The Shunt-In Shunt-Out Problem in Rail Freight Transport: an Event-Based Simulation Framework for Sustainable Rolling Stock Management

Tommaso Bosi [a,1], Federico Bigi [b], Andrea D'Ariano [a], Francesco Viti [b]
[a]Department of Civil, Computer Science and Aeronautical Technologies Engineering,
University of Roma Tre
00154, Rome, Italy
[1] E-mail: tommaso.bosi@uniroma3.it, Phone: +39 392 984 8809
[b] Faculty of Science, Technology and Medicine (FSTM), University of Luxembourg
L-4364, Esch-Sur-Alzette, Luxembourg

The type of a submission

Type A: Research paper.

Abstract
The European Union plans to double the freight rail traffic by 2050 both to cut pollution emissions and to mitigate congestion by shifting traffic from road to rail networks. One of the challenges is to minimize emissions and high costs associated with shunting yard operations while maintaining an acceptable service level. In this context, we propose an Event-Based Simulation Framework for wagon Shunt-in and Shunt-out operations. The Event-Based Simulation Framework exploits programming tools and a MILP model to minimize the number of Shunt-in and Shunt-out operations performed and, consequently, both strategic and tactical objectives such as the clustering rate, the wagon fleet size, departure delays and emissions of shunting locomotives. Several versions of the MILP model are described based on the Shunt-in policy applied. Each Shunt-in policy has different criteria for wagon's choice and has shown a strong goal orientation. To test the MILP model's effectiveness, we have considered short and long-term real train timetables for freight trains in the Bettemburg Eurohub Sud Terminal (Luxembourg) and we have assessed different KPIs linked to tactical and strategic objectives. Computational results show that the criteria for choosing which wagons should be taken-out from the inbound train and should be inserted into the outbound train might significantly impact multiple rail system KPIs analyzed. The Event-Based Simulation Framework is part of the ANTOINE national project financed by CFL (Chemins de fer luxembourgeois) and is considered an add-on tool to Shunty, an industrial software project for rail decision-makers.

Keywords
Freight Rail Operation, Shunting Operation, Rolling Stock Maintenance, Multimodal Transport Systems, Decision Support System.

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 main 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, 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. Some of these costs are related to operations performed inside shunting yards, namely, $shunting\ operations$.

1.1 SISO Operations and Impacts

The shunting operation plays a crucial role in daily railway operations and requires careful management of resources. The $Shunt\text{-}In\ Shunt\text{-}Out\ Problem$ (SISO) involves assigning wagons from a heterogeneous fleet to timetabled services in order to meet the wagon demand, considering factors such as the composition and final destination of each departing train. Additionally, the SISO operations can take into account the condition-based maintenance requirements specified by the leasing contracts for each individual wagon. This upstream problem of the shunting yard *Classification Problem* consists both of defining the selection criteria we take out wagons from the inbound train due to condition-based maintenance and demand-matching constraints ($Shunt\text{-}Out$, SO), and of replacing them with shunting yard's wagons in order to make up the outbound train ($Shunt\text{-}In$, SI) while taking into account parameters such as the total number of SISO operations performed and the resulting overall operational costs, the time to shunt, and the shunting yard's supplies availability (Figure 1). The SISO optimization streamlines the shunting and maintenance activities by improving the *clustering rate*. This leads to more efficient cost management as it allows for adjacent wagons shunted out to be treated as a single entity, resulting in a single cost for the entire group. These operations must balance a range of strategic and tactical objectives, such as reducing the operational costs of shunting while adhering to maintenance constraints outlined in leasing contracts and wagon demand, minimizing departure delays to maintain acceptable service levels, avoiding cancellations due to missed deadlines or insufficient wagons, optimizing the size of the wagon fleet to reduce overhead costs, and reducing emissions from shunting locomotives powered by diesel engines. Our research proposes an Event-Based Simulation Framework (EBSF) dealing with the SISO problem as currently this, and other shunting yard issues, are solved based on practitioners' experience. The EBSF is an add-on tool for $Shunty$, an industrial software of the ANTOINE project, financed by the Chemins de Fer Luxembourgeois (CFL).

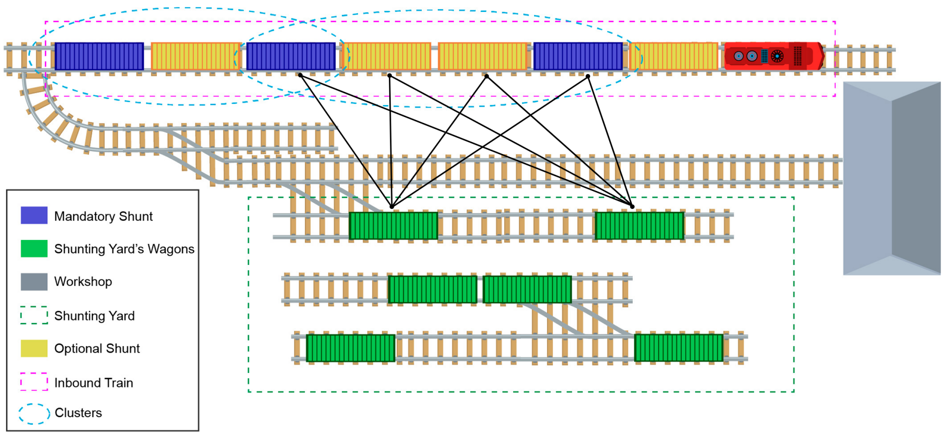


Figure 1: SISO problem representation. For the Shunt-Out problem, the blue circles are the possible clusters of shunts and the decision falls on activating optional shunts or not. In this case, the Shunt-In possible assignments have been considered for the second cluster and the green wagons closer to the train.

1.2 The Event-Based Simulation Framework

The proposed EBSF has been developed to incorporate Operations Research (OR) and programming tools, and process the timetable by simulating terminal operations and the passage of time. The built-in MILP model considers rolling stock maintenance and timetable constraints as well as a multi-component objective function aiming for the minimization of the number of shunts performed. The objective function further considers: weighted delay terms of outbound trains; the SI policy applied; weighted terms associated with shunting binary variables used to avoid a quick shortage of wagons in the shunting yard. For the SI problem, different policies are described, each of which is characterised by specific wagon assignment criteria. Due to their different assignment criteria, each SI policy has pros and cons, therefore, they should be used considering the goal that practitioners want to achieve. 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. The general idea is to provide a framework that can be easily implemented at zero cost and with multiple and new SI criteria, without prior knowledge of the shunting yard configuration, wagon fleet and so forth. The rest of this paper is structured as follows: Section 2 reviews the relevant literature on shunting yard issues and maintenance operations; Section 3 provides a formal description of the SISO Problem and states our assumptions; Section 4 describes the MILP model for the SISO problem; Section 5 shows the performance of each SI policy taking into account multiple KPIs; Section 6 summarizes the conclusions and suggests where to focus future research.

2 Literature Review

This section presents the relevant literature related to the SISO problem, focusing on: how shunting operations are performed in the freight rail industry; where wagon maintenance impacts and how it is generally performed; the research gap we address with our research.

2.1 Shunting Operations

The shunting operation is defined as the movement of one or multiple wagons within the shunting yard. These manoeuvres can be done for multiple reasons and are modelled in the literature through the *Rolling Stock Problem* (RSP) and the *Train Unit Shunting Problem* (TUSP). The RSP consists of planning a service time for each rolling stock and can be seen as the management of wagon and/or train units in order to reduce the cost to supply the services or fulfil the demand. Instead, the process of parking unused rolling stock units and the related manoeuvres within a shunting yard is called shunting, with the corresponding planning problem defined as the TUSP (Freling et al. (2005)). For Giacco et al. (2014), the management of rolling stock is the major cost factor for each railway company, and hence the greatest competitive advantage, since the quality of the service depends on it. Giacco et al. solved the RSP with a two-step approach that combines scheduling tasks related to train services, short-term maintenance operations, and empty runs. Usually, for shunting operations on both passenger and freight trains, the RSP is solved first, followed by the TUSP. Given the clear connection between these two problems, an integrated approach, the *Integrated Rolling Stock and Unit Shunting Problem* (IRSUSP), has been proposed by Haahr and Lusby (2017). Similarly, Li et al. (2020) implemented a simulation approach through multiple shunting policies aiming to reduce operating costs for the entry and exit system of wagons.

2.2 Wagon Maintenance

The cost of wagon maintenance is a significant factor in the rail freight system (Jaehn et al. (2015)). Wagons spend a significant amount of downtime undergoing maintenance and repairs, leading to overhead costs such as the need for a larger pool of wagons and variable costs such as storage fees. Maintenance can be triggered by mileage, time, or condition monitoring (Lin and Lin (2017)). Condition- or mileage-based maintenance is now more common due to the fact that rolling stock spends 70% of its time unused in the shunting yard, leading to additional, inefficient maintenance operations. To minimize costs, Budai et al. (2006) proposes a solution for the *Preventive Maintenance Scheduling Problem* using a modified scheduling problem and a greedy heuristic. Herr et al. (2017) solves both the *Rolling Stock Problem* and maintenance scheduling for passenger trains by considering both preventive maintenance scheduling and degradation based on the distance traveled, with the aim of maximizing each train's useful life. When organizing maintenance, practitioners must take into account both the inbound train wagons that require shunting and replacement, as well as ensuring traffic safety. Our study focuses on the first decision problem, as the number and type of wagons to be replaced is usually assumed in most literature (Chuijiang (2021)).

2.3 Open Issues and Contributions

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

- State-of-the-art (e.g. Chuijiang (2021)) assumes 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);
- Most of the early studies, aim to minimize the operational time and/or cost resulting from shunting in the short term (daily). The shunting operation was never seen as a single entity itself which is why it is often neglected the long-term impacts. This research performs long-term analysis such as the wagon fleet estimation or the level monitoring of the shunting yard capacity for an n-years scenario;
- As far as we know, no paper takes into account maintenance constraints for shunting operation optimization. Indeed, the maintenance scheduling problem is usually treated as a separate tactical problem. This could lead to sub-optimal solutions that disregard KPIs such as the wagon fleet size.

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

- We introduce a new shunting yard issue by highlighting its practical impact on several KPIs;
- We propose an Event-Based Simulation Framework to process real timetables by simulating terminal operations and the passage of time. The EBSF exploits a MILP model, populated by maintenance rules and demand-matching constraints as well as a multi-component objective function. The model minimizes the number of clusters of shunts activated and the possible departure delay caused by the time to shunt. Moreover, it considers terms concerning the SI policy applied and the shunting convenience cost, in order to avoid an infeasible state of the shunting yard capacity;
- We describe three SI policies implemented as add-on terms of the MILP objective function. Each SI policy exploits a wagon selection criteria and has proven to have a strong goal orientation.

3 Problem Description

The SO problem can be influenced by various constraints such as maintenance rules, operational costs, seasonal wagon demand, etc. The cost of the make-ready stage of the shunting locomotive is the biggest part of the SO operation's cost, so creating clusters of shunts by triggering optional shunts that aren't caused by maintenance or demand constraints can bring cost benefits. However, the time to shunt a cluster is equal to the number of wagons in the cluster multiplied by the unit time cost of 15 minutes. Overly large clusters created by too many optional shunts can lead to delays or cancellations. A mathematical model can help practitioners determine when to cluster. The outbound train's composition must be fulfilled and wagons must not be moved for maintenance unless their mileage is within the lease range, taking into account the resource level of the shunting yard and workshop to prevent a shortage of wagons. The SI problem is complementary to the SO problem, aiming to minimize time and economic costs by replacing each shunted-out wagon with a suitable wagon from the shunting yard. A suitable wagon must have enough mileage for the next trip and be of the correct type. The basic problem is replacing a shunted wagon with one that takes less time to shunt in the shunting yard. However, considering just one parameter is short-sighted. For example, the shunted-in wagons might already be close to the mileage limit, which will lead to another SO operation when the train returns. A multi-component objective function that focuses on economic costs can avoid this additional cost. The SISO problem's event flowchart is described in Figure 2. The inbound train is moved to the Arrival/Departure yard queue, where inspections take about 35 minutes. If shunting operations are needed, it is moved to the shunting queue and each SO operation takes 15 minutes. Each shunt costs € 350 and can only be performed if the shunting yard is not busy. The maintenance for one wagon costs € 10500 and the range is from 150,000 km to 172,500 km. Wagons shunted for demand matching are sent to the shunting yard, while those for maintenance go to the workshop for 3 days. Shunted-out wagons are then replaced through SI operations. After all SISO operations, the train is moved to the TLA queue for loading and then to the ADY for departure (unless cancelled).

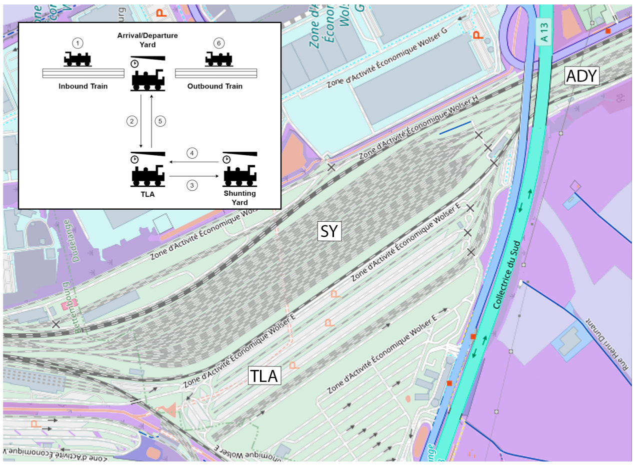


Figure 2: The flowchart outlining the event sequence for the SISO Problem, with TLA denoting the *Train Unloading/Loading Area* and SY referring to the *Shunting Yard*.

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, good type restrictions and so on;
- The operational time to shunt out a wagon is on average 15 minutes, 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 exceed the maximum mileage defined by the corresponding leasing contract. The virtual mileage is equal to the kilometres 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 in the objective function of the MILP;
- 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 relative gaussian distribution has been developed through data history provided by CFL practitioners;
- 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 concerns the methodology and is structured as follows: Subsection 4.1 lists the nomenclature used to model the SISO problem; Subsection 4.2 explains in-depth a basic version of the MILP model, with its constraints and objective function, where it is applied a first example of SI policy, named MIN; Subsection 4.3 presents several SI policies translatable as different versions of the MILP objective function.

4.1 Nomenclature

Sets	
Name	**Description**
$\mathcal{T}$	set of inbound train's wagons
S	set of shunting yard's wagons
K	set of wagon types
Parameters	
Name	**Description**
a_T	integer value expressing the inbound train's arrival time

Parameters	
Name	**Description**
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 kilometres 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 $\mathcal{T}$
ms_j	integer value expressing the current mileage of the wagon j-th inside the shunting yard $\mathcal{S}$
m_{max_i}	integer value expressing the maximum mileage before the maintenance of the wagon i-th on the inbound train $\mathcal{T}$ based on the leasing contract
ms_{max_j}	integer value expressing the maximum mileage before the maintenance of the wagon j-th inside the shunting yard $\mathcal{S}$ based on the leasing contract
m_{min_i}	integer value expressing the minimum mileage to shunt the wagon i-th on the inbound train $\mathcal{T}$
$type_{in_i}$	integer value equal to 1 or 2 expressing the type of the wagon i-th on the inbound train $\mathcal{T}$
$type_{S_j}$	integer value equal to 1 or 2 expressing the type of the wagon j-th in the shunting yard $\mathcal{S}$
$code_{in_i}$	integer value expressing the unique code associated with the wagon i-th on the inbound train $\mathcal{T}$
$code_{S_j}$	integer value expressing the unique code associated with the wagon j-th inside the shunting yard $\mathcal{S}$
$type_r$	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 $\mathcal{T}$
$rise$	integer value expressing the surplus of wagons of the type $type_r$ in the outbound train new composition, compared to the inbound train $\mathcal{T}$
n_{ms_j}	float value expressing the $virtual\ rate$ of the wagon j-th inside the shunting yard $\mathcal{S}$, namely, the ratio between the kilometres covered once the outbound train's next trip has been performed (virtual mileage) and the maximum 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 $\mathcal{S}$
$c_{s_{i,j}}$	float value expressing the temporal cost to replace the wagon i-th on the inbound train $\mathcal{T}$ with the wagon j-th inside the shunting yard $\mathcal{S}$, normalized through the Min-Max normalization
M	a Big-M coefficient
Decision Variables	
Name	**Description**
ad_T	integer value expressing the actual departure time of the outbound train once all the shunting operations are performed
$code_{out_i}$	integer value expressing the unique code associated with the wagon i-th on the outbound train
β	float value between 0 and 1 expressing the percentage of operational time left before the outbound train's deadline once all the SO operations are performed
α	float value equal to 1 - β
y_i	binary variable equals to 1 if on the wagon i-th on the inbound train $\mathcal{T}$ a maintenance or optional shunt is performed
$x_{i,k}$	binary variable equals to 1 if on the wagon i-th on the inbound train $\mathcal{T}$ a demand shunt is performed and it is replaced by a shunting yard's wagon of type k

Decision Variables	
Name	**Description**
$z_{i,j}$	binary variable equal to 1 if the wagon i-th on the inbound train $\mathcal{T}$ is replaced by the wagon j-th inside the shunting yard $\mathcal{S}$
γ_i	binary variable equals to 1 if the wagon i-th on the inbound train $\mathcal{T}$ is shunted out, regardless of the shunt type
σ_1	binary variable equals to 1 if $dd_T \geq ad_T > d_T$, and to 0 if $d_T \geq ad_T$
σ_2	binary variable equals to 1 if $ad_T > dd_T$, and to 0 if $ad_T \leq dd_T$.
σ_3	real variable equals to $\frac{ad_T - d_T}{dd_T - d_T}$ if $\sigma_1 = 1$, and to 0 if $\sigma_1 = 0$
$adj_{i,i+1}$	binary variable equals to 1 if both the wagon i-th and its adjacent wagon $i+1$-th on the inbound train $\mathcal{T}$ are shunted out

4.2 Mathematical Model

Objective Function

The MILP model for the SISO problem aims to minimize an objective function made up of three main terms. These three terms orient the model to: minimize the number of clusters created and, consequently, the emissions and operational costs; look at the delay produced by the SISO operations; consider both the SI policy applied, and its decision criteria, and the shunting yard state in terms of available wagons. The last two terms are associated with different weights based on the practice priority.

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

The first term (1) represents the actual number of shunts performed considering the assumption on clustering (Subsection 3.1). It is the overall number of wagons shunted out, regardless if due to maintenance, optional or demand shunt, minus the adjacencies activated. That way, we count the number of clusters of shunts activated rather than single wagons shunted out, taking into account the assumption that states that two or more adjacent wagons shunted out are considered as a single economic cost. SO operations are not multiplied by a weight, since we have set the other objective function terms so that they are all comparable in number of wagons.

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

On the other hand, the second term (2) expresses the penalty related to the lowering of service level due to the possible departure delay $\mu(ad_T)$ of the outbound train. Due to its non-linear behaviour, $\mu(ad_T)$ has been handled by introducing three different binary variables σ used for the constraints (4)-(13). If ad_T is smaller than d_T, σ_2 and σ_3 will be both equal to zero, and $\mu(ad_T)$ will be equal to zero as well; if ad_T ranges between the outbound train's planned departure time d_T and deadline dd_T, σ_2 will be equal to zero, while σ_3 will be equal to $\frac{ad_T - d_T}{dd_T - d_T}$, as well as $\mu(ad_T)$. $\mu(ad_T)$ is then multiplied by a weight proportional to the number of wagons on the inbound train. The weight can not be proportional to the number of wagons shunted out, because it would both produce a non-linear term and enter into contradiction with the clustering assumption. Indeed, that way the solver would aim to always cluster, because the clustering advantage and the time to shunt would go hand in hand. With the penalty, the solver is forced to choose to activate the optional shunts only if the corresponding cluster does not produce an excessive departure delay.

$$\frac{|\mathcal{T}|}{4}(\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{S}}(c_{s_{i,j}}\alpha + n_{ms_j}\beta)z_{i,j} \;+\sum_{i\in\mathcal{T}} y_i c_{u_i}) \tag{3}$$

According to practice priority, the last term (3) is weighted with a halved value compared to the second term. This term considers two aspects: the SI policy applied, in this case, the MIN policy, and the shunting convenience costs c_{u_i}. The MIN policy aims to both minimize the temporal costs to shunt in wagons stored inside the shunting yard and replace shunted out wagons with the ones with the minimum *virtual rate*. The temporal costs and the minimum virtual rate are then multiplied, respectively, by the percentage of time to shunt out required and of time to shunt in left before the outbound train's deadline. That way, if the time left after the SO operations is too short (α higher than β), the model will choose wagons inside the shunting yard with lower temporal costs rather than the ones with lower virtual rates. Otherwise, β will be higher than α, thus the model will shunt in wagons with lower virtual rates. The second aspect is basically a preemptive tool to avoid too many optional shunts and, therefore, an unfeasible state of the shunting yard capacity. The value c_{u_i} forces the model to weigh carefully optional shunts to avoid the possibility that the shunting yard will quickly run out of wagons. Therefore, if the current mileage of the wagon i-th in the outbound train is low and the number of wagons in the shunting yard is below a certain threshold, the cost c_{u_i} is computed so that will be high, and the model will opt not to shunt out the wagon.

Time Constraints

These constraints represent the ones related to the delay $\mu(ad_T)$ function. Based on ad_T, they state if the outbound train is on time, late or if it will be cancelled due to the exceeding of the planned deadline. This is done by using three temporal variables σ, as described in the nomenclature, in order to add a delay penalty in the objective function.

$$ad_T \;\leq\; d_T \;+\sigma_1 M + \sigma_2 M \tag{4}$$

$$dd_T \;+\; (1 \;-\; \sigma_1)M \;\geq\; ad_T \tag{5}$$

$$ad_T \;>\; d_T \;-\; (1-\sigma_1)M \tag{6}$$

$$ad_T \;\leq\; dd_T \;+\; \sigma_2 M \tag{7}$$

$$ad_T \;>\; dd_T \;-\; (1-\sigma_2)M \tag{8}$$

$$\sigma_1 \;+\; \sigma_2 \;\leq\; 1 \tag{9}$$

$$\sigma_3 \;\leq\; \sigma_1 M \tag{10}$$

$$\sigma_3 \;\geq\; \frac{a_T - d_T}{dd_T - d_T}\sigma_1 \tag{11}$$

$$\sigma_3 \;\leq\; \frac{ad_T - d_T}{dd_T - d_T} \;+\; (1-\sigma_1)M \tag{12}$$

$$\sigma_3 \;\geq\; \frac{ad_T - d_T}{dd_T - d_T} \;-\; (1-\sigma_1)M \tag{13}$$

Constraints (4)-(6) assure that the conditions of σ_1 are satisfied, namely: if both σ_1 and σ_2 are equal to 0 then the outbound train must be on time; if σ_1 is equal to 1 then ad_T must range between d_T (not included) and dd_T. Instead, constraints (7) and (8) express σ_2's conditions, saying that if σ_2 is equal to 0 then ad_T has not already reached the outbound train's deadline, otherwise, the outbound train will be cancelled. While constraint (9) links σ_1 to σ_2 by forcing them not to be simultaneously active, constraints (10)-(13) link σ_1 and σ_3 and represent σ_3's conditions. The latter states that if σ_1 is equal to 0 then σ_3 will be equal to 0 as well, while if σ_1 is equal to 1 then ad_T ranges between d_T and dd_T, and σ_3 will be equal to $\frac{ad_T - d_T}{dd_T - d_T}$.

$$a_T \; + \; \sum_{i \in \mathcal{T}} \gamma_i ts \; + \; \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{S}} c_{s_{i,j}} z_{i,j} \; = \; ad_T \tag{14}$$

Constraint (14) sets the ad_T equal to the inbound train arrival time plus the time required to perform both the SO and SI operations.

$$\frac{(a_T \; + \; \sum_{i \in \mathcal{T}} \gamma_i ts)}{dd_T} \; = \; \alpha \tag{15}$$

$$1 - \alpha = \; \beta \tag{16}$$

Constraints (15) and (16) set α and β as described in the nomenclature. The value of α ranges between 0 and 1, and is set as the arrival time plus the time to shunt out, divided by the outbound train's deadline.

SO Constraints

The following constraints assure that Shunt-out operations are performed properly, namely, based on the assumptions.

$$y_i \geq \frac{m_i + r_T}{m_{max_i}} - 1 - (\sum_{k \in \mathcal{K}} x_{i,k}) M \qquad \forall i \in \mathcal{T} \tag{17}$$

$$y_i \leq (1 - \sum_{k \in \mathcal{K}} x_{i,k}) \frac{m_i + r_T}{m_{min_i}} \qquad \forall i \in \mathcal{T} \tag{18}$$

The above constraints (17) and (18) represent three conditions: (i) if m_{max_i} is exceeded by performing the next trip, then the wagon i-th must be shunted out, thus y_i is equal to 1; (ii) the wagon i-th must be shunted out only if exceeds m_{max_i} required, otherwise, y_i is equal to 0; (iii) if a demand shunt is already activated for the wagon i-th, it is not possible to perform a maintenance or an optional shunt.

$$y_i \leq \frac{m_i + r_T}{m_{max_i}} \qquad \forall i \in \mathcal{T} \tag{19}$$

Instead, constraint (19) should be considered if we want to force the model to activate only mandatory shunts once m_{max_i} is exceeded (CFL approach).

$$\sum_{i \in \mathcal{T} : type_{in_i} \neq type_r} x_{i,type_r} = rise \tag{20}$$

$$\sum_{k \in \mathcal{K}} x_{i,k} = 0 \qquad \forall i \in \mathcal{T} : type_{in_i} = type_r \tag{21}$$

$$\sum_{i \in \mathcal{T}} x_{i,k} = 0 \qquad \forall k \in \mathcal{K} : k \neq type_r \tag{22}$$

Constraints (20)-(22) assure that the new composition on the outbound train will be satisfied. Therefore, the sum of $x_{i,k}$ with a k different from the $type_r$ (the wagon type that must increase on the outbound train) must be equal to $rise$, the additional wagons of $type_r$ required by the new composition. Moreover, the sum of $x_{i,k}$ of $type_r$ type must be equal to 0, because in this way the other part of the composition will not be altered.

$$|type_{out_i} - type_{in_i}| = p_i \qquad \forall i \in \mathcal{T} \quad (23)$$

In certain circumstances, the practice might require an orderly composition of the departure train, namely, for each wagon position a specific type is mandatory, as indicated by the timetable. To express this necessity, constraints (20)-(22) must be replaced by constraint (23), where $type_{out_i}$ represents the type of wagon i required on the outbound train and the binary variable p_i is equal to 1 when the inbound train wagon i is shunted out due to demand matching. However, constraint (23) is only valid when there are exactly two wagon types.

SI Constraints
Constraints (24)-(27) allow the model to perform the Shunt-in operations.

$$\sum_{j \in \mathcal{S}: type_{S_j} = type_{in_i}} z_{i,j} = y_i \qquad \forall i \in \mathcal{T} \quad (24)$$

$$\sum_{j \in \mathcal{S}: type_{S_j} \neq type_{in_i}} z_{i,j} = x_{i,type_r} \qquad \forall i \in \mathcal{T} \quad (25)$$

$$z_{i,j} \leq 2 - \frac{ms_j + r_T}{ms_{max_j}} \qquad \forall i \in \mathcal{T}, \forall j \in \mathcal{S} \quad (26)$$

$$\sum_{i \in \mathcal{T}} z_{i,j} \leq 1 \qquad \forall j \in \mathcal{S} \quad (27)$$

Constraints (24) and (25) force the model to activate z_{ij} with the proper type j, such that while the wagons shunted out with maintenance or optional shunt will be replaced by shunting yard's wagons of the same type, the ones shunted-out with demand shunt will be replaced by wagons of the opposite type. Instead, constraints (26) and (27) assure both that no wagon with insufficient residual mileage to perform the next trip r_T will replace the inbound train wagons, and that the same shunting yard wagon will not replace more than one inbound train wagon. These constraints define the feasible region where the SI policy applied will select shunting yard wagons.

Adjacency Constraints
The following constraints allow the model to comply with the clustering assumption.

$$\sum_{k \in \mathcal{K}} x_{i,k} + y_i = \gamma_i \qquad \forall i \in \mathcal{T} \quad (28)$$

$$2adj_{i,i+1} \leq \gamma_i + \gamma_{i+1} \qquad \forall i = 1, ..., |\mathcal{T}| - 1 \quad (29)$$

Constraints (28) and (29) assure that if two or more wagons on the inbound train will be shunted-out, then the respective adjacency variables will be activated and counted in the objective function. This is done by summarizing in a single variable γ_i both the demand, the optional and the maintenance shunts performed, and by forcing the activation of $adj_{i,i+1}$ only if both γ_i and γ_{i+1} are equal to 1.

$$\sum_{j \in \mathcal{S}} z_{i,j} code_{S_j} + (1 - \gamma_i) code_{in_i} = code_{out_i} \qquad \forall i \in \mathcal{T} \quad (30)$$

To conclude, constraint (30) is an optional one, not strictly necessary to optimize the problem, but quite useful to keep track of wagons' codes that will be on the outbound train once all the SO and SI operations are performed. If the wagon i-th has been replaced by activating z_{ij}, this constraint associates the wagon j-th's code to the position i-th, alternatively, the wagon i-th's code remains unchanged.

4.3 Shunt-In Policies

In this section, we propose different SI policies, namely, add-ons to the multi-component objective function (1)-(3).

Policy Modelling

The basic version of the SI model takes into account only the time cost required to move a wagon i-th into the position j-th. This could be a limiting approach in a strategic vision, as it just looks at the short-term decisions, while avoiding any predictive approach to future shunting. SI policies allow the model to exploit different features impacting the number of future shunting operations. These features are directly linked to assumptions (Subsection 3.1), such as those on clustering, and have proved to be impactful also on the wagon fleet size, departure delays, and average mileage performed by each wagon. The general structure of an SI policy is the following (31). Based on the policy criteria, weights W_1 and W_2 can assume different meanings, in order to head the objective function towards specific tactical and strategic goals.

$$\min \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{R}} (\alpha W_1 + \beta W_2) z_{i,j} \quad (31)$$

For our case study, we will refer to W_1 as the cost $c_{S_{i,j}}$, while W_2 will change depending on the policy applied. α and β are complementary parameters describing the temporal state of the system. They are given in real-time by MILP constraints. Specifically, β represents the remaining percentage of the operational time once all the SO operations have been performed (the latter is expressed by α). Depending on the value assumed by α and, consequently, by β, the solver will decide whether to weigh more the cost $c_{S_{i,j}}$ or the policy criteria W_2. Given that we are talking about two different measurement units, it is necessary to normalize both W_1 and W_2 (we used a $Min\text{-}Max\ Normalization$, Patro and Sahu (2015)).

MIN Policy

This policy aims to shunt in the shunting yard's wagon with the minimum virtual rate v_{S_i}. Therefore, the objective function is the following:

$$\min \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{R}} (\alpha c_{S_{i,j}} + \beta v_{S_i}) z_{i,j} \quad (32)$$

Based on (32), if the percentage of operational time required by the SO operations is predominant, the solver will opt to shunt in wagons with a lower cost $c_{S_{i,j}}$. Otherwise, the solver will be directed toward the policy criteria, by choosing wagons with the lowest v_{S_i}.

AVG L-S Policy

This policy aims to make the most of the wagon's mileage capacity, based on its services assignment record. When a wagon i has been assigned with many long-trip services (n_{long_i}), it's reasonable to assign it with short-trip ones (n_{short_i}), to fully exploit its maximum mileage ms_{max_i}. Its assignment record is expressed by the *degree of unbalance*, as described by (33). We want to choose the suitable i-th wagon $\in \mathcal{S}$ with the maximum degree of unbalance, defined as:

$$\theta_i = n_{long_i} - n_{short_i} \tag{33}$$

The objective function will therefore become:

$$\min \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{R}} (\alpha c_{S_{i,j}} - \beta((-1)^{long} \theta_i)) z_{i,j} \tag{34}$$

Where, by assuming $r_{threshold}$ as the 25^{th} percentile of the distribution of the trips considered, the binary data $long$ is equal to 1 if r_T for the outbound train is greater than $r_{threshold}$, 0 otherwise. If the next service will perform a long trip, the solver will opt for wagons that have performed more short trips than long ones, otherwise, wagons with more long trips than short ones will be chosen.

NCLD Policy

The $NCLD$ policy helps the solver to create clusters as wide as possible. This approach is strictly related to the clustering assumption and aims to cluster the entire train by activating a single shunting variable γ_i. We define the virtual rate, both for the inbound train and shunting yard wagons, as follows:

$$v_i = \frac{m_i + r_T}{m_{max_i}} \tag{35}$$

$$v_{S_j} = \frac{ms_j + r_T}{ms_{max_j}} \tag{36}$$

If wagons with homogeneous virtual rates are shunted in, then it is likely that, during the next trips, the inbound train wagons will be shunted out altogether. The add-on in the multi-component objective function will be:

$$\min \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{R}} (\alpha c_{S_{i,j}} + \beta |A_{SO} - A_{SI}|) z_{i,j} \tag{37}$$

With A_{SO} and A_{SI} equal, respectively, to the average virtual rate on the inbound train once all the SO operations have been performed and to the average virtual rate of the wagons shunted in, as expressed by (38) and (39).

$$A_{SO} = \frac{\sum_{i \in \mathcal{A}} (1 - \gamma_i) v_i}{\sum_{i \in \mathcal{A}} (1 - \gamma_i)} \tag{38}$$

$$A_{SI} = \frac{\sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{R}} z_{i,j} v_{S_j}}{\sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{R}} z_{i,j}} \tag{39}$$

That way, if β is higher than α, (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.

5 Computational Results

The simulations refer to two different 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 MILP model performance with real data on emissions, delays and train cancellations provided by practitioners. The long-term analysis's benchmark on the wagon fleet size and emissions produced consists of running the simulator with the activation of 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, based on the tactical and strategic objectives affected by SISO operations:

- The $wagon\ fleet\ size$, impacting overhead costs, given that the rolling stock cost is one of the most competitive factors for a railway company (Giacco et al. (2014));
- $Delays$ and $train\ cancellations$, in order to guarantee an acceptable service level;
- $Emissions$, which minimization is directly related to shunting operations performed and operational costs.

The integration of the MILP model into the Event-Based Simulation Framework was achieved using the CPLEX Python API in *PyCharm* 11.0.15, along with the callable library *cplex* release 22.1.0. The computer used for the simulations was equipped with an Intel(R) Core(TM) i9-10885H, CPU 2.40GHz, and 32 GB of RAM.

5.1 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 both to replace the wagons shunted out due to maintenance and to fulfil the outbound train planned composition. In the CFL case, the request of wagons by type is usually unbalanced (18% of supplies for the Simple type and 82% for the Double one), thus, effective wagons management is required to avoid infeasible solutions. For the simulations, we have considered a pool of 1100 wagons. If a wagon is used, during the run, for at least one service, it will be counted as part of the solution. In Table 1 data relative to the fleet used resulting from the 2022-2050 simulation are gathered. For a most comprehensive analysis, we have tracked as well the average current mileage of wagons available in the shunting yard (Figures 3a-3b).

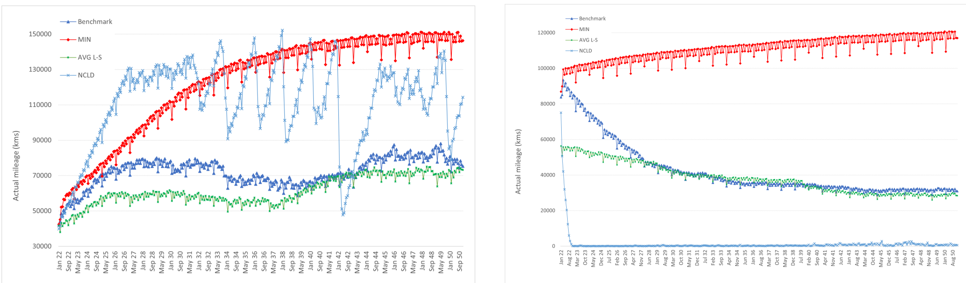
This is a good index of fleet reliability. Indeed, although MIN performs better in the short term, its short-sighted approach pays in the long term. Especially for the Simple

Table 1: Distribution of mileage 2022-2050

	Wagon Usage Distribution (Km)			**Wagon Fleet Used**	
	STD	Median	Mean	Number of Wagons	Cost (€ Mln)
			Benchmark		
SIMPLE	326368	125530	329237	170	28
DOUBLE	402884	301088	437303	755	297
			AVG L-S		
SIMPLE	218118	152137	267500	209	34
DOUBLE	1094474	1683600	1469344	224	88
			MIN		
SIMPLE	286682	250105	331470	169	28
DOUBLE	854882	662145	939078	351	138
			NCLD		
SIMPLE	388391	265365	388710	122	20
DOUBLE	678124	1388309	1082490	313	123

Table 1 Distribution of mileage on rolling stocks and wagon fleet usage for the 2022-2050 simulation.

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 a positive behaviour by keeping, for the Simple wagons, the average mileage around 70000 kilometres.



(a) (b)

Figure 3: Average mileage for the Simple (a) and Double (b) wagons available in the shunting yard.

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 the supply chain/inventory management, called *Bullwhip Effect*. 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 Figure 3a. 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 higher median, mean and standard deviation. This translates into fewer wagons used to the fullest, and savings of € 203 mln over 30 years.

5.2 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. 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 minutes later than the scheduled departure time. When the train exceeds the deadline of 180 minutes, it is cancelled except in the case that carries high-value goods. In our instances, arrival delays may be due to the combination of trips and shunting operations. To consider trip delays, we have combined the Machine Learning model developed by Pineda-Jaramillo et al. (2021) into the framework, which computes travel times based on several wagon attributes (weight, length, volume and so forth). 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 2 data comparing SISO policies and CFL benchmark delay and train cancellation rates for the 2021 simulation are gathered. Policies provide a significant improvement in the percentage of delays and train cancellations, and the 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.

Table 2: Delay and train cancellation rates for the 2021 simulation

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

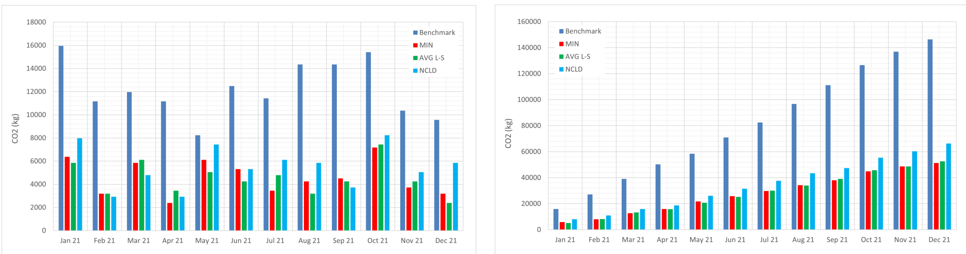
The SISO model here proves its strength by reducing the delays and cancellations rate up to 4% for MIN and AVG L-S compared to the 40% of the benchmark. Results show how NCLD and MIN keep a steady cumulative delay throughout the simulation while AVG L-S, once passed a warm-up phase where the shunting activity is heavier due to the setup of the *degree of unbalance*, reduces the delays rate over time by increasing the clustering rate.

5.3 Emissions

In this subsection, we gather and analyze data concerning volumes of fuel and emissions generated by the shunting activity of the simulations. 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. 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 travelling (Rymaniak et al. (2019)). 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. Therefore, by defining a standard operational time to shunt given by the sum of the shunt out and shunt in average times (55 minutes), the cumulative number of hours of shunting activity per month/year, and the fuel consumed, are computed as follows:

$$TOT_f = (N_s \frac{15}{60}) f_s \tag{40}$$

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 UIC (2010). This study provides an assessment of the diesel locomotive fleet in Europe and average emission factors. By using fuel data as the primary activity indicator, we can extract the annual volumes produced for each GHG. As we can see in Figure 4, there are significant effects in terms of fuel



(a) (b)

Figure 4: Average and cumulative kilograms of CO_2 produced by the benchmark and each policy for the 2021 simulation.

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 2021 benchmark of CFL 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 following growing branch between the benchmark and MIN. For the 2022-2050 simulation, the total reduction of CO_2 produced by applying AVG L-S policy instead of the basic MILP model is around 8% translatable as 1,52 tons per year.

5.4 Computation Time

While our contribution primarily focuses on a long-term analysis, the computation time to process a single train may be useful in addressing and integrating the real-time problem with short-term issues like the Classification or the Train Makeup problems. To address this, we present the running times of all simulations in Table 3 along with the average computation time to process a standard-size inbound train.

	Benchmark	**MIN**	**AVG L-S**	**NCLD**
Simulation	Computation Time			
2021	00:10:21,05	00:11:10,32	00:10:48,11	00:11:59,55
2022-2050	05:00:23,12	05:24:03,23	05:13:07,18	05:47:22,11
Single Train	00:00:01,02	00:00:01,30	00:00:01,20	00:00:01,50

Table 3: Computation time for simulations and average processing time for a standard-size inbound train.

6 Conclusions

We present a Decision Support System (DSS) designed to solve the Shunt-In Shunt-Out (SISO) problem in shunting yards. Currently, this problem is only tackled based on the experience of CFL practitioners. Our DSS serves as an auxiliary tool for the industrial software *Shunty* which is part of the ANTOINE national project. The DSS optimizes multiple strategic and tactical KPIs, such as fleet size, shunting operations, delays, and cancellations. It employs an Event-Based Simulator Framework (EBSF) that integrates a MILP model and programming tools, with Python setting up the simulation and the MILP model processing inbound trains requiring SISO. The MILP model considers technical feasibility and demand matching, and proposes various SISO policies, each with its own selection criteria, resulting in unique behavior and a clear goal orientation. These policies can potentially be combined for further optimization and improved outcomes. The results from the 2021 simulation show a 65% reduction in fuel consumption and emissions, a €3.8 mln reduction in costs, and a 36% reduction in delayed trains. Future work includes optimization for multi-train SISO, a *Column Generation* algorithm for a SI heuristic, and the integration of all ANTOINE sub-problems into Shunty. Shunty provides practitioners with a user-friendly interface for uploading timetables, applying SISO policies, and analyzing KPIs. Our goal is to offer practitioners both real-time and long-term decision support.

7 Acknowledgements

This research was supported by 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.

References

Agency, E. E., 2021. *Greenhouse gas emissions from transport in europe*, European Environment Agency, Copenhagen, Denmark.

Birol, F., 2021. *Net zero by 2050 - a roadmap for the global energy sector*, International Energy Agency, Paris, France.

Budai, G., Huisman, D., and Dekker, R., 2006. "Scheduling preventive railway maintenance activities", *Journal of the Operational Research Society*, vol. 57, pp. 1035–1044.

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*, vol. 9, pp. 579–594.

Freling, R., Lentink, R. M., Kroon, L. G., and Huisman, D., 2005. "Shunting of passenger train units in a railway station", *Transportation Science*, vol. 39(2), pp. 261–272.

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

Haahr, J. and Lusby, R. M., 2017. "Integrating rolling stock scheduling with train unit shunting", *European Journal of Operational Research*, vol. 259(2), pp. 452–468.

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

Jaehn, F., Rieder, J., and Wiehl, A., 2015. "Minimizing delays in a shunting yard", *OR Spectrum*, vol. 37, pp. 407–429.

Li, B., Qu, W., Xuan, H., Xu, C., Wang, X., and Geng, Y., 2020. "Shunting Strategy for Placing-In and Taking-Out Wagons on Branch-Shaped Freight Operation Network with Simulation Method", *Journal of Advanced Transportation*, vol. 2020, pp. 1–25.

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

Pagand, I., Carr, C., and Doppelbauer, J., 2020. *Fostering the railway sector through the European Green Deal*, European Union Agency For Railways, Valenciennes, France.

Patro, S. G. K. and Sahu, K. K., 2015. "Normalization: A preprocessing stage", *arXiv*.

Pineda-Jaramillo, J., McDonald, W., Zheng, W., and Viti, F., 2021. "Identifying the major causes associated to rail intermodal operation disruptions using causal machine learning.".

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

UIC, 2010. *Ecological Transport Information Tool for Worldwide Transports*, International Union of Railways, Berlin, Germany.