Towards Safer Freight Rail Shunting: Integrating MILP and ML Classification Models in a Risk Management Framework.

Bigi Federico[1], Bosi Tommaso[2], D'Ariano Andrea[2], and Viti Francesco[1]

[1] University of Luxembourg, Esch-Sur-Alzette, Luxembourg,
federico.bigi@uni.lu,
[2] Department of Civil, Computer Science and Aeronautical Technologies Engineering, Roma Tre University, Rome, Italy

Abstract. This paper proposes a novel risk analysis framework for rail freight shunting operations, challenging the direct application of Machine Learning (ML) as input for operational decision-making. Our approach integrates ML model's performance metrics into a Mixed-Integer Linear Programming (MILP) model for shunting operation, improving the reliability and adaptability of rail yard management. A comparative analysis based on real data from the Luxembourgish rail freight company CFL Multimodal across various destinations reveals that a full-risk function approach significantly outperforms the ML input model, reducing the risk term and cancellation rates. This study demonstrates that utilizing a risk assessment framework mitigates the potential for operational inefficiencies and unfeasibility inherent in ML-dependent models, ensuring continuous data accuracy and operational robustness.

Keywords: Rail Freight Operations, Shunting Operations, Risk Management, Machine Learning, Mixed-Integer Linear Programming, Data-driven modeling

1 Introduction

The transportation sector, particularly road transport, is a significant contributor to greenhouse gas emissions and air pollution, highlighting the need for sustainable alternatives. Rail freight transport, accounting for only 5.4% of total transport as of 2021, is recognized by the European Union as a key component in reducing its environmental impact. The EU aims to double rail freight's contribution within the next 30 years, emphasizing its role in achieving carbon neutrality Agency (2021). However, the shift to rail freight brings challenges, particularly in the logistics and operational costs associated with operations in rail yards. Shunting, the process of rearranging wagons for train composition and maintenance, is crucial for efficient service delivery and contributes to operational delays, and cancellations, which can spread also in the network Spanninger et al. (2022), causing a chain effect throughout it. As rail freight aims to expand its role in sustainable transportation, optimizing these shunting operations becomes critical to enhancing service reliability and minimizing environmental and economic impacts (Cadarso and Marín, 2011; Jaehn et al., 2015; Giacco et al., 2014). To achieve this, new data-driven models have been developed to predict delays in the freight train sector, as well as a model to optimize shunting operations Pineda-Jaramillo and Viti (2023). However, there is a notable lack of research on how unexpected rolling stock disruptions affect shunting operations. Addressing this gap is the goal of this research, as such unplanned events can cause further delays, cancellations, and additional maintenance. This study introduces a novel risk analysis framework to improve the decision-making process in shunting yard operations. We focus on the integration of Machine Learning (ML) performance and its input into a Mixed-Integer Linear Programming (MILP) model. We propose an extension of the shunt-out/shunt-in model proposed in Bosi et al. (2023) and Bigi et al. (2024), by incorporating a more sophisticated representation of the dynamic operational environment of shunting yards. Our methodology aims to synergize ML analytics and MILP optimization, opening a new dimension of analysis that focuses on enhancing the MILP decision through the assessment of the ML model's trustworthiness within shunting yard operations. By incorporating metrics like True Positive Rate (TPR) and True Negative Rate (TNR), we can quantify the model's performance in terms of system vulnerability and improve the reliability of its predictions. This incorporation is particularly beneficial in guiding maintenance decision-making, and optimizing shunting operations while mitigating the risk of disruptions.

2 Literature Review

While many definitions of shunting operations exist, we will adhere to the one proposed by Cadarso and Marín (2011), which defines the shunting operation as the movement of one or more rolling stocks within a shunting yard. These operations, while being expensive and time-consuming, are usually performed for a specific reason, which is for example fulfilling a service for an outbound train that has to be created in the departure track (demand) from the available wagons parked in the classification tracks (supply). Once this matching has been performed, then for each wagon that needs to be shunted a specific routing has to be created within the tracks of the shunting yard to connect these rolling stocks to the actual outbound train. Thus, the shunting optimization problem, as for Haahr and Lusby (2017), is usually modeled through two interrelated and sequential sub-problems: the *Rolling Stock Problem* (RSP) and the *Train Unit Shunting Problem* (TUSP). The RSP addresses the planning of each wagon's service time and aims to optimize the management of rolling stocks and reduce costs, while the TUSP is concerned with the routing of different rolling stocks throughout the shunting yard, parking of these in the shunting yard Richard Freling and ERIM (2002); Kroon et al. (2008) and all the operations that have to be performed on the rolling stock to get it ready for service Lentink (2006). Generally, the RSP is solved before tackling the TUSP in both passenger and freight train operations. The sequencing is logical, as the TUSP depends on the outcomes of the RSP for essential inputs, particularly the alignment of rolling stock for inbound and outbound trains. Recognizing this distinction, various studies have proposed methodologies to integrate these two problems together. Kamenga et al. (2019, 2021) proposes the Generalized Train Unit Shunting problem (G-TUSP), which is composed of 4 subproblems: the Train Matching Problem (TMP), the problem of matching arriving and departing train units; the Track Allocation Problem (TAP), the problem of choosing train units location; the Shunting Routing Problem (SRP), the problem of determining train units routing during shunting movement; the Shunting Maintenance Problem (SMP), the problem of defining train units maintenance scheduling. The Shunt-in Shunt-out (SISO) problem, as presented in Bosi et al. (2023); Bigi et al. (2024), introduces a MILP model for optimizing rail freight operations and integrating maintenance operations. This model addresses two key components: Shunt-out (SO), which involves removing wagons from an inbound train based on factors like maintenance needs, operational costs, and seasonal demand, and Shunt-in (SI), which focuses on efficiently selecting suitable wagons from the shunting yard for the outbound train. Suitability is determined by the wagon's remaining mileage and type compatibility. This model's primary goal is to minimize the number of shunting operations, ensuring cost-effectiveness while maintaining train composition integrity. While existing literature, such as Pineda-Jaramillo and Viti (2023) and Spanninger et al. (2022), extensively explores disruption prediction using data-driven modeling, there remains a distinct gap. Specifically, there is a lack of research directly linking ML models to the disruptions of rolling stocks, and none that connects these predictions to shunting operations in freight rail transport. Our work aims to bridge this gap by integrating ML models for rolling stock disruption prediction within MILP models for rolling stock management and shunting operations, offering a new perspective in this field.

3 Methodology

In this section, we present an improved stochastic optimization model that addresses the limitations identified in earlier MILP models as presented in Bosi et al. (2023) and Bigi et al. (2024). This study extends the previously presented models and implements a risk assessment framework to evaluate the ML model's performance to represent the trustworthiness of the model. For this study, we use a Binary Classification ML model to predict the disruption of rolling stock, whose predictions are then assessed through a risk assessment approach. All the definitions of parameters and variables can be found in Section A.

3.1 Risk Assessment

Risk assessment identifies hazards and evaluates risks in a system, leading to control measures to mitigate them. Risk is classically defined as:

$$Risk = probability\, of\, occurrence * impact\, of\, event \tag{1}$$

In the FAIR Risk Taxonomy (FAIR Institute (2023)), the risk analysis includes the identification of the assets at risk for the targeted event and the analysis of the likelihood of the event, to assess the frequency of potential threats. The impact of the event, also defined as Loss Magnitude (LM), proposes a set of losses that one might consider in their risk assessment.
As the target of our analysis is the rolling stock, the *event* that we're trying to prevent is an unexpected disruption. To predict a potential disruption, we use a Binary Classification ML model. Given that this model is the source of our information on the disruption, this is the *asset* that we are interested in assessing against its failure, as this could lead to a rolling stock disruption, setting off a cascade effect of other disruptive events.
Therefore, the likelihood of our event is defined as the probability that the ML model could provide a wrong prediction, leading to an event that affects both rolling stocks and shunting operations. To each prediction, we can always associate its relative performance of the classification model, False Positive Ratio, True Positive Ratio, True Negative Ratio, and False Negative Ratio. These rates define how trustworthy the model is on its prediction versus the actual event happening and can be computed based on the validation data.
The losses analyzed in this study are relative to the potential events that are triggered by the failure of the ML model: the **Productivity loss** (PL) is defined as the loss of availability due to disruptions occurring; the **Replacement loss** (RL) is defined as the cost for replacing the disrupted rolling stock; **Monetary loss** (MnL) is defined as fees associated with the disruptive event. With this, we can define two types of risks, the **Risk of Pointless Disruption** and the **Risk of Disruption**.
The Risk of Pointless Disruption, R_{point}, represents the risk of performing unnecessary maintenance, which we formalize as:

$$R^{point} = \overbrace{\bar{t}_{out}^{prev} * f^{dest,yr} * R^{dest} * rev^{dest}}^{\text{PL}} + \overbrace{C_{single}^{shunt}}^{\text{RL}} \tag{2}$$

The PL in this case is associated with the event of sending a rolling stock out for preemptive maintenance, defined as the economic impact of having the rolling stock out for maintenance.

The Risk of Disruption, R_{grav}^{disr}, defines instead the risk associated with encountering a disruptive event, which we formalize as:

$$R_{grav}^{disr} = \overbrace{(D_i - a_i) * rev^{dest}}^{\text{PL}} + \overbrace{C_{maint}^{corr,grav} + C_{single}^{shunt}}^{\text{RL}} + \overbrace{P(grav == 4) * C^{canc}}^{\text{MnL}} \tag{3}$$

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

A conceptual framework is also presented for clarity in Figure 1.

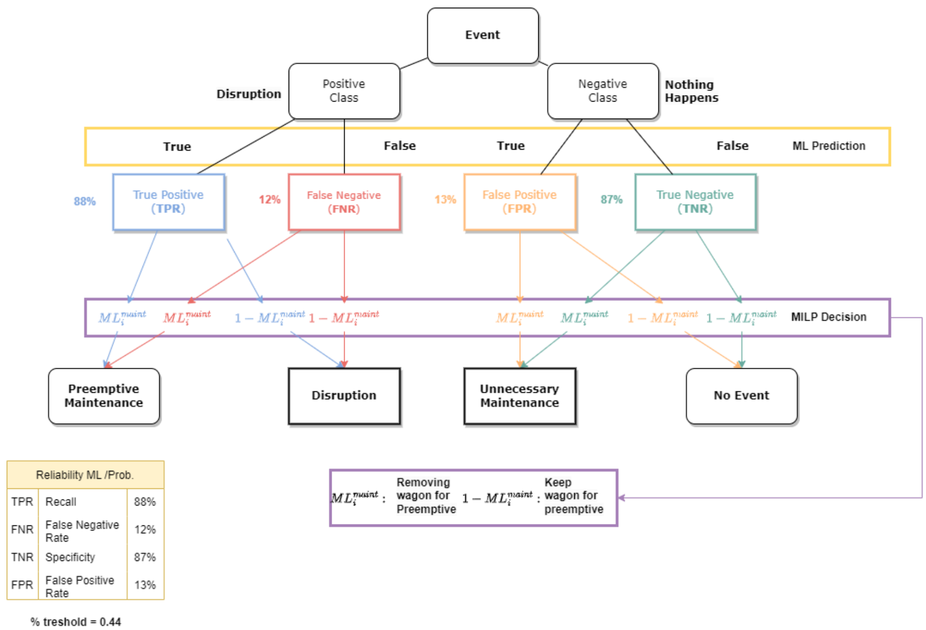

Fig. 1: Conceptual framework of the risk assessment.

We can see the impact of the assessment on the decision of the shunting model, as well as the proposed mild integration of the ML inside the MILP. The rolling stock arrives with an unknown status, defined in our case as the *event*. The ML model is then run, providing its prediction regarding the status of the wagon. This prediction is then not executed right away, but instead is placed inside the MILP, together with the performance metric of the ML model, to make the former decide whether to agree with the ML model based on the associated risks, operations to perform, delays, and cancellations. The strength of this framework is that the MILP model can make much more informed decisions based on how good is the ML model, as well as being able to tune the ML according to the impact of each risk. In the example provided in Figure 1, we can see that the proposed model presents good metrics of Recall and Specificity. Recall is a metric that measures the proportion of actual positive cases correctly identified as positive by the model, while Specificity is a metric that measures the proportion of actual negative cases correctly identified as negative by the model. Within the framework, as for these metrics, the MILP model will tend to trust more the ML prediction. Let's suppose a rolling stock arrives in the station, with our ML model providing us with a *False*, meaning either that it could be a False Negative or a True Negative. Based on the performance of the model, as well as the operations that have to be conducted in parallel to this one, the MILP will decide whether to shunt it out or not for preemptive maintenance weighting in this case the Risk of Disruption only by 12%, rather than the expected probability of disruption of the wagon.

3.2 Operational Research Model

The objective function 4 formulates a revenue maximization problem, defined as the potential revenue of the train minus the costs and risk associated with the MILP decision of shunting, maintenance, and potential cancellation.

$$\max \quad Rev^{train} - (C_{\text{maint}} + shunts \cdot C_{\text{shunt}} + Risk + Canc^{trig} \cdot C_{canc}) \tag{4}$$
$$\text{s.t.}$$

The set of constraints 5 models the behavior of the model in terms of risk and revenue.

$$\sum_{i \in \text{rail}} (1 - w_i^{SO}) \cdot rev_i + \sum_{j \in \text{tr}} \sum_{k \in \text{pos}} w_{i,j,k}^{\text{SI}} * rev_{j,k} = Rev^{train} \tag{5a}$$

$$\sum_{i \in \text{rail}} R_i^{\text{SI}} + R_i^{SO} = Risk \tag{5b}$$

$$\begin{aligned} & pred_i^{ML} \cdot \{TPR \cdot [(\sum_{g \in \text{gr}} P_g \cdot R_{i,g}^{disr}) \cdot \\ & (1 - ML_i^{maint}) + (C_{shunt} + C^{prev,maint}) * ML_i^{maint}] + \\ & FPR * [(C_{shunt} + C^{prev,maint} + R^{point}) * ML_i^{maint}]\} \\ & +(1 - pred_i^{ML}) \cdot \{TNR \cdot (C_{shunt} + R^{point} + C^{prev,maint}) \cdot ML_i^{maint} \\ & +FNR \cdot [(\sum_{g \in \text{gr}} P_g * R_{i,g}^{disr}) \cdot (1 - ML_i^{maint}) \\ & +(C_{shunt} + C^{prev,maint}) \cdot ML_i^{maint}]\} = R_i^{SO} \quad \forall i \in \text{rail} \end{aligned} \tag{5c}$$

$$\begin{aligned} & \sum_{j \in \text{tr}} \sum_{k \in \text{pos}} \{[pred_{j,k}^{ML} \cdot TPR \cdot (\sum_{g \in \text{gr}} P_g \cdot R_{i,j,k}^{disr}) \cdot w_{i,j,k}^{\text{SI}}] + \\ & +(1 - pred_{j,k}^{ML}) \cdot FNR \cdot [(\sum_{g \in \text{gr}} P_g \cdot R_{i,j,k}^{disr}) \cdot w_{i,j,k}^{\text{SI}}]\} = R_i^{\text{SI}} \qquad \forall i \in \text{rail} \end{aligned} \tag{5d}$$

Equation 5a calculates the train's revenue, accounting for the revenue from wagons post-operations and shunting. Equation 5b defines the overall risk with two components for the Shunt-Out, R_i^{SO}, and the Shunt-In, R_i^{SI}, respectively.

- R_i^{SO}, formalized in Equation 5c, defines whether to remove wagons based on ML model prediction. When $pred_i^{ML} = 1$, the equation weighs disruption risk, $R_{i,g}^{disr}$, against shunting and maintenance costs, considering the model's TPR and FPR. For $pred_i^{ML} = 0$, it compares the risk of keeping the wagon against removal costs, using TNR and FNR as weights.

– R_i^{SI}, described in Equation 5d, evaluates the risk associated with wagons from the shunting yard joining the outbound train. The focus here is on selecting wagons for shunting based on the minimum risk for their intended destination.

Equations 5 forms the decision-making framework for train operations, integrating revenue considerations with risks and costs of wagon shunting and maintenance, guided by the insights from an ML model.

The set of constraints 6 defines the time, cost, and number of shunt constraints.

$$\sum_{i\in\mathrm{rail}} \cdot \sum_{g\in\mathrm{gr}} (P_g \cdot C_g^{corr,maint}) \cdot maint_i + C^{prev,maint} \cdot ML_i^{maint} = C_{maint} \tag{6a}$$

$$t_{dep} - t_{op} \leq M \cdot Canc^{trig} \tag{6b}$$

$$shunts \cdot t_{shunt} = t_{dep} \tag{6c}$$

$$shunts^{SI} + shunts^{SO} = shunts \tag{6d}$$

The following set of equations 7 models the shunt-out operations, the decision of which rolling stock to remove from the inbound train.

$$\sum_{i\in\mathrm{rail}} w_i^{SO} - \sum_{i=1}^{rail-1} adj_{i,i+1} = shunts^{SO} \tag{7a}$$

$$2 \cdot adj_{i,i+1}^{SO} \leq w_i^{SO} + w_{i+1}^{SO} \quad \forall i \in [\mathrm{rail}, \mathrm{rail}+1] \tag{7b}$$

$$maint_i + ML_i^{maint} + dem_i \leq 1 \quad \forall i \in \mathrm{rail} \tag{7c}$$

$$dem_i \leq (1.05 - prob_i^{ML}) \quad \forall i \in \mathrm{rail} \tag{7d}$$

$$\sum_{i\in\mathrm{rail}} dem_i \leq \sum_{l\in\mathrm{types}} |diff_l| \tag{7e}$$

$$maint_i + ML_i^{maint} + dem_i = w_i^{SO} \quad \forall i \in \mathrm{rail} \tag{7f}$$

The set of equation 8 connects the shunt-in and shunt-out variables, as well as the type and demand management.

$$\sum_{l\in\mathrm{types}} diff_l - \sum_{i\in\mathrm{rail}} (w_i^{SO} - \sum_{j\in\mathrm{tr}} \sum_{k\in\mathrm{pos}} w_{i,j,k}^{\mathrm{SI}}) = 0 \tag{8a}$$

$$\sum_{j\in\mathrm{tr}} \sum_{k\in\mathrm{pos}} w_i^{SO} \cdot typ_{i,j,k} + (1 - w_i^{SO}) \cdot WI_{i,l} \geq w_{i,l}^{OUT} \quad \forall i \in \mathrm{rail}, \quad \forall l \in \mathrm{types} \tag{8b}$$

$$\sum_{i\in\mathrm{rail}} w_{i,l}^{OUT} - WO_{i,l} = 0 \quad \forall l \in \mathrm{types} \tag{8c}$$

$$\sum_{l\in\mathrm{types}} w_{i,l}^{OUT} = 1 \quad \forall i \in \mathrm{rail} \tag{8d}$$

Finally, the set of equations 9 defines the shunt in constraints.

$$\sum_{\hat{i}=1}^{rail-1} w^{SI}_{\hat{i},j,k} \cdot \sum_{\tilde{i}=\hat{i}+1}^{rail} w^{SI}_{\tilde{i},j,k} \cdot adj_{j,k} \geq adj^{SI}_{j,k} \quad \forall j \in \text{tr}, \quad \forall k \in \text{Pos} \tag{9a}$$

$$\sum_{i\in \text{rail}} \sum_{j\in \text{tr}} \sum_{k\in \text{pos}} w^{\text{SI}}_{i,j,k} - \sum_{k=1}^{pos-1} adj^{SI}_{j,k} = shunts^{SI} \tag{9b}$$

$$\sum_{i\in \text{rail}} w^{\text{SI}}_{i,j,k} \leq 1 \quad \forall j \in \text{tr}, \quad \forall k \in \text{Pos} \tag{9c}$$

$$\sum_{j\in \text{tr}} \sum_{k\in \text{pos}} w^{\text{SI}}_{i,j,k} \leq 1 \quad \forall i \in \text{rail} \tag{9d}$$

$$w^{\text{SI}}_{i,j,k} \leq \text{suit}_{j,k} \quad \forall i \in \text{rail}, \quad \forall j \in \text{tr}, \quad \forall k \in \text{Pos} \tag{9e}$$

3.3 Comparison with the classical approach

To compare our approach to the traditional approach to ML implementation, we apply the above-mentioned MILP model by treating ML^{maint}_i as an input rather than a decision variable. This is modeled as Eq. 10:

$$ML^{maint}_i = pred^{ML,in}_i \quad \forall i \in \text{rail} \tag{10a}$$

This allows for a direct comparison between the two models in terms of both risk and key performance indicators. The modifications include the introduction of a new input, $pred^{ML,in}_i$, set equal to $pred^{ML}_i$ if $maint_i = 0$, and 0 otherwise; the removal of Equation 7d, to provide more flexibility in demand management due to the forced acceptance of the input.

This modification allows the use of Equations 5c and 5d as the risk model, enabling the computation of risk associated with ML-driven decisions.

3.4 Case Study

For our case study, we used a binary classification supervised ML model that predicts rolling stock disruption, training it on real data provided by CFL Multimodal for the years 2021 and 2022. We opted for a Decision Tree (DT) model approach, chosen for better performance in the Recall, which fits our research goal. The model then was tuned to optimize the F1 score, which combines Precision and Recall. The final model, a Tuned Decision Tree (TDT), demonstrated good performance in predicting unplanned maintenance events, with the monthly TEU count and actual mileage being significant predictive features. For this methodology, the probability threshold choice in the TDT model is critical for balancing safety and operational efficiency. This threshold, set between 0 and 1, determines the cutoff at which a data point is classified into one of the two categories based on the predicted probability. A lower threshold leads to more conservative predictions with higher false positives, triggering more preventive maintenance actions. It is important to note that to predict rolling stock disruptions, a variety of ML models, with different input datasets, could have been effectively employed, without any issue in the framework implementation. The instance for the study was created using real statistics from CFL Multimodal for the year 2023, covering 7 different destinations, listed in Table 1. This dataset includes information such as the average number of wagons required per destination, associated revenue, frequency of cancellations, distance required for each journey, and expected travel times. For the case study, we run the models for both the full Risk Function and the ML as an input for the 1-year-old fleet, and each probability threshold ranges between 0.05 and 0.95. The cancellation threshold is 180 minutes, with 15 minutes per shunting operation.

4 Results

	Instance Informations		Risk Function			ML Input		
	MWC	**ML Wag**	**AVG Shunts**	**Risk Term**	**Cancellation**	**AVG Shunts**	**Risk Term**	**Cancellation**
Antwerp	10	0	25.17	1153286.6	100%	18.5	1104504.3	100%
Halkali	12	18	9.78	250061.3	0%		**Unfeasible**	
Kiel	4	6.67	8.56	707510.2	22%	10.72	773810.7	33%
Lyon	6	0	7.33	718661.3	0%	7.83	786285.0	0%
Poznan	8	16.89	18.67	1344728.4	100%	19	1342142.2	100%
Rostock	2	11.61	11.89	644566.5	44%	12.94	619173.1	61%
Trieste	8	0	12	691160.4	0%	12.89	701581.3	89%

Table 1: Risk Function vs ML Input analysis.

Table 1 presents a comparative analysis of the full risk function versus using ML as input for the different destinations. The MWC refers to *Minimum wagon change* and is defined as the minimum number of wagons that have to be changed between the inbound and outbound train to fill the demand. The ML Wag refers to *Wagon Flagged by ML*, representing the average number of wagons flagged as "in risk of disruption" for that specific instance. The risk term is consistently lower for the Risk Function model, except for 2 destinations (Antwerp and Rostock), compared to the other model, showing a better management of the scenarios. Overall, the model exhibits improved risk performance management, averaging a reduction of approximately 5%. Cancellation rates from the Risk Function are either matching or outperforming those of the ML Input. Specifically, in cases like Trieste and Rostock, we have a decrease in the Cancellation rate of 89% and 17% respectively. Regarding the Shunts, the Risk function model consistently performs fewer shunting operations compared to the full ML as an Input model, which is also explanatory of why fewer Cancellations are seen for this instance. Finally, for the destination of Halkali, the ML Input returned Unfeasible solutions due to an elevated number of ML requests (18). Instead, the Risk Function manages not only to offer an optimal solution but also without cancellation. This underscores the resilience and reliability of the Risk Function, specifically in scenarios where ML Input may not yield usable results.

Overall, the Risk Model appears to deliver a more robust and consistent assessment, particularly in terms of risk management and operational reliability.

5 Conclusion

This study presents a framework for using Risk Assessment techniques to improve shunting operations in freight rail transportation. We developed a MILP and a Binary Classification ML model for optimizing shunting operations in this context. We leverage Classification model metrics to enhance the performance of the MILP model. Our findings demonstrate that while ML can provide valuable insights, reliance on these predictions without a risk-informed buffer may lead to operational inefficiencies and even unfeasible scenarios. The robustness of the Risk Function model, which consistently shows improved performance in terms of risk management, highlights the necessity for a broader approach to data-driven models that values reliability and continuous data validation rather than solely inputting its results. This study supports the assumption that a comprehensive risk assessment framework, paired with an OR model, can make the model adaptable and flexible to the dynamic nature of rail operations. Further research includes an improvement of the MILP model, a more long-term analysis for asset assessment, and new data-driven models for rolling stock disruption prediction.

Bibliography

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

Bigi, F., Bosi, T., Pineda-Jaramillo, J., Viti, F., and D'Ariano, A. (2024). Long-term fleet management for freight trains: Assessing the impact of wagon maintenance through simulation of shunting policies. *Journal of Rail Transport Planning & Management*, 29:100430.

Bosi, T., Bigi, F., D'Ariano, A., Viti, F., and Pineda-Jaramillo, J. (2023). Optimal Management of Full Train Load Services in the Shunting Yard: A Comprehensive Study on Shunt-In Shunt-Out Policies. *Computers & Industrial Engineering*, page 109865.

Cadarso, L. and Marín, (2011). Robust rolling stock in rapid transit networks. *Computers & Operations Research*, 38(8):1131–1142.

FAIR Institute (2023). FAIR Institute.

Giacco, G. L., Carillo, D., D'Ariano, A., Pacciarelli, D., and Marín, G. (2014). Short-term rail rolling stock rostering and maintenance scheduling. In *Transportation Research Procedia*.

Haahr, J. and Lusby, R. M. (2017). Integrating rolling stock scheduling with train unit shunting. *European Journal of Operational Research*.

Jaehn, F., Rieder, J., and Wiehl, A. (2015). Minimizing delays in a shunting yard. *OR Spectrum*, 37(2):407–429.

Kamenga, F., Pellegrini, P., Rodriguez, J., and Merabet, B. (2021). Solution algorithms for the generalized train unit shunting problem. *EURO Journal on Transportation and Logistics*, 10:100042.

Kamenga, F., Pellegrini, P., Rogriguez, J., Merabet, B., and Houzel, B. (2019). Train Unit Shunting : Integrating rolling stock maintenance and capacity management in passenger railway stations. *8th International Conference on Railway Operations Modelling and Analysis*.

Kroon, L. G., Lentink, R. M., and Schrijver, A. (2008). Shunting of passenger train units: An integrated approach. *Transportation Science*, 42(4):436–449.

Lentink, R. M. (2006). *Algorithmic Decision Support for Shunt Planning*. PhD thesis, School of Economics, Erasmus University Rotterdam.

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

Richard Freling, Ramon M. Lentink, L. G. K. D. H. and ERIM (2002). Shunting of Passenger Train Units in a Railway Station. *ERIM Report Series Research in Management*.

Spanninger, T., Trivella, A., Büchel, B., and Corman, F. (2022). A review of train delay prediction approaches. *Journal of Rail Transport Planning & Management*, 22:100312.

A Appendix

A.1 Parameters

Shunting yard (SY) parameters:

- $rail$: number of wagons in the train, $rail \in \mathbb{R}$
- tr: number of tracks in the SY $tr \in \mathbb{R}$
- pos: number of positions for tracks in the SY $pos \in \mathbb{R}$
- typ: number of types available in the fleet, $typ \in \mathbb{R}$
- gr: possible number of gravity of disruption, $gr \in \mathbb{R}$
- rev_{sy}: potential revenue of each wagon in the SY, $rev_{sy} \in \mathbb{R}^{tr*pos}$
- $R^{SY,gr}_{disr}$: risk of disruption for each wagon in the SY by gravity, $R^{SY,gr}_{disr} \in \mathbb{R}^{tr*pos*gr}$
- adj_{sy}: adjacent wagons in the SY, $adj_{sy} \in \mathbb{R}^{tr*pos}$
- $suit_{sy}$: suitable wagons in the SY, $suit_{sy} \in \mathbb{R}^{tr*pos}$. This is defined by pre-computing if the wagons in the SY can perform the next trip without surpassing the maximum mileage threshold.
- typ_{sy}: wagon in the shunting yard per type, $typ_{sy} \in \mathbb{R}^{tr*pos*typ}$
- $pred^{ML}_{sy}$: $prob^{ML}_{sy}$ prediction of the ML model for the wagon in the SY. $prob^{ML}_{sy} \in 0, 1^{tr*pos}$

Train parameters:

- $difference$: difference in the wagons of the train by type, $difference \in \mathbb{R}^{rail*typ}$
- $WI_{rail,type}$: type composition for the inbound train, $WI_{rail,type} \in \mathbb{R}^{rail*typ}$
- $WO_{rail,type}$: type composition for the outbound train, $WO_{rail,type} \in \mathbb{R}^{rail*typ}$
- t_{op}: expected operational time of the train, defined as the time before cancellation happens, $t_{op} \in \mathbb{R}$
- $R^{disr}_{rail,gr}$: risk of disruption for each wagon in the train, $R^{disr}_{rail,gr} \in \mathbb{R}^{\setminus\partial\sqsupset\lessdot*\eth\setminus}$
- R^{point}: risk of pointless disruption, $R_{point} \in \mathbb{R}$
- rev_{rail}: potential revenue of the train, $rev_{rail} \in \mathbb{R}^{rail}$
- $maint_{rail}$: wagons to be removed due to maintenance constraint,
- $prob^{ML}_{rail}$: the probability of disruption as computed from the ML model for the wagons in the train, $prob^{ML}_{rail} \in \mathbb{R}^{\setminus\partial\sqsupset\lessdot}$
- $pred^{ML}_{rail}$: $prob^{ML}_{r}ail$ prediction of the ML model for the wagon in the inbound train. $pred^{ML}_{rail} \in 0, 1^{rail}$

ML & Risk parameters:

- C_{shunt}: cost of a single shunt, $C_{shunt} \in \mathbb{R}$
- $C^{corr,maint}_{gr}$: cost of corrective maintenance per gravity, $C^{corr,maint}_{gr} \in \mathbb{R}^{gr}$
- $C^{prev,maint}$: average cost of preemptive maintenance, $C^{prev,maint} \in \mathbb{R}$.
- t_{shunt}: time to perform one shunting operation, $t_{shunt} \in \mathbb{R}$
- C_{canc}: cost of cancelling one train, $C_{canc} \in \mathbb{R}$
- TPR: True Positive Ratio from the ML model, $TPR \in \mathbb{R}$
- TNR: True Negative Ratio from the ML model, $TNR \in \mathbb{R}$

- FPR: False Positive Ratio from the ML model, $FPR \in \mathbb{R}$
- FNR: False Negative Ratio from the ML model, $TPR \in \mathbb{R}$
- P^{gr}: probability of gravity disruption, $P^{gr} \in \mathbb{R}^{gr}$

Specifically, $C^{prev,maint}$ is computed by taking the average cost of preemptive maintenance and adding to it the loss due to unavailability for that period, namely $L^{prev,maint} = t_{out}^{prev} * f_{daily} * R * rev_{rail}$, where t_{out}^{prev} is the expected time out for preemptive maintenance in days, f_{daily} is the daily frequency for a destination. Therefore, $C^{prev,maint} = \mathbb{E}[C^{prev,maint}] + L^{prev,maint}$.

A.2 Sets

- GRAVITE: $\{x \mid x \in \mathbb{Z}, 1 \leq x \leq gr\}$
- TRAIN: $\{x \mid x \in \mathbb{Z}, 0 \leq x \leq rail\}$
- TRACKS: $\{x \mid x \in \mathbb{Z}, 0 \leq x \leq tr\}$
- POS_TRACKS: $\{x \mid x \in \mathbb{Z}, 0 \leq x \leq pos\}$
- TYPES: $\{x \mid x \in \mathbb{Z}, 1 \leq x \leq typ\}$

A.3 Decision Variables

- t_{dep}: expected departure time, $t_{dep} \in \mathbb{Z}^+$
- $Canc^{trig}$: cancellation variable for costs, $Canc^{trigg} \in [0, 1]$
- $Risk$: risk term overall between SI & SO, $Risk \in \mathbb{Z}^+$
- C_{maint}: overall maintenance cost, $C_{maint} \in \mathbb{Z}^+$
- Dem_{rail}: wagons to be SO for demand reason, $Dem_{rail} \in [0, 1]$
- w_{rail}^{SO}: wagons to be shunted out from the train, $w_{rail}^{SO} \in [0, 1]$
- $w_{rail}^{IN,staying}$: wagons that are staying after the SO, $w_{rail}^{IN,staying} \in [0, 1]$
- $shunt^{SO}$: number of shunting operations performed for SO, $shunt^{SO} \in \mathbb{Z}^+$
- adj^{SO}: number of adjacent wagons moved for the SO, $adj^{SO} \in \mathbb{Z}^+$
- $adj_{rail,rail}$: check if the wagons to SO are adjacent, $adj_{rail,rail} \in [0, 1]$
- R_{rail}^{SO}: risk of the train after the SO, $R_{rail}^{SO} \in \mathbb{Z}^+$
- ML_{rail}^{maint}: wagons to be SO due to the preemptive maint, $ML_{rail}^{maint} \in [0, 1]$
- $w_{rail,tr,pos}^{SI}$: wagons to be shunted in *rail* from the SY at *(tr,pos)*, $w_{rail,tr,pos}^{SI} \in [0, 1]$
- adj^{SI}: number of adjacent wagons moved for the SI, $adj^{SI} \in \mathbb{Z}^+$
- $adj_{tr,pos}^{SI}$: flag for adjacent wagon in the SY, $adj_{tr,pos}^{SI} \in [0, 1]$
- $shunt^{SI}$: number of shunting operations performed for SI, $shunt^{SI} \in \mathbb{Z}^+$
- $shunts$: overall number of shunting operations performed, $shunts \in \mathbb{Z}^+$
- w^{SI}: number of wagons to SI, $w^{SI} \in \mathbb{Z}^+$
- w^{moved}: overall number of wagons moved, $w^{moved} \in \mathbb{Z}^+$
- $W_{rail,type}^{out}$: train overall outbound type composition, $W^{out} \in [0, 1]$
- $diff^{SISO}$: difference between SI and SO, $diff^{SISO} \in \mathbb{Z}^+$
- $R_{tr,pos}^{SI}$: risk term of the wagon to SI, $R_{tr,pos}^{SI} \in \mathbb{Z}^+$
- R_{rail}^{SI}: risk term of the wagon to SI (connect with the train), $R_{rail}^{SI} \in \mathbb{Z}^+$
- $R_{rail}^{staying}$: overall risk term of the wagon staying in the train between SI and SO, $R_{rail} \in \mathbb{Z}^+$
- Rev_{rail}: overall train revenue, $Rev_{rail} \in \mathbb{Z}^+$