of the optimization model;
- The wagon object keeps its attributes, such as the distance covered before maintenance and the number of maintenance performed, updated such that the collected data can be used for further analysis and decision-making.

The proposed model is designed to optimize short-term operational efficiency by addressing the clustering problem related to shunt-out operations and maintenance constraints, as well as the shunt-in operations performed based on demand-matching constraints and the policy version applied to the objective function. Nevertheless, It is important to note that when a shunting operation is performed, it will affect all the shunting operations performed on successive single trains. Indeed, while the shunt-out policy is strongly oriented to the assumptions on clustering and the maintenance constraints stated by the wagon leasing contract, the shunt-in policies work on the single train to act on the long term and impose a behavior that seems reasonable with respect to the constraints and objective. For the shunt-out operations, it is reasonable to assume that will be a positive correlation between the clustering rate and the monthly or yearly number of shunts performed and emissions, as well as a negative correlation between the clustering rate and the average delays per month/year, and the percentage of train cancellations or the probability of unfeasibility of the shunting yard capacity in the long term. For the shunt-in operations, the policy decision criteria applied will impact both the clustering rate, delays, and the wagon fleet size.

The following logical connectives can be assumed:

See Table C1

## References


Agency, E. E. (2021). *Greenhouse gas emissions from transport in Europe*. European Environment Agency.

Commission, E. (2015). *Shift2Rail strategic master plan*. European Commission.

Kearns, S., Albrecht, H., D'Ariano, A., Franco, G., & Tzanidaki, J. (2016). *Network and traffic management systems*. Strategic Transport Research and Innovation Agenda (STRIA).

Pagand, I., Carr, C., & Doppelbauer, J. (2020). *Fostering the railway sector through the european green deal*. European Union Agency For Railways.

Boysen, N., Emde, S., & Fliedner, M. (2016). The basic train makeup problem in shunting yards. *OR Spectrum, 38*, 207–233.

Bohlin, M., Gestrelius, S., Dahms, F., Mihalak, M., & Flier, H. (2016). Optimization methods for multistage freight train formation. *Transportation Science, 50*, 823–840.

Guglielminetti, P., Lagraulet, M. F., Artuso, D., Lunadei, E., Musso, A., Fusco, G., Licciardello, R., Piccioni, C., & Seno, F. (2015). Study on single wagonload traffic in Europe – Challenges, prospects and policy options. *European Commission*.

Bohlin, M., and Hansmann, R., (2018) .Optimization of Railway Freight Shunting, *Handbook of Optimization in the Railway Industry*, pp. 181–212.

Deleplanque, S., Hosteins, P., Pellegrini, P., & Rodriguez, J. (2022). Train management in freight shunting yards: Formalisation and literature review. *IET Intelligent Transport Systems, 16*, 1286–1305.

Jaehn, F., Rieder, J., & Wiehl, A. (2015). Single-stage shunting minimizing weighted departure times. *Omega, 52*, 133–141.

Meinert, M., Prenleloup, P., Schmid, S., & Palacin, R. (2015). Energy storage technologies and hybrid architectures for specific diesel-driven rail duty cycles: Design and system integration aspects. *Applied Energy, 157*, 619–629.

Denari, R., & Derossi, M. (2019). An innovative maintenance methodology for freight wagons combining proportional hazard model and life cycle cost. *Polimi*.

Van Den Broek, R., Hoogeveen, H., Van Den Akker, M., & Huisman, B. (2021). A local search algorithm for train unit shunting with service scheduling. *Transportation Science, 56*, 141–161.

Lin, B., Liu, S., Lin, R., Wang, J., Sun, M., Wang, X., Liu, C., Wu, J., & Xiao, J. (2019). The location-allocation model for multi-classification-yard location problem. *Transportation Research Part E: Logistics and Transportation Review, 122*, 283–308.

Zhang, Y., Song, R., He, S., Li, H., & Guo, X. (2018). Optimization of classification track assignment considering block sequence at train marshalling yard. *Journal of Advanced Transportation*.

Boysen, N., Scholl, J., & Stephan, K. (2017). When road trains supply freight trains: Scheduling the container loading process by gantry crane between multi-trailer trucks and freight trains. *OR Spectrum, 39*, 137–164.

Murali, P., Ordóñez, F., & Dessouky, M. M. (2016). Modeling strategies for effectively routing freight trains through complex networks. *Transportation Research Part C: Emerging Technologies, 70*, 197–213.

Falsafain, H., & Tamannaei, M. (2019). A novel dynamic programming approach to the train marshalling problem. *IEEE Transactions on Intelligent Transportation Systems, 21*, 701–710.

Xu, X., & Dessouky, M. M. (2022). Train shunting with service scheduling in a high-speed railway depot. *Transportation Research Part C: Emerging Technologies, 143*, Article 103819.

Jaehn, F., Rieder, J., & Wiehl, A. (2015). Minimizing delays in a shunting yard. *OR Spectrum, 37*, 407–429.

Lin, B. and Lin, R., (2017). An approach to the high-level maintenance planning for emu trains based on simulated annealing, *arXiv*.

Luan, X., Miao, J., Meng, L., Corman, F., & Lodewjiks, G. (2017). Integrated optimization on train scheduling and preventive maintenance time slots planning. *Transportation Research Part C: Emerging Technologies, 80*, 329–359.

Herr, N., Nicod, J., Varnier, C., Zerhouni, N., Herr, N., Nicod, J., Varnier, C., Zerhouni, N., & Cherif, M. (2017). Joint optimization of train assignment and predictive maintenance scheduling. *International Conference on Railway Operations Modelling and Analysis*.

Gerum, P. C. L., Altay, A., & Baykal-Gürsoy, M. (2019). Data-driven predictive maintenance scheduling policies for railways. *Transportation Research Part C: Emerging Technologies, 107*, 137–154.

Zagorskikh, A. A., Alsova, O. K., & Gavrilov, A. V. (2020). A simulation model of the process of organizing the movement and maintenance of trains at a railway station. *Journal of Physics: Conference Series, 1661*.

Chuijiang, G. (2021). Optimization model and algorithm for the placing-in and taking-out of wagons in railway stations with branch-shaped goods operating sites. *International Journal of Rail Transportation, 9*, 579–594.

Adlbrecht, J., Hüttler, B., Zazgornik, J., Gronalt, M., (2015). The train marshalling by a single shunting engine problem, *Transportation Research Part C: Emerging Technologies*, vol. 58, pt. A, pp. 56-72.

Di Loreto, S., Malikova, L., Tomekova, I., Mirkovis, T., Ohlin, M., & Maliacek, J. (2018). Overview of priority rules in operation. *RailNetEurope*.

Cabral, M. O., Shires, J., Wardman, M., Teklu, F., & Harris, N. (2021). The use of recovery time in timetables: Rail passengers' preferences and valuation relative to travel time and delays. *Transportation, 48*, 337–368.

Jovanovic, P., Kecman, P., Bojovic, N., & Mandic, D. (2017). Optimal allocation of buffer times to increase train schedule robustness. *European Journal of Operational Research, 256*, 44–54.

Corman, F., D'Ariano, A., Marra, A., Pacciarelli, D., & Samà, M. (2017). Integrating train scheduling and delay management in real-time railway traffic control. *Transportation Research Part E: Logistics and Transportation Review, 105*, 213–239.

Islam, D. M. Z., Laparidou, K., & Burgess, A. (2016). Cost effective future derailment mitigation techniques for rail freight traffic management in Europe. *Transportation Research Part C: Emerging Technologies, 70*, 185–196.

Navajas Cawood, E., Rotoli, F., & Soria, A. (2016). *Capacity assessment of railway infrastructure : Tools, methodologies and policy relevance in the eu context*. Publications Office.

Cao, Y., Zhang, Z., Cheng, F., & Su, S. (2022). Trajectory optimization for high-speed trains via a mixed integer linear programming approach. *IEEE Transactions on Intelligent Transportation Systems, 23*, 17666–17676.

Wang, P., Goverde, R. M. P., & van Luipen, J. (2019). A connected driver advisory system framework for merging freight trains. *Transportation Research Part C: Emerging Technologies, 105*, 203–221.

Ursavas, E., & Zhu, S. X. (2017). Integrated passenger and freight train planning on shared-use corridors. *Transportation Science, 6*, 1376–1390.

Kurhan, M. and Kurhan, D., (2018.) Problems of providing international railway transport, *MATEC – Web of Conferences*, vol. 230, 2018.

P. Daszkiewicz and M. Andrzejewski, "Preliminary analyzes in terms of the possibility of reducing energy consumption by the sm42 locomotive used in track works," *MATEC Web of Conferences*, vol. 118, 2017.

Feo-Valero, M., Garcia-Menendez, L., & del Saz-Salazar, S. (2016). Rail freight transport and demand requirements: An analysis of attribute cut-offs through a stated preference experiment. *Transportation, 43*, 101–122.

Mo, P., Yang, L., D'Ariano, A., Yin, J., Yao, Y., & Gao, Z. (2020). Energy-efficient train scheduling and rolling stock circulation planning in a metro line: A linear programming approach. *IEEE Transactions on Intelligent Transportation Systems, 21*(9), 3621–3633.

Merkisz, J., Fuc, P., Lijewski, P., Ziolkowski, A., Galant, M., & Siedlecki, M. (2016). Analysis of an increase in the efficiency of a spark ignition engine through the application of an automotive thermoelectric generator. *Journal of Electronic Materials, 45*, 4028–4037.

Rymaniak, L., Ziolkowski, A., Gallas, D., (2017). Particle number and particulate mass emissions of heavy duty vehicles in real operating conditions, *MATEC Web of Conferences*, vol. 118.

Wang, D., Zhao, J., & Peng, Q. (2022). Optimizing the loaded train combination problem at a heavy-haul marshalling station. *Transportation Research Part E: Logistics and Transportation Review, 162*, Article 102717.

Giacco, G. L., Carillo, D., D'Ariano, A., Pacciarelli, D., & Marìn, G. (2014). Short-term rail rolling stock rostering and maintenance scheduling. *Transportation Research Procedia, 3*, 651–659.

Patro, S., Sahu, K.,(2015). Normalization: A preprocessing stage, *arXiv*.

Wang, H., Mao, W., & Eriksson, L. (2019). A three-dimensional dijkstra's algorithm for multi-objective ship voyage optimization. *Ocean Engineering, 186*.

Casas, P., Kraus, L., Sedeno-Noda, A., Borndorfer, R., (2021). Targeted multiobjective dijkstra algorithm, *arXiv*.

Birol, F. (2021). *Net zero by 2050 – A roadmap for the global energy sector*. International Energy Agency.

Wang, X., & Disney, S. M. (2016). The bullwhip effect: Progress, trends and directions. *European Journal of Operational Research, 250*, 691–701.

Barbour, W., Martinez Mori, J. C., Kuppa, S., & Work, D. B. (2018). Prediction of arrival times of freight traffic on US railroads using support vector regression. *Transportation Research Part C: Emerging Technologies, 93*, 211–227.

Pineda-Jaramillo, J., & Viti, F. (2023). Identifying the rail operating features associated to intermodal freight rail operation delays. *Transportation Research Part C: Emerging Technologies, 147*, Article 103993.

Rymaniak, L., Daszkiewicz, P., Merkisz, J., & Bolzhelarskyi, Y. (2019). Method of determining the locomotive engine specific fuel consumption based on its operating conditions. *AIP Conference Proceedings* (vol,, 2078.

Biemann, K., & Notter, B. (2023). Environmental Methodology and Data. *EcoTransIT World (EWI)*.

F. Kamenga, P. Pellegrini, J. Rodriguez, and B. Merabet, "Solution algorithms for the generalized train unit shunting problem," *EURO Journal on Transportation and Logistics*, vol. 10, 2021.

Asghari, M., Fathollahi-Fard, A. M., Al-e-hashem, S. M. J., & Dulebenets, M. A. (2022). Transformation and linearization techniques in optimization: A state-of-the-art survey. *Mathematics, 10*, 283.

Sahinidis, N. V. (2018). Mixed-integer nonlinear programming. *Optimization and Engineering, 20*, 301–306.